

Plaģiātu noteikšana

Plaģiātu atklāšana ir process, kurā var atrast darbā vai dokumentā plaģiāta gadījumus. Plaģi izplatīta datoru izmantošana un interneta parādīšanās ir atvieglējusi citu cilvēku darba plagēšanu. Lielākā daļa plaģiātu gadījumu ir atrodami akadēmiskajā vidē, kur dokumenti parasti ir esejas vai atskaites. Tomēr plaģiātu var atrast praktiski jebkurā jomā, ieskaitot romānus, zinātniskos darbus, mākslas dizainu un pirmkodu.

Plaģiātu atklāšana var būt manuāla vai ar programmatūru saistīta palīdzība. Manuāla noteikšana prasa ievērojamas pūles un lielisku atmiņu, un tas ir nepraktisks gadījumos, kad jāsalīdzina pārāk daudzi dokumenti, vai arī oriģinālie dokumenti nav pieejami salīdzināšanai. Izmantojot ar programmatūru saistītu atklāšanu, iespējams, tiek salīdzinātas plašās dokumentu kopijas, tādējādi daudz sekmīgāk atklāt.

Plaģiātu prakse, izmantojot pietiekamu vārdu aizvietošanu, lai izvairītos no atklāšanas programmatūras ir pazīstama kā rogeting. [1]

Ar programmatūru saistīta atklāšana

Ar datorizētu plaģiātu noteikšana (CaPD) ir informācijas izguves (IR) uzdevums, kuru atbalsta specializētas IR sistēmas, ko dēvē par plaģiāta noteikšanas sistēmu (PDS).

Teksta dokumentos

Teksta plaģiātes noteikšanas sistēmas izmanto vienu no divām vispārīgām atklāšanas metodēm, no kurām viena ir ārēja, otrā ir būtiska [2]. Ārējās detektoru sistēmas salīdzina aizdomīgu dokumentu ar atsauces kolekciju, kas ir dokumentu kopums, ko uzskata par īstu. [3] Pamatojoties uz izvēlētu dokumenta modeli un iepriekš definētiem līdzības kritērijiem, atklāšanas uzdevums ir iegūt visus dokumentus, kuros ir teksts, kas ir līdzīgs grādam virs izvēlēta sliekšņa, lai teksts aizdomīgā dokumentā. [4] Iekšējā PDS analizē tikai novērtējamo tekstu, neveicot salīdzināšanu ar ārējiem dokumentiem. Šīs pieejas mērķis ir atpazīt autora unikālā rakstīšanas stila izmaiņas kā potenciālā plaģiāta indikatoru [5]. PDS nespēj droši identificēt plaģiātu bez cilvēka sprieduma. Līdzības tiek aprēķinātas ar iepriekš noteiktu dokumentu modeļu palīdzību un var būt nepatiesi pozitīvi. [6] [7] [8] [9] [10]

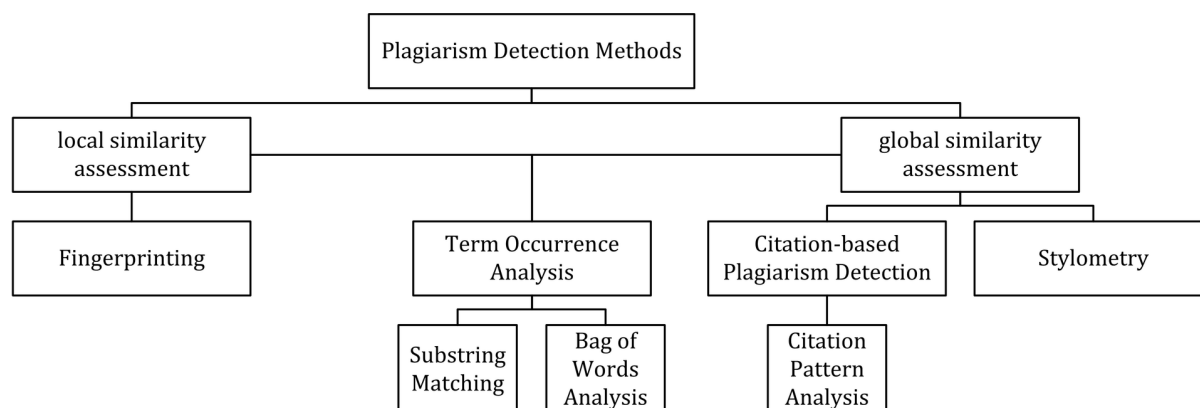
Augstākās izglītības apstākļu efektivitāte

