


## Métricas para Verificación de Autoría y emulación de procesos cognitivos Metrics to assess Authorship Verification and emulation of cognitive processes

O. Pérez-Cortés <sup>a,\*</sup>, H. Gómez-Pozos <sup>a</sup>, H. D. Molina-Ruiz <sup>b</sup>

<sup>a</sup>Área Académica de Computación y Electrónica, Universidad Autónoma del Estado de Hidalgo, 42184, Pachuca, Hidalgo, México.

<sup>b</sup>Escuela Superior de Tepeji del Río, Universidad Autónoma del Estado de Hidalgo, 42855, Noxtongo Tepeji del Río Ocampo, Hidalgo, México.

### Resumen

Se proponen dos medidas cuantitativas, el parámetro de Hurst y la legibilidad, para realizar la verificación de autor. Se determinan las cantidades mencionadas para seis diferentes autores considerando seis obras de cada uno de ellos. Con estos valores se construye un espacio de dos dimensiones donde cada punto corresponde a un único autor; midiendo la *distancia* entre dos puntos de dicho espacio es posible decidir si un texto es atribuible a un autor o no. Adicionalmente dichas medidas proporcionan una interpretación cualitativa, es decir, los resultados son expresables en términos como la *facilidad* al leer un texto y si existen palabras, asociados, a pensamientos, que *persisten* en un texto.

**Palabras Clave:** Verificación de autoría, clasificadores, dimensión fractal, legibilidad.

### Abstract

Two quantitative measures, the Hurst parameter and legibility, are proposed to perform author verification. The quantities mentioned are determined for six different authors considering six works of each of them. With these values, a two-dimensional space is constructed where each point corresponds to a single author; by measuring the *distance* between any two points, it is possible to decide whether a text is attributable to an author or not. Additionally, these measures provide a qualitative interpretation, that is, in terms such as the *easiness* of reading a text and if there are words, associated with thoughts, persisting all over the document.

**Keywords:** Authorship verification, classifiers, fractal dimension, legibility.

### 1. Introducción

Gracias a las Tecnologías de Información y Comunicación (TIC) existen diversas actividades que antes no eran posibles por ejemplo el *e-commerce*, o la mensajería instantánea y otras se facilitan, como trámites gubernamentales, acceso a bases de datos, etc. (Mnushka and Savchenko, 2020; Issac et al., 2020; Kaushik and Bagga, 2021; Ahmad et al., 2019). Estos avances traen consigo diversos escenarios que se transforman en potenciales inconvenientes para los usuarios (Abbasi and Chen, 2005; Eduardo, 2017; Negoita et al., 2019; Agazzi, 2015; Mioara, 2012). Algunos de ellos se engloban en lo que se denomina como *el problema de autoría*. Ejemplos de esto tienen que ver con el enorme flujo de información existente en blogs, redes sociales, *e-mail*, etc. cuyo contenido puede ser una amenaza, *spam*, mensajes racistas, etc. por lo que resulta vital conocer su autoría, o bien extraer meta-información útil en toma de

decisiones (Nirkhi Smita, 2016; Brocardo et al., 2015; Abbasi and Chen, 2005; Nurmohamed et al., 2021; Khezr and Mohan, 2022; Haeussler and Sauermann, 2013).

El *problema de la autoría* se divide en tres ramas. En primer lugar la denominada *Identificación de Autoría* (IA) se emplea para determinar la probabilidad de que un autor sea el responsable de un escrito dado, de entre un conjunto de autores sospechosos (se asume que el verdadero autor se encuentra entre los posibles sospechosos, lo que no es cierto en la mayoría de los escenarios del mundo real) (Okuno et al., 2014; Alhijawi et al., 2018). En segundo lugar la *Verificación de Autoría* (VA) ayuda a confirmar si un sospechoso es o no el autor de un texto particular (Nirkhi Smita, 2016; White and Sprague, 2021; Theophilo et al., 2021; Boenninghoff et al., 2019). Finalmente, la *Caracterización de Autoría* (CA) se utiliza para recopilar meta-información del autor, como el género, antecedentes lingüísticos, edad, etc. a partir de un documento anónimo

\*Autor para correspondencia: obed\_perez@uaeh.edu.mx

**Correo electrónico:** obed\_perez@uaeh.edu.mx (Obed Pérez-Cortés), hmolina@uaeh.edu.mx (Hector Daniel Molina-Ruiz), gpzos@uaeh.edu.mx (Heberto Gómez-Pozos).

