

BAB IX

Mendeteksi Penulis dalam Sengketa Kepengarangan: Studi Analisis Kepengarangan Barang Bukti Teks Digital



Devi Ambarwati Puspitasari

A. Urgensi Kolaborasi Teknologi dan Linguistik Forensik

Penelitian bahasa kian berkembang, khususnya terkait pengumpulan data bahasa. Penelitian bahasa telah melibatkan banyak teori dan pendekatan, baik pendekatan dari sisi bahasa seperti linguistik, analisis wacana, linguistik fungsional sistemik (LFS), maupun kolaborasi antarbidang teori tersebut. Beberapa waktu terakhir, banyak penelitian bahasa yang sudah menerapkan kolaborasi antara ilmu bahasa dan teknologi. Oleh karena itu, sering dijumpai penelitian yang menghasilkan sebuah sistem informasi, aplikasi, dan produk-produk digital bertema kebahasaan lainnya. Kamus dan korpus adalah salah satu contoh hasil inovasi kolaborasi bahasa dan teknologi yang paling sering ditemukan.

Tidak dapat dimungkiri bahwa sebuah sistem informasi hasil penelitian bahasa yang kemudian didukung oleh teknologi informasi mampu menghasilkan suatu manfaat yang lebih dari penelitian bahasa

D. A. Puspitasari

Badan Riset dan Inovasi Nasional, e-mail: devi018@brin.go.id

© 2023 Badan Riset dan Inovasi Nasional

Puspitasari, D. A. (2023). Mendeteksi penulis dalam sengketa kepengarangan: Studi analisis kepengarangan untuk barang bukti teks digital. Dalam N. P. Sitanggang, *Hukum dalam teroka linguistik* (151–161). Penerbit BRIN.

DOI: 10.55981/brin.737.c682, ISBN: 978-623-8372-44-7, E-ISBN: 978-623-8372-43-0

itu sendiri. Kolaborasi ini menghasilkan manfaat langsung kepada pengguna dari hasil penelitian bahasa. Meskipun belum banyak, periset mulai ramai melakukan hal tersebut. Beberapa yang sering ditemukan adalah penelitian tentang pengumpulan data bahasa, yang kemudian dikemas secara digital dan dilengkapi dengan fitur-fitur otomatis untuk pengguna (Annisa & Hidayatullah, 2017; Indiana & Ramadhani, 2019; Sardiarinto dkk., 2021). Terdapat pula penelitian yang berupaya mengembangkan pangkalan data forensik kebahasaan dengan fitur-fitur otomatis untuk memudahkan pengguna mengenali lini masa kejadian, barang bukti, dan lain-lain. Penelitian semacam ini masih banyak dilakukan untuk mengumpulkan dan menyajikan data bahasa yang kemudian diberi wadah sistem informasi yang sering disebut aplikasi.

Namun, ada hal lain yang sudah harus mulai dilakukan, yaitu membawa penelitian data bahasa ke tahap yang lebih tinggi, yang disebut kecerdasan buatan. Pentingnya membawa kolaborasi bahasa dan teknologi ke ranah kecerdasan buatan semata-mata bertujuan untuk memberikan manfaat yang lebih besar, baik kepada data bahasa itu sendiri maupun kepada komunitas pengguna bahasa. Saat kita menengok platform *dataset* dunia, sudah kita temukan *dataset* bahasa Inggris, Prancis, Jerman, Italia, Spanyol, Hungaria, Arab, dan Mandarin. Dapat kita temukan pula aplikasi yang menggunakan *dataset* tersebut sehingga dapat kita rasakan langsung manfaatnya. Salah satu yang paling umum adalah mesin penerjemah otomatis, mesin analisis sentimen, dan mesin pengubah suara menjadi teks. Terdapat pula aplikasi yang menjembatani komunikasi antara manusia dan komputer dengan mengandalkan *dataset* bahasa dan kecerdasan buatan.