Pētījums tika veikts, lai pārbaudītu plaģiātistikas noteikšanas programmatūras efektivitāti augstākajā izglītībā. Viena pētījuma daļa piešķīra vienai studentu grupai rakstīt papīru. Šie skolēni vispirms izglītoti par plaģiātu un informēja, ka viņu darbs jāīsteno, izmantojot plagija noteikšanas sistēmu. Otrai studentu grupai tika piešķirts rakstīt papīru bez informācijas par plaģiātu. Pētnieki cerēja atrast zemākas likmes pirmajā grupā, bet abās grupās atrada aptuveni tādas pašas plaģiāta likmes. [11]

Pieejas

Tālāk sniegtajā attēlā ir aprakstīta visu atklāšanas pieeju klasifikācija, kas pašlaik tiek izmantota datorizētā plaģiāta noteikšanai. Pieejas raksturo līdzības novērtēšanas veids, ko tās apņemas: globāla vai vietēja.

Globālās līdzības novērtēšanas pieejas izmanto pazīmes, kas ņemtas no lielākām teksta daļām vai dokumenta kopumā, lai aprēķinātu līdzību, savukārt vietējās metodes pārbauda tikai iepriekš atlasītos teksta segmentus kā ievadi.



Datorizētu plaģiātu noteikšanas metožu klasifikācija

Pirkstu nospiedums

Fingerprinting ir pašlaik visplašāk pielietotā pieeja plaģiātu noteikšanai. Šī metode veido reprezentatīvu dokumentu apkopojumu, izvēloties vairāku apakšstrāvu kopu (n gramus) no tiem. Komplekti atspoguļo pirkstu nospiedumus, un to elementi sauc par minutiae. [12] [13]

Pareizu dokumentu pārbauda par plaģiātu, aprēķinot tā pirkstu nospiedumu un uzdodot minimālas piezīmes, izmantojot iepriekš sagatavotu pirkstu nospiedumu indeksu visiem atsauces kolekcijas dokumentiem. Minūtes, kas atbilstu citiem dokumentiem, norāda kopīgus teksta segmentus un ierosina potenciālu plaģiātu, ja tie pārsniedz izvēlēto līdzības sliekšni [14]. Aprēķinu resursi un laiks ir ierobežojošie faktori pirkstu nospiedumiem, tāpēc šī metode parasti salīdzina tikai daļēju elementu apakškopu, lai paātrinātu aprēķinu un ļauj veikt pārbaudes ļoti lielā vākšanā, piemēram, internetā. [12]

Stringa saskaņošana

Stīgu saskaņošana ir izplatīta pieeja, ko izmanto datorzinātnēs. Pieliekot plaģiātu atklāšanas problēmu, dokumenti tiek salīdzināti ar vārdnīcas tekstu pārklāšanos. Lai risinātu šo uzdevumu, ir piedāvātas daudzas metodes, no kurām dažas ir pielāgotas ārējai plaģiātistikas noteikšanai. Pārbaudot aizdomīgu dokumentu šajā iestatījumā, ir nepieciešams aprēķināt un saglabāt efektīvi salīdzināmus pārskatus par visiem referācijas kolekcijas dokumentiem, lai tos salīdzinātu pa pāriem. Parasti šim uzdevumam ir izmantoti sufiksu dokumentu modeļi, piemēram, sufiksu koki vai piedēkļu vektori. Tomēr apakšvirknes atbilstība joprojām ir dārga, un tas padara to par neefektīvu risinājumu lielu dokumentu krājumu pārbaudīšanai. [15] [16] [17]