(Theophilo et al., 2021; Vivitha Vijayan, 2019). En este trabajo se propone una metodología para atacar el problema de la VA y potencialmente la IA.

La VA es abordada por los investigadores desde dos enfoques. En primer lugar desde el punto de vista lingüístico el texto en cuestión es analizado en función de características como el léxico, la estructura sintáctica, idiosincrasia, análisis de errores, entre otras (Nirkhi Smita, 2016; Ahmed, 2018; Yang et al., 2018). En segundo lugar, las técnicas computacionales enfrentan el problema al convertir el texto en un conjunto de números y tratarlo como una tarea de clasificación (Stanisz et al., 2019; Benzebouchi et al., 2018). Ambos enfoques proporcionan resultados que permiten resolver, parcialmente, el problema de la VA; no obstante, con el primero no existe consenso sobre cuáles, en cantidad y cualidad, son los parámetros lingüísticos que deben considerarse, mientras que en el segundo algunos métodos aún están sujetos a verificación cruzada (Halvani et al., 2019).

En este trabajo se propone una metodología que tiene dos finalidades: contribuir a resolver el problema de la VA por medio de técnicas computacionales y, al mismo tiempo, vincular los datos entregados por un algoritmo numérico y procesos cognitivos como la memoria y aprendizaje. En relación al segundo, se consideran dos cualidades, la *dificultad* al leer un texto y la *persistencia de ideas* las cuales pueden cuantificarse por medio de la *legibilidad*,  $\mu$ , y el parámetro de Hurst,  $H$ , respectivamente. De manera que dado un autor y alguno de sus escritos éste es procesado y mapeado como un punto  $(\mu_0, H_0)$  en un plano cuyos ejes representan las cualidades mencionadas previamente. Por medio de la distancia euclidiana entre puntos será posible hacer inferencias que auxiliarían a la VA; simultáneamente, esta representación ofrece una interpretación en términos de *dificultad* y *persistencia*, cualidades que están ligadas a procesos cognitivos como la memoria, la comprensión-aprendizaje, entre otros (Smith and Kelly, 2016).

### 1.1. La escritura como un proceso cognitivo

Algunos autores consideran que la escritura no es una actividad mecánica que reproduce el pensamiento, sino es una herramienta epistemológica y cognitiva que lo transforma. La acción de escribir pone en marcha muchos procesos que involucran aspectos cognitivos, neurológicos y educativos (Illes et al., 2020; Omidian et al., 2021; White and Sprague, 2021). Particularmente, la legibilidad y el parámetro de Hurst tienen impacto directo procesos cognitivos como la memoria y la comprensión lectora respectivamente. Existe evidencia que indica que que a mayor *dificultad* de lectura menor comprensión (Johnson et al., 2019; Xu et al., 2019; Daneman and Carpenter, 1980) y para una mejor comprensión es indispensable la *persistencia de una idea* a lo largo de un texto. (McNamara et al., 2010; Crossley, 2020).

La *legibilidad* se entiende como la facilidad de leer un texto (Collins-Thompson, 2014; Kincaid et al., 1975). Una escala bien aceptada por la comunidad científica es la reportada por (Kincaid et al., 1975), presenta de manera resumida en la tabla 1. De otra parte, el parámetro de Hurst, cuantifica la *persistencia*, *auto-similaridad* ó *memoria* de un fenómeno (Mishura and Zili, 2018; Fernandes et al., 2015; Mitra, 2012). Se dice que un proceso dado tiene *memoria* cuando sus observaciones en

un tiempo  $t_0$  (presente) dependen de observaciones pasadas, y simultáneamente éste influirá en su estado futuro (Fernandes et al., 2015). Particularmente, para el fenómeno de la escritura (López-Ortega et al., 2020) han planteado que es posible cuantificar la *persistencia de ideas* con el parámetro de Hurst. Esto es, al analizarse lo escrito en ciertos documentos efectivamente depende fuertemente de lo escrito previamente (en el mismo texto), y la probabilidad de que se mantenga ese orden de ideas (dentro del mismo documento) es mayor que aquella de modificarla. Los fenómenos que presentan esta cualidad obtienen un parámetro de Hurst en el intervalo  $[0.5, 1]$ .