Kembali pada penelitian bahasa yang umumnya adalah pengumpulan dan analisis data, kini muncul tren untuk membawanya ke tahap yang lebih tinggi ke dalam sistem informasi, memanfaatkan korpus linguistik, mewedahi dalam bentuk aplikasi, bahkan menggunakan alat bantu perangkat lunak. Sayangnya, penelitian-penelitian tersebut belum menggunakan data raksasa (*big data*) dan

dibuat hanya untuk satu tujuan khusus sehingga data bahasa yang ada dalam penelitian dan manfaat penelitian tersebut hanya berhenti pada tingkat generalisasi yang cenderung kecil dan umumnya tidak bisa dimanfaatkan di luar penelitian.

Dalam dua dekade terakhir di Eropa dan Amerika, muncul penelitian dengan memanfaatkan data bahasa yang menghasilkan manfaat lebih dan dapat dipakai berulang-ulang, yaitu pembuatan *dataset* bahasa. Penelitian pembuatan *dataset* bahasa tidak terbatas pada bahasa dengan karakter alfabet saja, tetapi juga telah banyak dikembangkan untuk bahasa-bahasa yang menggunakan aksara, seperti Mandarin dan Arab. Bahasa Arab, contohnya, telah dikembangkan sampai pada bahasa isyarat (Latif dkk., 2019). Pengembangan tersebut memanfaatkan *machine learning* dengan tujuan mengembangkan aplikasi berbasis kecerdasan buatan. Begitu banyak *dataset* bahasa yang dibuat di dunia dengan berbagai metode. Hal tersebut disebabkan oleh kondisi dunia yang sedang mempersiapkan diri dengan kemajuan teknologi.

Data bahasa Indonesia dalam bentuk digital sesungguhnya sangat dibutuhkan pada masa mendatang. Saat ini, data bahasa yang dikumpulkan dan diwujudkan dalam bentuk *dataset* sangat dibutuhkan dalam pembuatan teknologi berbasis kecerdasan buatan. Telah banyak *dataset* bahasa dari bahasa-bahasa asing yang dipakai sebagai komponen aplikasi atau program kecerdasan buatan. Namun, sampai saat ini belum banyak penelitian di Indonesia yang melakukan penelitian bahasa dengan menggunakan *big data* dan sampai pada tahap pengklasifikasian data bahasa dengan kecerdasan buatan.

Penelitian teknologi informasi kini memiliki perhatian besar pada bahasa. Ini terbukti dari munculnya cabang ilmu dalam kecerdasan buatan yang fokus pada pemrosesan bahasa alami manusia, yaitu *Natural Language Processing* (NLP) yang makin dibutuhkan untuk menghasilkan sebuah program kecerdasan buatan. Oleh sebab itu, kolaborasi antara penelitian teknologi informasi dan penelitian bahasa makin dibutuhkan. Dewasa ini, *dataset* bahasa adalah salah satu bentuk penelitian hasil kolaborasi keduanya. Penelitian ini sudah ramai dilakukan di luar negeri, tetapi masih sangat jarang di Indonesia.

Terkait manfaat kolaborasi antara teknologi dan linguistik forensik, penelitian linguistik forensik dapat memanfaatkan *machine learning* untuk pengembangan model analisis kepengarangan. Analisis kepengarangan yang dimaksud adalah alat bantu analisis linguistik yang mengolaborasikan unsur semantis dan korpus linguistik untuk menjabarkan suatu pola teks. Lebih khusus, analisis kepengarangan dimanfaatkan dalam bidang linguistik forensik, yaitu untuk mengidentifikasi profil suatu teks. Dalam bidang linguistik forensik, analisis kepengarangan umumnya digunakan untuk mengidentifikasi kebenaran kepemilikan suatu teks semacam surat wasiat, surat pribadi dalam kasus kriminal, dan berbagai jenis teks lainnya.