Vārdu maisiņš

Vārdu analīzes soma atspoguļo vektoru telpas iegūšanas - tradicionālās IR koncepcijas - pieņemšanu plaģiātu noteikšanas jomā. Dokumenti tiek attēloti kā viens vai vairāki vektori, piemēram, dažādām dokumenta daļām, kuras tiek izmantotas pāriem saprotamiem aprēķiniem. Līdzības aprēķināšana tad var balstīties uz tradicionālo līdzību kosmētikas līdzekli vai sarežģītākiem līdzības pasākumiem. [18] [19] [20]

Atsauču analīze

Citation-based plagiarism detection (CbPD) [21] balstās uz atsauču analīzi un ir vienīgā pieeja plaģiātu noteikšanai, kas nebalstās uz teksta līdzību. [22] CbPD izskata citātu tekstu un atsauces informāciju tekstos, lai identificētu līdzīgus paraugus atsauces secībās. Tādējādi šī pieeja ir piemērota zinātniskiem tekstiem vai citiem akadēmiskiem dokumentiem, kuros ir citāti. Atsauču analīze, lai atklātu plaģiātu, ir relatīvi jauna koncepcija. To nav pieņēmusi komerciāla programmatūra, bet ir pirmais no citātu balstītas plaģiātiskās atklāšanas sistēmas

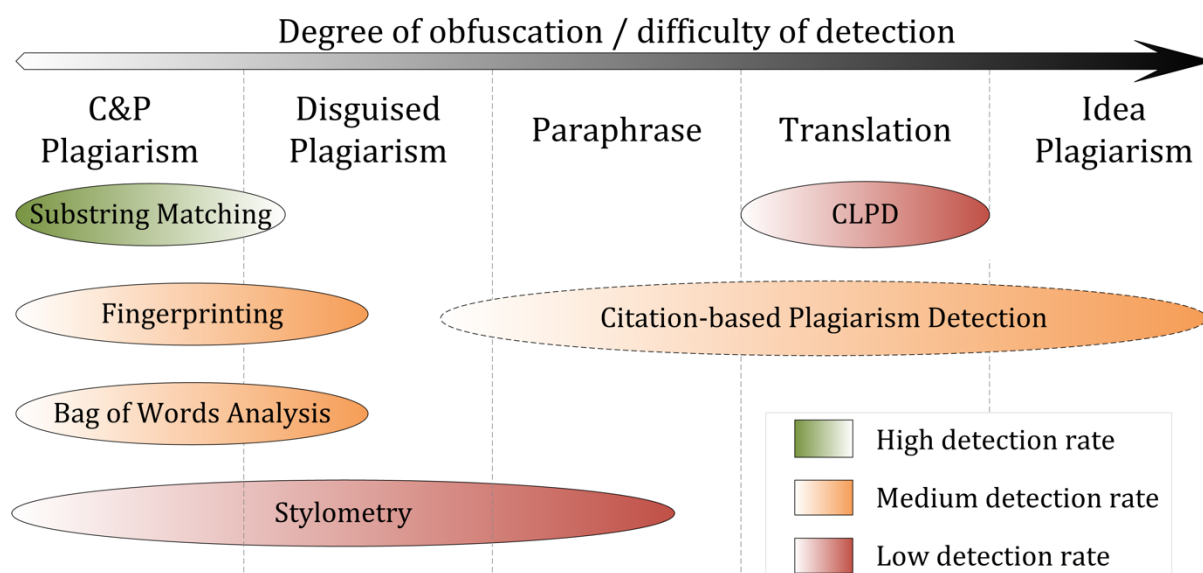
prototips [23]. Pārbaudīto dokumentu citāds secinājums un tuvums ir galvenie kritēriji, kas tiek izmantoti, lai aprēķinātu atsaucē modeļa līdzību. Citēšanas modeļi atspoguļo pēctecības, kas ne tikai satur citātus, kas salīdzināti ar salīdzinātajiem dokumentiem. [22] [24] Tiek uzskatīts, ka faktori, ieskaitot kopējo citātu absolūto skaitu vai relatīvo daļu modeli, kā arī varbūtība, ka citāti kopīgi parādās dokumentā, ir arī modeļu līdzības pakāpes kvantificēšana. [22] [24] [25] [26]