Tabla 1: Algunos valores de la escala Flesh-Kinkard

Valor	Nota
90-100	Fácil de leer, entendible para un niño de 11 años.
60-70	Entendible para estudiantes de entre 13-15 años.
30-50	Difícil de leer.
0-30	Muy difícil de leer

Con todo esto, las métricas de *legibilidad* y *parámetro de Hurst* representan un medio para cuantificar la *facilidad* que presentará el leer cierto autor y en qué medida las ideas que presenta *persisten* a lo largo del documento. De tal manera que, tentativamente, para promover una mejor comprensión de lo leído es deseable que el texto de un autor presente altos valores de  $\mu$  y  $H$  (Johnson et al., 2019; Xu et al., 2019; Daneman and Carpenter, 1980; McNamara et al., 2010; Crossley, 2020). Luego, es posible vislumbrar un sistema computacional que incorpore estas métricas,  $\mu$  y  $H$ , y pueda inferir el impacto en los procesos cognitivos citados como ya es realizado por algunos (Davidson, 1994; Sugawara et al., 2020). Al mismo tiempo, al medir la distancia euclidiana en el plano  $\mu$ - $H$  es posible inferir si dos puntos son atribuibles a un mismo autor, contribuyendo así a la VA.

### 1.2. Organización del artículo

La organización de este artículo es la siguiente: en la sección II se presenta la metodología propuesta y un breve resumen del formalismo empleado para calcular las métricas adoptadas. En la sección III se muestran los resultados encontrados y se discuten en la sección IV. Finalmente en la sección V se presentan las conclusiones y trabajo a futuro.

## 2. Metodología

A fin de contribuir a resolver el problema de la VA en el contexto del espacio propuesto,  $\mu$ - $H$ , se debe adoptar la metodología que se describe a continuación. En primer lugar se debe contar con un *corpus* de textos conocidos de diversos autores conocidos. Después debe procesarse cada uno de éstos para calcular su legibilidad y parámetro de Hurst para construir, por cada autor, un conjunto  $A = \{(\mu_0^A, H_0^A), (\mu_1^A, H_1^A) \dots\}$ . Finalmente, se debe calcular un *centroide* para cada conjunto atribuible a un único autor y *mapearse* en el plano  $\mu$ - $H$ ; posteriormente se mide la distancia euclidiana entre cada punto y se debe establecer que efectivamente estén separados. La metodología se representa esquemáticamente por medio de la figura 1. Las ecuaciones empleadas para calcular la legibilidad y el parámetro de Hurst se describen en las secciones 2,1 y 2,2 respectivamente.