Dewasa ini, kasus kriminal kerap menghadirkan barang bukti digital dan membutuhkan identifikasi profil dari teks tersebut. Selama ini, identifikasi profil teks dilakukan dengan pendekatan grafologi, yang meyakini bahwa faktor di luar teks dan profil goresan dalam tulisan tangan menjadi hal yang sangat menentukan. Dalam teks digital, yaitu teks yang diproduksi dengan alat elektronik seperti komputer dan telepon seluler, identifikasi profil teks menjadi lebih sulit dan tidak bisa lagi menggunakan pendekatan grafologi.

Authorship analysis adalah pendekatan identifikasi profil teks yang sering dipakai dalam ranah kajian linguistik forensik. Atribusi pada teks anonim tidak didasarkan pada faktor di luar teks, seperti kertas, tinta, jenis, atau dokumen asal. *Authorship analysis* lebih menekankan pertimbangan gaya bahasa sebagai dasar identifikasi (Grieve dkk., 2019). Analisis kepengarangan menentukan penulis dokumen yang kepengarangannya dipermasalahkan, seperti dalam mendeteksi plagiarisme dan kolusi, menghubungkan teks kriminal ke penulis dari daftar potensial penulis, serta membuat profil penulis yang tidak dikenal berdasarkan karakteristik linguistik (Coulthard, 2004). Analisis kepengarangan merupakan proses penelaahan ciri-ciri penggalan teks dalam sebuah konteks forensik untuk menarik simpulan tentang kepengarangannya.

B. Sejarah Analisis Kepengarangan

Kelahiran analisis kepengarangan dalam ranah forensik dimulai oleh Jan Svartvik pada tahun 1968 dalam kasus pernyataan pengakuan dari Timothy Jan Evans. Svartvik menggunakan analisis linguistik dari struktur wacana, yaitu struktur sintaksis, gaya bahasa, register atau daftar kata, dan perbandingan antara bahasa lisan dan bahasa tulisan (Coulthard, 2004). Svartvik mengungkap daftar kata yang biasa digunakan polisi dalam pernyataan Evans. Hal tersebut terbukti dari pernyataan Evans yang ditulis menggunakan register formal, yaitu register tipikal dari teks polisi. Selain itu, Svartvik juga menemukan bahwa isi pidato pernyataan Evans tidak konsisten dengan sisa transkrip wawancara. Pernyataan itu telah dibuktikan sebagai produk dari pengaruh polisi karena ada perbedaan yang sangat tajam dalam gaya dan struktur pernyataan Evans.

Dasar linguistik bahasa diatur oleh aturan yang mencakup produksinya dalam bentuk tulisan dan lisan. Adapun bahasa lisan dan tulisan memiliki bentuk organisasi yang berbeda. Bahasa lisan memiliki klausa yang cenderung pendek dan rasio leksikal dan gramatikal yang rendah. Daftar kata pada bahasa lisan mewakili apa yang terjadi sebagai proses melalui dominasi penggunaan kata kerja. Bahasa tertulis cenderung memiliki klausa yang lebih panjang, kepadatan leksikal yang lebih tinggi, dan mewakili apa yang terjadi sebagai produk melalui nominalisasi. Setiap penutur asli memiliki versi yang berbeda dan memiliki individual bahasa yang mereka gunakan dan tulis, yang disebut dengan idiolek. Idiolek tersebut akan memanasifestasikan dirinya melalui ciri khas dan pilihan idiosinkratik dalam teks (Coulthard, 2004).

Gaya bahasa adalah cara seseorang berbicara atau menulis. Gaya bahasa menggabungkan pola yang terdiri atas sejumlah leksikal, sintaksis, dan fitur karakter yang cenderung muncul secara kolektif dalam komunikasi seseorang. Gaya bahasa itulah yang membedakan setiap individu (Juola, 2008). Pemilihan fitur linguistik adalah kecenderungan, bukan aturan. Jika tidak diperbaiki, penggunaan fitur yang sama akan muncul sepanjang waktu. Namun, setiap fitur

individu mungkin tidak muncul di setiap teks yang diberikan. Analisis kepengarangan membutuhkan fitur kata yang sering dan untuk itu dibutuhkan teks panjang atau banyak contoh teks, yang sayangnya jarang didapatkan dalam kegiatan forensik. Sampai saat ini, penanda gaya bahasa yang andal dan valid belum diidentifikasi. Artinya, belum ada yang disebut dengan sidik linguistik (Juola, 2008).