Stilometrija

Stilometrijā tiek izmantotas statistiskās metodes, lai noteiktu autora unikālo rakstīšanas stilu [27] [28], un to galvenokārt izmanto autorības piešķiršanai vai raksturīgajai CaPD. Veidojot un salīdzinot stilometriskos modeļus dažādiem teksta segmentiem, var noteikt fragmenti, kas stilistiski atšķiras no citiem, tādējādi potenciāli plaģiātisti. [5]

Veiktspēja

Plaģiātistikas noteikšanas sistēmu salīdzinošie novērtējumi [3] [29] [30] [31] [32] [33] norāda, ka to veiktspēja ir atkarīga no tā, kāda veida plaģiāts pastāv (sk. Attēlu). Izņemot citēšanas modeļa analīzi, visas atklāšanas metodes balstās uz teksta līdzību. Tādēļ simptomātiski ir tas, ka noteikšanas precizitāte mazina plaģiāta gadījumus.



Atklāšanas izpildījums CaPD pieejām, atkarībā no uzrādītā plaģiātisma tipa

Literārās kopijas, zināms kā kopēšanas un ielīmēšanas (c&p) plaģiātisms, vai nedaudz apslēpti plaģiātisma gadījumi var būt atklāti ar augstu precizitāti, izmantojot pašreizējo ārējo PDS, ja avots ir pieejams programmatūrai. It īpaši apakšvirkņu salīdzināšanas procedūras sasniedz labu sniegumu c&p plaģiātismam, tā kā tās parasti izmanto bezzudumu dokumentu modeļus, tādus kā sufiksu kokus. Veiktspēja sistēmai, kas izmanto pirkstu nospiedumu noņemšanas vai vārdu somas analīzi, lai atklātu kopijas, ir atkarīga no informācijas zudumiem, kas radušies, lietojot dokumenta paraugu. Piemērojot elastīgās gabalu un atlasēs stratēģijas, tās ir vairāk spējīgas noteikt apslēpta plaģiātisma mērenas formas, kad salīdzina ar apakšvirkņu salīdzināšanas procedūrām.

Iekšējā plaģiātisma noteikšana, izmantojot stilometriju, var zināmā mērā pārvarēt tekstuālās līdzības robežas, salīdzinot valodniecisko līdzību. Ņemot vērā, ka stilistiskās atšķirības starp plaģiātisma un oriģinālajām daļām ir ievērojamas un var tikt uzticami identificētas, stilometrija var palīdzēt identificēt apslēptu un pārfrāzētu plaģiātismu. Stilometrijas veida salīdzinājumus visticamāk neizdosies izmantot gadījumos, kad daļas ir tik spēcīgi pārfrāzētas, ka tās sāk vairāk līdzināties plaģiatora personīgajam rakstīšanas stilam, vai tekstu ir sastādījuši vairāki autori. Rezultāti no Starptautiskajām Plaģiātisma Noteikšanas Sacensībām, kas norisinājās 2009., 2010. un 2011. gadā [3][32][33], tā pat kā Stein veiktie eksperimenti [34] liecina, ka stilometriskās analīzes rezultāti ir uzticami tikai dokumentiem, kuru garums ir ap dažu tūkstošu vai desmitiem tūkstošu vārdu, kas ierobežo šīs metodes piemērotību CaPD iestatījumiem.