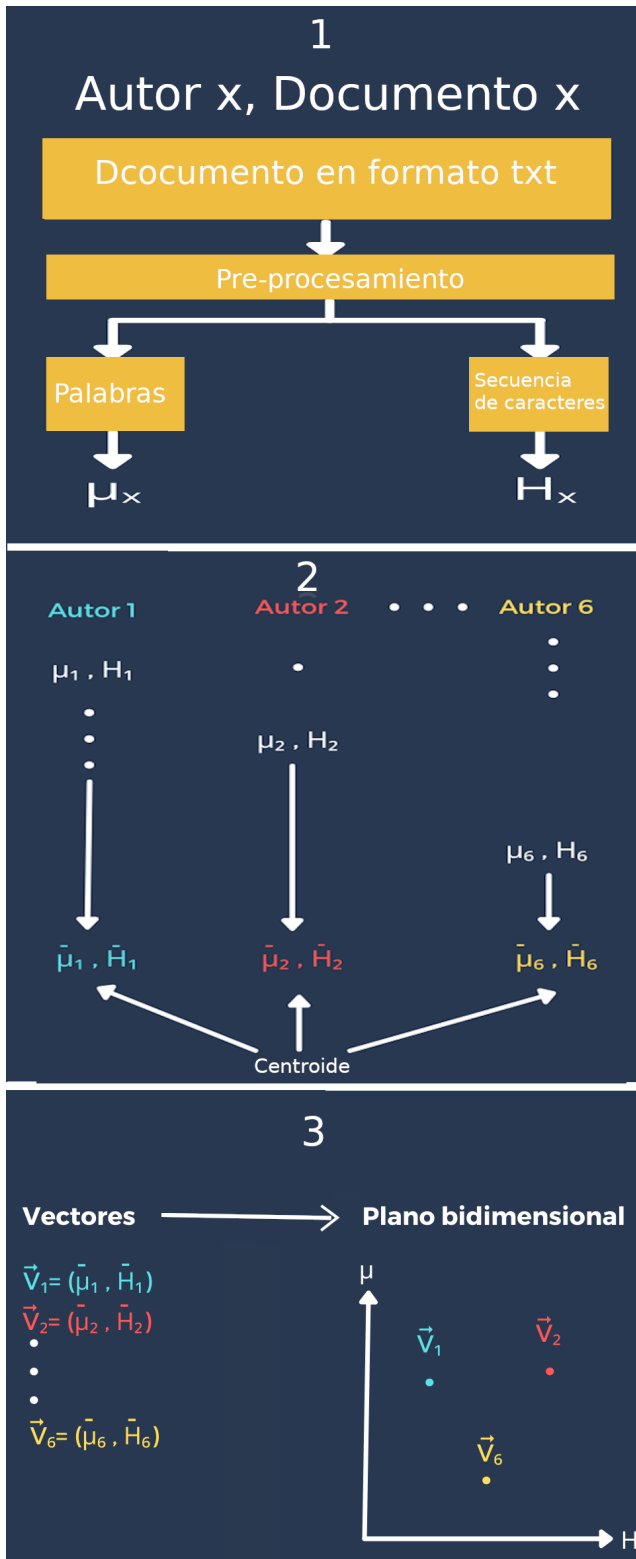


Figura 1: Metodología propuesta para calcular puntos en un espacio bidimensional que identificamos como un único autor.

Para implementar la metodología propuesta se construyó un *corpus* con seis autores y seis textos de cada uno de ellos (36 documentos en total). Todos estos autores son de habla inglesa, vivieron durante los siglos XIX-XX y sus escritos son catalogados como *narrativa*; el listado completo se presenta en el apéndice A. Para obtener las métricas deseadas,  $\mu$  y  $H$ , se procesa cada uno de los textos de manera individual y se lee todo el

documento. Particularmente, al calcular  $\mu$  el texto se considera como un conjunto de palabras, cada una con un cierto número de letras. Para el caso de  $H$ , cada documento se considera como un conjunto de caracteres y cada uno se convierte a su respectivo código ascii, de manera que un documento es expresado como una serie de tiempo.

### 2.1. Legibilidad

Para calcular la legibilidad se emplea la ecuación 1 definida en (López-Ortega et al., 2020).

$$\mu = 100 \frac{N}{N-1} \frac{\bar{x}}{\sigma^2}. \tag{1}$$

Donde  $N$  es la cantidad de palabras en el texto,  $\bar{x}$  es la media del número de letras por palabra, y  $\sigma^2$  es la desviación estándar de dicha media.

### 2.2. Parámetro de Hurst

Existen diversas técnicas para determinar  $H$ , algunas de ellas pueden encontrarse en (Clegg, 2006). En este trabajo se empleó la librería *hurstExp*, creada por Christopher Scholzel, disponible para el software python V=3.8. Tal y como se describió en la sub-sección 1.1  $H$  es un número que se encuentra en el intervalo  $[0, 1]$ . Si  $H = 0,5$  entonces el fenómeno analizado es equivalente a un *random walk*. Si  $0,5 < H < 1$  significa que el fenómeno es *persistente*, esto es, la ocurrencia de secuencias numéricas tiene una gran probabilidad de repetirse. Finalmente, si  $0 < H < 0,5$  entonces se estudia un fenómeno *anti-persistente*, que significa mayor ruido y mayor volatilidad (Zeinali and Pourdarvish, 2022; Bui and Slepaczuk, 2022; Tsionas, 2021; Fenoy et al., 2021).