Pembicara atau penutur cenderung membuat pilihan yang khas dan individual dari kata-kata yang disukai. Jumlah kata leksikal akan sangat banyak muncul, tetapi hanya sedikit memunculkan tata bahasa atau gramatikal. Dengan demikian, seorang pembicara memiliki cakupan yang sangat luas terhadap pilihan kata-kata dan konten, tetapi pilihan tata bahasa yang sangat terbatas terhadap kata-kata tetap tidak bisa dihindari. Untuk alasan ini, atribusi kepenulisan linguistik, terutama teks yang pendek, cenderung fokus pada variasi dalam pemilihan kata-kata leksikal dan seberapa banyak tumpang tindih antara pilihan otoritatif dalam teks yang diketahui dan teks-teks yang dipertanyakan.

Untuk menunjukkan persamaan atau perbedaan antara serangkaian penulis, ahli bahasa forensik mengidentifikasi fitur dari idiolek yang tumpang tindih dengan orang lain. Makin banyak *string* yang dibagikan dalam dua dokumen, makin banyak pula kesamaan pengodean maknanya. Oleh karena itu, makin kecil pula kemungkinan bahwa teks tersebut ditulis oleh pengarang atau individu yang berbeda (Coulthard, 2004). Telah banyak riset mengenai penanda kepengarangan yang telah menunjukkan bahwa kata-kata atau kosakata memiliki kemampuan untuk membedakan teks yang ditulis oleh penulis yang sama atau oleh penulis yang berbeda, terutama *string* kata (Coulthard, 2004; Grant & Baker, 2001; Larner, 2014; Macleod & Grant, 2011; Perkins & Grant, 2018).

Grant (2013) mengaitkan serangkaian pesan teks dalam penyelidikan pembunuhan melalui analisis sistematis terhadap terjadinya ejaan (Macleod & Grant, 2011). Nini (2018) mengukur

kesamaan huruf-huruf pendek yang dihubungkan dengan kasus *Jack the Ripper* berdasarkan urutan kata yang dibagikan. Grieve dkk. (2019) menghubungkan *Bixby Letter* yang terdiri atas 139 kata dengan menerapkan pendekatan kuantitatif baru untuk atribusi kepengarangan suatu teks pendek, yaitu penelusuran *n-gram*. Solusi untuk masalah analisis teks pendek dalam konteks linguistik forensik adalah dengan mempertimbangkan ada tidaknya fitur yang bertentangan dengan frekuensi kata, seperti yang pada awalnya diusulkan oleh Grant (2013) dan Nini (2018). Pendekatan ini terdiri atas kuantifikasi kesamaan di antara dua teks dengan menggunakan koefisien Jaccard, yaitu menghitung fitur di antara dua teks yang kemudian dibagi dengan jumlah total fitur di dalamnya.

C. Langkah Analisis Kepengarangan di dalam Teks Digital

Teks digital adalah teks personal yang ditulis individu dalam sebuah alat digital, seperti unggahan media sosial, surat elektronik, pesan singkat, dan teks-teks tertulis lainnya yang diproses dan hanya dapat dibaca dengan alat digital. Alat digital yang dimaksud adalah perangkat komputer, telepon pintar, dan telepon seluler. Teks digital tidak memiliki personalisasi goresan seperti halnya tulisan tangan di media konvensional. Satu-satunya hal personal yang melekat pada teks digital adalah identitas digital dan gaya bahasa dari teks tersebut. Namun, di era teknologi ini identitas digital dapat direkayasa atau dimodifikasi, yang pembuktiannya juga melibatkan internet serangkaian alat digital lainnya.