Arvien vairāk palielinās pētījumu skaits metodēm un sistēmām, kas ir spējīgas noteikt tulkotos plaģiātismus. Patreiz starpvalodu plaģiātismu noteikšana (CLPD) tiek uzskatīta par nepabeigtu tehnoloģiju [35], un atbilstošas sistēmas nav bijušas spējīgas sasniegt apmierinošus rezultātus praksē. [31]

Uz citātiem balstīta plaģiātisma noteikšana, izmantojot citēšanas paraugu analīzes, ar labākiem rezultātiem spēj identificēt spēcīgākas parafrāzes un tulkojumus, salīdzinot ar pārējām plaģiātismu noteikšanas pieejām, jo tā ir neatkarīga no tekstuālajām īpašībām. [22][25] Tomēr tā kā uz citātiem balstīta analīze ir atkarīga no pietiekamas atsauces informācijas pieejamības, tā ir izmantojama tikai uz akadēmiskiem tekstiem. Tā atpaliek no uz tekstu bāzētām pieejām, ja nepieciešams atklāt īsakus plaģiātisma fragmentus, kuras ir piemērotas nokopēšanas un

ielīmēšanas plaģiātismiem; pēdējais attiecas uz nedaudz izmainītu vairāku fragmentu sajaukšanu no dažādiem avotiem. [36]

Programmatūra

Plaģiātismu atklāšanas programmatūras projektēšanu lietošanai ar teksta dokumentiem raksturo vairāki faktori:

| Faktors | Apraksts un alternatīvas |
|--|---|
| Meklēšanas telpa | Publiskā internetā, lietojot meklētājprogrammas / institucionālas datubāzes / vietējās, sistēmas specifiskas datubāzes. |
| Analīzes laiks | Laika starpība starp brīdi, kad dokuments ir iesniegts un brīdi, kad ir pieejami rezultāti. |
| Dokumenta ietilpība / daudzuma apstrāde | Skaits dokumentu, kurus sistēma var apstrādāt konkrētā laika vienībā. |
| Intensitātes pārbaude | Cik bieži un kuriem dokumentu fragmentu tipiem (paragrāfiem, teikumiem, teikumiem ar noteiktu vārdu skaitu) sistēma vaicā ārējiem resursiem, tādiem kā meklētājprogrammām. |
| Salīdzināšanas algoritmu tips | Algoritmi, kas nosaka veidu, kā dokumenti tiek savstarpēji salīdzināti. |
| Precizitāte un pamanīšana | Dokumentu skaits, kas tika pareizi atzīmēti kā plaģiātismi, salīdzinot ar kopējo skaitu atzīmēto dokumentu un kopējo skaitu dokumentu ar plaģiātismiem. Augsta precizitāte nozīmē to, ka tikai daži kļūdaini pozitīvi tika atrasti, un augsta pamanīšana nozīmē to, ka tikai daži kļūdaini negatīvi palika neatklāti. |

Vairums liela mēroga plaģiātismu atklāšanas sistēmu izmanto lielas, iekšējas datubāzes (papildus citiem resursiem), kas turpina palielināties pēc katra analīzei iesniegtā dokumenta. Tomēr šī iezīme tiek uzskatīta kā studentu autortiesību pārkāpumu.

Pirmkodā

Arī plaģiātisms datora pirmkodā ir bieži sastopams, un tā atklāšanai nepieciešams izmantot citādākus rīkus nekā dokumentos esošo tekstu salīdzināšanai. Ir veltīts nozīmīgs pētījums akadēmiska avota koda plaģiātismam. [37]

Atšķirīgs pirmkoda plaģiātisma aspekts ir tas, ka nav pieejamu esejas dzirnavu, kādas var atrast tradicionālā plaģiātismā. Tā kā vairumā programmēšanas uzdevumu studentiem ir noteiktas īpašas prasības, ir sarežģīti atrast eksistējošas programmas, kuras tām jau atbilst. Vairums studentu plaģiātiem izvēlas izmantot savu vienaudžu darbus, jo integrēt ārēju kodu ir bieži vien sarežģītāk nekā rakstīt to no praktiski nekā.