## 3. Resultados

Los valores encontrados para la legibilidad,  $\mu$ , se distribuyen de manera más o menos homogénea en un pequeño intervalo. Por ejemplo para el autor Ernest Hemingway  $91,8 < \mu < 109,9$ , lo que resulta en una media de 98.8. Para el resto de los autores el resultado es semejante. Considerando al mismo autor, pero ahora para  $H$ , resulta que  $0,59 < H < 0,74$ , por lo que la media aproximada es 0.68. La gráfica completa de los autores se presenta en la figura 2.

También se determinó la *distancia euclidiana* entre los puntos del plano (es decir, entre todos los autores y consigo mismo), dando como resultado la figura 3. La diagonal representa la distancia de un autor consigo mismo, por lo que es 0, la diagonal superior e inferior son simétricas. Para una adecuada representación gráfica se calculó el logaritmo base 10 de dicha distancia (la distancia mínima tiene un *offset* de 0.1 para evitar el 0); de manera que el valor  $-1$  representa la distancia cero.

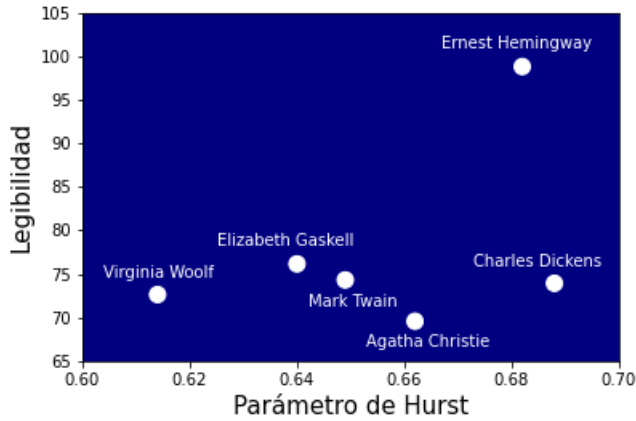


Figura 2: Valores mapeados.

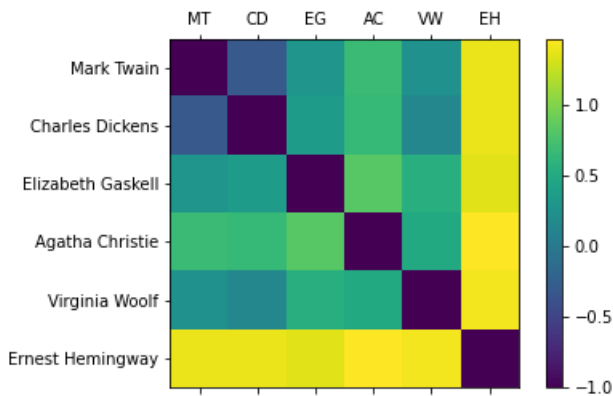


Figura 3: Distancia euclidiana entre los autores.

#### 4. Discusión

La gráfica de la figura 2 sugiere que, al menos parcialmente, es posible asignar un punto en el espacio  $\mu-H$  a cada uno de los autores. Numéricamente es posible distinguirlos y cuantificar su separación, como se demuestra en la figura 3; tal y como se plantea en el trabajo a futuro, sección 5.1, estas cantidades pueden y deben mejorarse (al incorporar un número mayor de autores y obras, entre otras). Sin embargo, no debe olvidarse que para la Inteligencia Artificial el emular los procesos y obtener resultados que un humano puede desempeñar por medio de un sistema computacional no es el único objetivo, sino también que estos sean *indistinguibles* (Pelau et al., 2021; Llargues Asensio et al., 2014; Song et al., 2022). Para los autores de este trabajo este término hace referencia a la habilidad de un sistema artificial de expresar sus resultados, cuantitativos, de manera *amigable* para el ser humano; esto, en principio, aplica para todo individuo y no únicamente para aquellos altamente especializados en alguna materia. De manera que se considera que al proponer métricas que tienen un significado tangible para un individuo (Zraick et al., 2021) se contribuye al dotar de una metodología que a partir de mediciones objetivas expresa sus resultados en términos *amigables*. Dicho con otras palabras se