Untuk melacak sebuah identitas teks digital, diperlukan perangkat lunak dan perangkat keras yang berkaitan dengan asal teks digital. Beberapa kasus pelacakan identitas digital suatu teks sulit dilakukan, tetapi hampir pasti dapat ditemukan. Namun, profil pemilik atau penulis suatu teks digital tidak bisa diidentifikasi dengan cara yang sama seperti pelacakan identitas digital. Identifikasi penulis atau pemilik teks digital memerlukan pendekatan yang mampu mengenali ciri-ciri personal dalam suatu teks digital. Grant (2013) dan Grieve dkk. (2019) menawarkan suatu alternatif untuk mengidentifikasi profil

suatu teks digital, yaitu analisis sistematis ejaan dan pemanfaatan *n-gram*. Menurut mereka, teks digital dapat dianalisis dengan menganalisis kata-kata yang dihasilkan dengan memperhatikan ejaan dan menghitung kecenderungannya.

Pada langkah pertama analisis kepengarangan, sebuah teks digital yang dipertanyakan penulisnya (Q) perlu dibandingkan dengan sekurang-kurangnya dua referensi teks digital yang telah diketahui jelas penulisnya (K1 dan K2). Sebuah teks Q dapat berbentuk pesan singkat atau surat panjang dari seseorang yang berisi pemerasan, pesan bunuh diri, wasiat, dan hal-hal lainnya. Untuk itu, kita dapat menggunakan teks K1 dan K2 dalam pesan singkat, unggahan media sosial, surat personal, atau bahkan esai milik orang yang disangkakan sebagai penulis teks Q. Ilustrasi perbandingan ini terlihat pada Gambar 9.1.



Gambar 9.1. Ilustrasi Pemilihan Teks Digital

Pada langkah kedua, ekstraksi teks Q dan K dilakukan. Ekstraksi tersebut dapat dilakukan secara manual atau menggunakan bantuan *Corpus Tools*. Hampir semua peranti pengolah korpus memiliki fitur penghitung kata dan *n-gram* yang dapat dimanfaatkan untuk mendapatkan hasil yang lebih cepat dan akurat. *N-gram* merupakan unit analisis yang didefinisikan sebagai urutan dari satu atau lebih bentuk linguistik (misalnya 1-gram, 2-gram) pada setiap tingkat analisis linguistik (misalnya kata dan karakter). Contoh perhitungan *n-gram* dapat dilihat pada Tabel 9.1.

Tabel 9.1 Contoh Menghitung *n-gram*

1 gram	2 gram	3 gram
ya	ya emang	ya emang gitu
oke	oke deh	ok deh sob

Hasil analisis atau pembuktian kepenulisan dari dua set dokumen (teks Q dan K) tidak hanya dapat menemukan kesamaan, tetapi juga membangun sebuah kecenderungan bahwa kesamaan ini khas atau spesifik (Grant, 2013). Meskipun sulit untuk menetapkan ambang batas universal untuk kekhasan suatu teks, *n-gram* dapat menentukan muncul atau tidaknya suatu struktur *leksikogramatika* tertentu dalam teks digital, termasuk frekuensinya.

Langkah ketiga adalah penghitungan skor *n-gram* dengan menggunakan koefisien Jaccard (Jaccard Coefficient—JC). Skor dihitung antara pasangan terkait, yaitu hasil perbandingan antara teks Q dengan teks K1 dan K2. Penghitungan koefisien Jaccard akan menghasilkan ukuran jarak antara nol (0) dan satu (1). Hasil nol menunjukkan bahwa dua teks sepenuhnya berbeda dan hasil satu menunjukkan bahwa mereka identik. Adapun hasil desimal antara nol dan satu menunjukkan variasi antara keduanya. Dapat disimpulkan bahwa hasil analisis *n-gram* dan penghitungan koefisien Jaccard terhadap dua teks akan dikategorikan sebagai identik atau diasumsikan ditulis oleh penulis yang sama jika hasilnya adalah satu atau mendekati satu. Sebaliknya, jika hasilnya makin jauh dari angka satu dan bahkan mendekati nol, dua teks digital tersebut diasumsikan ditulis oleh dua penulis yang berbeda.