Saskaņā ar Roy un Cordy [38] pirmkoda līdzības noteikšanas algoritmi var būt klasificēti, balstoties uz vai nu

- Simbolu virknes – meklē konkrētas segmentu tekstuālās saistības, piemēram, piecu vārdu grupu. Ātri, taču tos var sajaukt, pārsaucot identifikatorus.
- Tokeni – līdzīgi kā ar simbolu virknēm, bet, izmantojot leksiku, lai vispirms programmu pārveidotu tokenos. Tas izņem atstarpes, komentārus un identifikatoru vārdus, padarot sistēmu spēcīgāku vienkāršām tekstu aizstāšanām. Šajā līmenī strādā vairums akadēmisko plaģiātismu atklāšanas sistēmu, izmantojot dažādus algoritmus, lai izmērītu līdzību starp tokenu secību.
- Parsēšanas koki – izveido un salīdzina parsēšanas kokus. Tas ļauj atrast augstāka līmeņa līdzības. Piemēram, koku salīdzinājums var normalizēt nosacītus izteikumus un detektēt savstarpēji ekvivalentas konstrukcijas.
- Programmas Atkarību Grafi (PDGs) – PDG uztver patieso kontroles plūsmu programmā un ļauj izvietot daudz augstāka līmeņa ekvivalences, lielāki izdevumi sarežģītībā un laika aprēķinos.
- Metrika – metrika uztver koda segmentu “punktus” attiecīgi konkrētiem kritērijiem; piemēram, “loku un nosacījumu skaits” vai “dažādu izmantoto mainīgo skaits”. Metrikas ir vienkāršas, lai izrēķinātu un ātri tiktu salīdzinātas, bet var arī novest pie kļūdainiem pozitīviem: divi fragmenti ar vienādu metriku skaitu var izpildīt pilnīgi atšķirīgas lietas.

- Hibrīda pieejas – piemēram, parsēšanas koki + sufiksu koki var apvienot parsēšanas koku detektēšanas ietilpību ar sufiksu koku piešķirto ātrumu, virknes saskaņošanas tipa datu struktūra.

Iepriekšējā klasifikācija bija izveidota koda uzlabošanai, pārveidošanai un nav domāta akadēmiska plaģiātisma atklāšanai (pārveidošanas ir būtiska nozīme, lai izvairītos no koda dublikātiem, literatūrā norādīti kā koda kloni). Augstāk minētās pieejas ir efektīvas pret dažādu līmeņu līdzībām; zema līmeņa līdzība attiecas uz identisku tekstu, kamēr augsta līmeņa līdzība var būt saistīta ar līdzīgām specifikācijām. Akadēmiskos uzstādījumos, kad visiem studentiem ir jāprogrammē, balstoties uz vienām un tām pašām specifikācijām, funkcionāli ekvivalents kods (ar augsta līmeņa līdzību) ir noteikti sagaidāms, un tikai zema līmeņa līdzība tiek uzskatīta kā pierādījums krāpšanai.

Atsauces

1. Grove, Jack (7 August 2014). "Sinister buttocks? Roget would blush at the crafty cheek Middlesex lecturer gets to the bottom of meaningless phrases found while marking essays". Times Higher Education. Retrieved 15 July 2015.
2. Stein, Benno; Koppel, Moshe; Stamatatos, Efstathios (Dec 2007), "Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection PAN'07" (PDF), SIGIR Forum, 41 (2), doi:10.1145/1328964.1328976
3. Potthast, Martin; Stein, Benno; Eiselt, Andreas; Barrón-Cedeño, Alberto; Rosso, Paolo (2009), "Overview of the 1st International Competition on Plagiarism Detection", PAN09 - 3rd Workshop on Uncovering Plagiarism, Authorship and Social Software Misuse and 1st International Competition on Plagiarism Detection (PDF), CEUR Workshop Proceedings, 502, pp. 1–9, ISSN 1613-0073, archived from the original (PDF) on 2 April 2012
4. Stein, Benno; Meyer zu Eissen, Sven; Potthast, Martin (2007), "Strategies for Retrieving Plagiarized Documents", Proceedings 30th Annual International ACM SIGIR Conference (PDF), ACM, pp. 825–826, doi:10.1145/1277741.1277928, ISBN 978-1-59593-597-7
5. Meyer zu Eissen, Sven; Stein, Benno (2006), "Intrinsic Plagiarism Detection", Advances in Information Retrieval 28th European Conference on IR Research, ECIR 2006, London, UK, April 10–12, 2006 Proceedings (PDF), Lecture Notes in Computer Science, 3936, Springer, pp. 565–569, doi:10.1007/11735106_66