considera que los resultados son expresables cualitativamente. Esto es, en el marco de las definiciones adoptadas en este trabajo, los textos de los seis autores estudiados son del tipo *persistente*. Esto significa que secuencias de caracteres se repiten a lo largo del texto, en otras palabras, un conjunto de ideas prevalece a lo largo de este; presumiblemente unas pocas bien definidas por el autor. Al mismo tiempo todos poseen una *facilidad* de lectura semejante, a excepción de Ernest Hemingway. Note que los autores estudiados poseen coincidencias, todos ellos vivieron durante un periodo de 150 años y su género es la narrativa. Es de esperarse que su escritura guarde semejanza.

La metodología propuesta permite caracterizar un texto en función de dos cantidades objetivas, de manera que si se determinan estas para dos documentos arbitrarios  $(\mu_1, H_1)$  y  $(\mu_2, H_2)$  es posible medir la distancia entre ambos puntos y así realizar la VA. Más aún, si se incrementa el número de autores y se genera un conjunto amplio de puntos en el espacio  $\mu-H$ , es posible que al calcular  $(\mu_0, H_0)$  de un texto de autor desconocido se proporcione una medida de semejanza con alguno de los autores conocidos y así aportar para la IA.

#### 5. Conclusiones

La metodología propuesta permite atacar el problema de la VA y potencialmente el de la IA empleando en ambos casos la *distancia euclidiana* en el espacio  $\mu-H$ . La distancia permitiría determinar si dos puntos son en realidad variaciones de un tercero o son puntos que no guardan relación. Potencialmente la metodología adoptada puede extenderse y, nuevamente, por medio de la *distancia*, cuantificar a qué punto de los existentes previamente estudiados se *asemeja más* un punto de interés y así ligarlo a uno en lugar de otro. De otra parte, las cualidades *dificultad* y *persistencia de ideas* en un texto, ligadas a procesos cognitivos, son cuantificadas por medio de las métricas propuestas.

##### 5.1. Trabajo a futuro

La metodología propuesta debe ser explorada considerando un mayor número de autores y un mayor número de textos para cada autor. Además que debe estudiarse si es factible el ampliar el número de dimensiones, es decir, si es viable proponer otras cualidades cuantificables, con lo cual se buscaría dotar de *inteligencia humana* a un sistema computacional. También se debe implementar un análisis para determinar cuál es el número mínimo de palabras por documento que son necesarias para obtener resultados confiables.

#### Referencias

Abbasi, A. and Chen, H. (2005). Applying authorship analysis to extremist-group web forum messages. *IEEE Intelligent Systems*, 20(5):67–75.

Agazzi, E. (2015). Bioethics as a paradigm of an ethics for a technological society. *Bioethics Update*, 1(1):3–21.

Ahmad, N., George, R. P., and Jahan, R. (2019). Emerging trends in iot for categorized health care. In *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, volume 1, pages 1438–1441.

Ahmed, H. (2018). The role of linguistic feature categories in authorship verification. *Procedia Computer Science*, 142:214–221. Arabic Computational Linguistics.

Alhijawi, B., Hriez, S., and Awajan, A. (2018). Text-based authorship identification - a survey. In *2018 Fifth International Symposium on Innovation in Information and Communication Technology (ISIICT)*, pages 1–7.