Sebagai contoh, dalam pembuktian penghitungan *n-gram* dan koefisien Jaccard, telah dipilih tiga teks berbeda, baik dari segi isi maupun sumbernya. Teks pertama (Q) adalah teks yang diidentifikasi penulisnya. Teks Q adalah sebuah surat pribadi yang dikirim melalui surel, sedangkan teks K adalah teks digital yang diasumsikan sebagai teks yang ditulis oleh penulis yang sama, yaitu penulis Q. Namun, untuk pembuktian secara jelas, dipilihlah dua jenis teks K. Teks K1 adalah teks yang sengaja dipilih dari penulis yang berbeda dari penulis

teks Q dan dari media yang berbeda, yaitu pesan WhatsApp. Teks K2 adalah teks dari penulis yang sama dari teks Q, tetapi dari media yang berbeda, yaitu pesan WhatsApp. Berikut dalam Tabel 9.2 adalah detail ketiga teks tersebut dan hasil perhitungannya.

Tabel 9.2 Data Teks Uji Coba

Teks	Asal Teks Digital	Jumlah Token
Q	surel pribadi	128
K1	pesan WhatsApp	146
K2	pesan WhatsApp	111

Tabel 9.3 Hasil Perhitungan Teks Uji Coba

Teks	Hasil perhitungan <i>n</i> -gram dan JC
Q dan K1	0,004
Q dan K2	0,846

Berdasarkan hasil perhitungan Tabel 9.3 sesuai dengan konsep *n*-gram dan koefisien Jaccard untuk menganalisis penulis atau pengarang suatu teks, dapat disimpulkan bahwa konsep Grant (2013) dan Grieve dkk. (2019) tersebut dapat dipercaya dengan akurasi yang cukup tinggi. Namun, langkah terakhir dalam analisis kepengarangan adalah analisis secara mendalam dengan memperhatikan kembali gaya bahasa dan faktor-faktor linguistik lainnya sebagai upaya untuk memperkuat hasil analisis kepengarangan tersebut.

D. Peran Ahli Bahasa sebagai Penyempurna Analisis Kepengarangan

Untuk membedakan satu penulis dengan penulis lain yang relevan atau menemukan sisi kekhasan teks, dapat dilakukan analisis dengan memilih *n*-gram yang optimal. Penentuan kluster *n*-gram dengan potensi yang besar, khususnya pada teks dengan jumlah token yang banyak, juga akan menentukan akurasi. Sebuah teks digital akan

menyulitkan dan membutuhkan waktu yang cukup lama untuk membedakan penulisnya. Namun, pemakaian sebuah kata atau frasa tertentu dan penggunaan ejaan yang spesifik sangat mungkin cukup sering diulang oleh seorang penulis karena di situlah letak ciri khas personal sebuah tulisan. Hal tersebut dapat ditemukan di kedua sampel yang disengketakan. Walaupun demikian, ada juga kemungkinan penulis cenderung tidak mengulangi kata atau frasa dengan ejaan yang spesifik tersebut. Oleh karena itu, dibutuhkan suatu upaya analisis mendalam untuk memeriksa apakah terdapat penggunaan kata-kata yang khusus, seperti kecenderungan memakai kata *aku* dibandingkan kata *saya*, kecenderungan menulis ejaan kata *aku* dengan dua huruf saja seperti *aq*, dan sebagainya.

Masih banyak faktor-faktor linguistik lainnya yang harus diperhatikan untuk memberikan akurasi dan kesimpulan dalam sebuah analisis kepengarangan teks digital. Masih banyak pula alternatif bantuan lainnya, seperti memanfaatkan *dataset* atau menyusun sebuah model *machine learning* dengan program tertentu. Kolaborasi antara teknologi informasi dan linguistik forensik sangat membuka peluang analisis barang bukti, khususnya barang bukti teks digital. Bantuan teknologi dan analisis linguistik yang mendalam akan menghasilkan suatu riset yang sangat bermanfaat dengan hasil yang cepat dan tepat.