6. Bao, Jun-Peng; Malcolm, James A. (2006), "Text similarity in academic conference papers", 2nd International Plagiarism Conference Proceedings (PDF), Northumbria University Press
7. Clough, Paul (2000), Plagiarism in natural and programming languages an overview of current tools and technologies (PDF) (Technical Report), Department of Computer Science, University of Sheffield, archived from the original (PDF) on 18 August 2011
8. Culwin, Fintan; Lancaster, Thomas (2001), "Plagiarism issues for higher education" (PDF), *Vine*, 31 (2): 36–41, doi:10.1108/03055720010804005, archived from the original (PDF) on 5 April 2012
9. Lancaster, Thomas (2003), Effective and Efficient Plagiarism Detection (PDF) (PhD Thesis), School of Computing, Information Systems and Mathematics South Bank University[permanent dead link]
10. Maurer, Hermann; Zaka, Bilal (2007), "Plagiarism - A Problem And How To Fight It", Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007, AACE, pp. 4451–4458
11. Youmans, Robert J. (November 2011). "Does the adoption of plagiarism-detection software in higher education reduce plagiarism?". *Studies in Higher Education*. 36 (7): 749–761. doi:10.1080/03075079.2010.523457.
12. Hoad, Timothy; Zobel, Justin (2003), "Methods for Identifying Versioned and Plagiarised Documents" (PDF), *Journal of the American Society for Information Science and Technology*, 54 (3): 203–215, CiteSeerX 10.1.1.18.2680 Freely accessible, doi:10.1002/asi.10170
13. Stein, Benno (July 2005), "Fuzzy-Fingerprints for Text-Based Information Retrieval", Proceedings of the I-KNOW '05, 5th International Conference on Knowledge Management, Graz, Austria (PDF), Springer, Know-Center, pp. 572–579
14. Brin, Sergey; Davis, James; Garcia-Molina, Hector (1995), "Copy Detection Mechanisms for Digital Documents", Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data (PDF), ACM, pp. 398–409, doi:10.1145/223784.223855, ISBN 1-59593-060-4
15. Monostori, Krisztián; Zaslavsky, Arkady; Schmidt, Heinz (2000), "Document Overlap Detection System for Distributed Digital Libraries", Proceedings of the fifth ACM conference on Digital libraries (PDF), ACM, pp. 226–227,

doi:10.1145/336597.336667, ISBN 1-58113-231-X, archived from the original (PDF) on 15 April 2012, retrieved 7 October 2011