- Benzebouchi, N. E., Azizi, N., Aldwairi, M., and Farah, N. (2018). Multi-classifier system for authorship verification task using word embeddings. In *2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP)*, pages 1–6.
- Boeninghoff, B., Hessler, S., Kolossa, D., and Nickel, R. M. (2019). Explainable authorship verification in social media via attention-based similarity learning. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 36–45.
- Brocardo, M. L., Traore, I., and Woungang, I. (2015). Authorship verification of e-mail and tweet messages applied for continuous authentication. *Journal of Computer and System Sciences*, 81(8):1429–1440.
- Bui, Q. and Slepaczuk, R. (2022). Applying hurst exponent in pair trading strategies on nasdaq 100 index. *Physica A: Statistical Mechanics and its Applications*, 592:126784.
- Clegg, R. G. (2006). A practical guide to measuring the hurst parameter.
- Collins-Thompson, K. (2014). Computational assessment of text readability: A survey of current and future research. *International Journal of Applied Linguistics*, 165(2):97–135.
- Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3):415–443.
- Daneman, M. and Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4):450–466.
- Davidson, J. (1994). The evaluation of computer-delivered natural speech in the teaching of reading. *Computers & Education*, 22(1):181–185.
- Eduardo, M. (2017). Threats of the internet of things in a techno-regulated society: A new legal challenge of the information revolution. *The ORBIT Journal*, 1(1):1–17.
- Fenoy, M., Muñoz-Fernández, G., Pareja Monturiol, J., and Sepúlveda, J. S. (2021). Healthy versus congestive heart failure patients. an approach via the hurst parameter. *Communications in Nonlinear Science and Numerical Simulation*, 103:106004.
- Fernandes, D. A., Neto, M., Soares, L. F., Freire, M. M., and Inácio, P. R. (2015). Chapter 10 - on the self-similarity of traffic generated by network traffic simulators. In Obaidat, M. S., Nicopolitidis, P., and Zarai, F., editors, *Modeling and Simulation of Computer Networks and Systems*, pages 285–311. Morgan Kaufmann, Boston.
- Haeussler, C. and Sauermaun, H. (2013). Credit where credit is due? the impact of project contributions and social factors on authorship and inventorship. *Research Policy*, 42(3):688–703.
- Halvani, O., Winter, C., and Graner, L. (2019). Assessing the applicability of authorship verification methods. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*. ACM.
- Illes, M., Wilson, P., and Bruce, C. (2020). Forensic epistemology: A need for research and pedagogy. *Forensic Science International: Synergy*, 2:51–59.
- Issac, K., Pranay, G., Bharanidharan, N., and Rajaguru, H. (2020). A study on real world implementation and future trends of internet of things. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pages 357–361.
- Johnson, J., Lubo, S., Yedla, N., Aponte, J., and Sharif, B. (2019). An empirical study assessing source code readability in comprehension. In *2019 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 513–523.
- Kaushik, N. and Bagga, T. (2021). Smart cities using iot. In *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pages 1–6.
- Khezz, P. and Mohan, V. (2022). The vexing but persistent problem of authorship misconduct in research. *Research Policy*, 51(3):104466.
- Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., and Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training Command Millington TN Research Branch.
- Llargues Asensio, J. M., Peralta, J., Arrabales, R., Bedia, M. G., Cortez, P., and Peña, A. L. (2014). Artificial intelligence approaches for the generation and assessment of believable human-like behaviour in virtual characters. *Expert Systems with Applications*, 41(16):7281–7290.
- López-Ortega, O., Pérez-Cortés, O., Castillejos-Fernández, H., Castro-Espinoza, F.-A., and González-Mendoza, M. (2020). Written documents analyzed as nature-inspired processes: Persistence, anti-persistence, and random walks-we remember, as along came writing-t. holopainen. *Applied Sciences*, 10(18).
- McNamara, D. S., Crossley, S. A., and McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1):57–86.
- Mioara, M. S. (2012). The impact of technological and communication innovation in the knowledge-based society. *Procedia - Social and Behavioral Sciences*, 51:263–267. The World Conference on Design, Arts and Education (DAE-2012), May 1-3 2012, Antalya, Turkey.
- Mishura, Y. and Zili, M. (2018). 2 - fractional and sub-fractional brownian motions. In Mishura, Y. and Zili, M., editors, *Stochastic Analysis of Mixed Fractional Gaussian Processes*, pages 31–73. Elsevier.