Sebuah *dataset* bahasa Indonesia yang dibuat dari teks-teks digital yang bersifat personal dan diiringi dengan akurasi anotasi yang tinggi serta sebuah model *machine learning* yang dirancang khusus untuk analisis kepengarangan akan membuat proses identifikasi barang bukti teks digital cepat dan efisien. Sebagai upaya penyempurnaan, analisis penguat dari para ahli bahasa terkait kepengarangan suatu teks digital akan membuat nilai kebenaran identifikasi tersebut tinggi.

Daftar Referensi

- Annisaa, R., & Hidayatullah, R. (2017). Aplikasi pembelajaran bahasa Inggris disertai kamus. *Jurnal Informatika, Manajemen dan Komputer*, 9(1), 36–43
- Coulthard, M. (2004). Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, 25(4), 431–447. <https://doi.org/10.1093/applin/25.4.431>

- Grant, T. (2013). Txt 4n6: Method, consistency, and distinctiveness in the analysis of SMS text messages. *Journal of Law and Policy*, 21(2), 467–494. <https://brooklynworks.brooklaw.edu/jlp/vol21/iss2/9>
- Grant, T., & Baker, K. (2001). Identifying reliable, valid markers of authorship: A response to Chaski. *International Journal of Speech, Language and the Law*, 8(1), 66–79. <https://doi.org/10.1558/sll.2001.8.1.66>
- Grieve, J., Clarke, I., Chiang, E., Gideon, H., Heini, A., Nini, A., & Waibel, E. (2019). Attributing the Bixby Letter using n-gram tracing. *Digital Scholarship in the Humanities*, 34(3), 493–512. <https://doi.org/10.1093/llc/fqy042>
- Indiana, B. D., & Ramadhani, I. (2019). Aplikasi pembelajaran bahasa Jawa berbasis android. *CAHAYA Tech: Journal of Computer, Engineering, and Technology* 8(1), 40–68. <https://doi.org/10.47047/ct.v8i1.18>
- Juola, P. (2008). Authorship attribution. *Foundation and Trends® in Information Retrieval*, 1(3), 233–334. <http://dx.doi.org/10.1561/1500000005>
- Larner, S. (2014). *Forensic authorship analysis and the world wide web*. Palgrave Pivot London. <https://doi.org/10.1057/9781137413758>
- Latif, G., Mohammad, N., Alghazo, J., AlKhalaf, R., & AlKhalaf, R. (2019). ArASL: Arabic alphabets sign language dataset. *Data in Brief*, 23, 103777. <https://doi.org/10.1016/j.dib.2019.103777>
- Macleod, N., & Grant, T. (2011, Juli). *Whose Tweet? Authorship analysis of micro-blogs and other short-form messages* [Presentasi]. 10th Biennial conference International Association of Forensic Linguists.
- Nini, A. (2018). An authorship analysis of the Jack the Ripper letters. *Digital Scholarship in the Humanities*, 33(3), 621–636. <https://doi.org/10.1093/llc/fqx065>
- Perkins, R., & Grant, T. (2018). Native language influence detection for forensic authorship analysis: Identifying L1 persian bloggers. *International Journal of Speech, Language and the Law*, 25(1), 1–20. <https://doi.org/10.1558/ijssl.30844>
- Sardiarinto, S., Sari, A. B. P., Iswahyuni, D., Andriani, A., Saputro, E., & Rapiyanta, P. T. (2021). Rancang bangun aplikasi pembelajaran bahasa Inggris berbasis android bagi guru SDIT Anak Sholeh. *Bianglala Informatika*, 9(2), 90–94. <https://doi.org/10.31294/bi.v9i2.10772>