- Mitra, S. K. (2012). Is hurst exponent value useful in forecasting financial time series? *Asian Social Science*, 8(8):111–120.
- Mnushka, O. and Savchenko, V. (2020). Security model of iot-based systems. In *2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, pages 398–401.
- Negoita, O. D., Purcarea, A. A., and Popescu, M. A.-M. (2019). Research on online promoting methods used in a technological society. *Procedia Manufacturing*, 32:1043–1050. 12th International Conference Interdisciplinarity in Engineering, INTER-ENG 2018, 4-5 October 2018, Tirgu Mures, Romania.
- Nirxhi Smita, R. V. Dharaskar, V. M. T. (2016). Authorship verification of online e messages for forensic investigation. *Procedia Computer Science*, 78:640–645.
- Nurmohamed, F. R. H., Voigt, I., Awadpersad, P., Matawlie, R. H., and Gadjradj, P. S. (2021). Authorship decision-making in the field of orthopedic surgery and sports medicine. *Journal of Clinical Orthopaedics and Trauma*, 21:101531.
- Okuno, S., Asai, H., and Yamana, H. (2014). A challenge of authorship identification for ten-thousand-scale microblog users. In *2014 IEEE International Conference on Big Data (Big Data)*, pages 52–54.
- Omidian, T., Siyanova-Chanturia, A., and Biber, D. (2021). A new multidimensional model of writing for research publication: An analysis of disciplinary, intra-textual variation, and I1 versus Ix expert writing. *Journal of English for Academic Purposes*, 53:101020.
- Pelau, C., Dabija, D.-C., and Ene, I. (2021). What makes an ai device human-like? the role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. *Computers in Human Behavior*, 122:106855.
- Smith, A. D. and Kelly, A. (2016). *The Encyclopedia of Adulthood and Aging*. John Wiley & Sons, Inc., first edition.
- Song, M., Xing, X., Duan, Y., Cohen, J., and Mou, J. (2022). Will artificial intelligence replace human customer service? the impact of communication quality and privacy risks on adoption intention. *Journal of Retailing and Consumer Services*, 66:102900.
- Stanisz, T., Kwapie, J., and Drod, S. (2019). Linguistic data mining with complex networks: A stylometric-oriented approach. *Information Sciences*, 482:301–320.
- Sugawara, S., Stenertorp, P., and Aizawa, A. (2020). Benchmarking machine reading comprehension: A psychological perspective.
- Theophilo, A., Giot, R., and Rocha, A. (2021). Authorship attribution of social media messages. *IEEE Transactions on Computational Social Systems*, pages 1–14.
- Tsionas, M. G. (2021). Bayesian analysis of static and dynamic hurst parameters under stochastic volatility. *Physica A: Statistical Mechanics and its Applications*, 567:125647.
- Vivitha Vijayan, S. G. (2019). A survey on author profiling techniques. *International Journal of Computer Sciences and Engineering*, 7:1065–1069.
- White, R. and Sprague, N. (2021). Deep metric learning for code authorship attribution and verification. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1089–1093.
- Xu, Z., Ellis, L., and Umphrey, L. R. (2019). The easier the better? comparing the readability and engagement of online pro- and anti-vaccination articles. *Health Education & Behavior*, 46(5):790–797. PMID: 31216882.
- Yang, M., Chen, X., Tu, W., Lu, Z., Zhu, J., and Qu, Q. (2018). A topic drift model for authorship attribution. *Neurocomputing*, 273:133–140.
- Zeinali, N. and Pourdardvish, A. (2022). An entropy-based estimator of the hurst exponent in fractional brownian motion. *Physica A: Statistical Mechanics and its Applications*, 591.
- Zraick, R. I., Azios, M., Handley, M. M., Bellon-Harn, M. L., and Manchaiah, V. (2021). Quality and readability of internet information about stuttering. *Journal of Fluency Disorders*, 67:105824.

**Apéndice A. Listado de autores y obras**

Autores y sus obras analizadas					
Mark Twain	Charles Dickens	Elizabet Gaskell	Agatha Christie	Virginia Woolf	Ernest Hemingway
Tom Sawyer	Oliver Twist	North and south	Mysterious affairs at styles	Jacob's room	The sun also raises
Adventures of Huckleberry Finn	Great Expectations	Ruth	The murder on the links	Monday or tuesday	For whom the bell tolls
American Clamait	Master Humphery's clock	Curious, if true	The secret adversary	Mrs. Dalloway in bond street	The old man and the sea
Connecticut Yankee in Artur's Court	Old curiosity shop	Wives and daugthers	The man in the brown suit	Night and day	Men without women
Life on the Mississipi	Little Dorrit	Silvia's lovers	Poirot investigates	The commond reader	In our time
The Prince and the Pauper	Christmas Carol	Round the sofa	The secrets of chimneys	The voyage out	A farewell to arms