16. Baker, Brenda S. (February 1993), On Finding Duplication in Strings and Software (Technical Report), AT&T Bell Laboratories, NJ, archived from the original (gs) on 30 October 2007
17. Khmelev, Dmitry V.; Teahan, William J. (2003), "A Repetition Based Measure for Verification of Text Collections and for Text Categorization", SIGIR'03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, pp. 104–110, CiteSeerX 10.1.1.9.6155 Freely accessible, doi:10.1145/860435.860456
18. Si, Antonio; Leong, Hong Va; Lau, Rynson W. H. (1997), "CHECK: A Document Plagiarism Detection System", SAC '97: Proceedings of the 1997 ACM symposium on Applied computing (PDF), ACM, pp. 70–77, doi:10.1145/331697.335176, ISBN 0-89791-850-9
19. Dreher, Heinz (2007), "Automatic Conceptual Analysis for Plagiarism Detection" (PDF), Information and Beyond: The Journal of Issues in Informing Science and Information Technology, 4: 601–614
20. Muhr, Markus; Zechner, Mario; Kern, Roman; Granitzer, Michael (2009), "External and Intrinsic Plagiarism Detection Using Vector Space Models", PAN09 - 3rd Workshop on Uncovering Plagiarism, Authorship and Social Software Misuse and 1st International Competition on Plagiarism Detection (PDF), CEUR Workshop Proceedings, 502, pp. 47–55, ISSN 1613-0073, archived from the original (PDF) on 2 April 2012
21. Gipp, Bela (2014), Citation-based Plagiarism Detection, Springer Vieweg Research, ISBN 978-3-658-06393-1
22. Gipp, Bela; Beel, Jöran (June 2010), "Citation Based Plagiarism Detection - A New Approach to Identifying Plagiarized Work Language Independently", Proceedings of the 21st ACM Conference on Hypertext and Hypermedia (HT'10) (PDF), ACM, pp. 273–274, doi:10.1145/1810617.1810671, ISBN 978-1-4503-0041-4
23. Gipp, Bela; Meuschke, Norman; Breitingner, Corinna; Lipinski, Mario; Nürnberger, Andreas (28 July 2013), "Demonstration of Citation Pattern Analysis for Plagiarism Detection", Proceedings of the 36th International ACM SIGIR Conference on

Research and Development in Information Retrieval (PDF), ACM,
doi:10.1145/2484028.2484214

24. Gipp, Bela; Meuschke, Norman (September 2011), "Citation Pattern Matching Algorithms for Citation-based Plagiarism Detection: Greedy Citation Tiling, Citation Chunking and Longest Common Citation Sequence", Proceedings of the 11th ACM Symposium on Document Engineering (DocEng2011) (PDF), ACM, pp. 249–258, doi:10.1145/2034691.2034741, ISBN 978-1-4503-0863-2
25. Gipp, Bela; Meuschke, Norman; Beel, Jöran (June 2011), "Comparative Evaluation of Text- and Citation-based Plagiarism Detection Approaches using GUTTENPLAG", Proceedings of 11th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'11) (PDF), ACM, pp. 255–258, doi:10.1145/1998076.1998124, ISBN 978-1-4503-0744-4
26. Gipp, Bela; Beel, Jöran (July 2009), "Citation Proximity Analysis (CPA) - A new approach for identifying related work based on Co-Citation Analysis", Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09) (PDF), International Society for Scientometrics and Informetrics, pp. 571–575, ISSN 2175-1935
27. Holmes, David I. (1998), "The Evolution of Stylometry in Humanities Scholarship", Literary and Linguistic Computing, 13 (3): 111–117, doi:10.1093/lilc/13.3.111
28. Juola, Patrick (2006), "Authorship Attribution" (PDF), Foundations and Trends Information Retrieval, 1: 233–334, doi:10.1561/15000000005, ISSN 1554-0669
29. Portal Plagiat - Softwaretest 2004 (in German), HTW University of Applied Sciences Berlin, retrieved 6 October 2011
30. Portal Plagiat - Softwaretest 2008 (in German), HTW University of Applied Sciences Berlin, retrieved 6 October 2011
31. Portal Plagiat - Softwaretest 2010 (in German), HTW University of Applied Sciences Berlin, retrieved 6 October 2011
32. Potthast, Martin; Barrón-Cedeño, Alberto; Eiselt, Andreas; Stein, Benno; Rosso, Paolo (2010), "Overview of the 2nd International Competition on Plagiarism Detection", Notebook Papers of CLEF 2010 LABs and Workshops, 22–23 September, Padua, Italy (PDF)
33. Potthast, Martin; Eiselt, Andreas; Barrón-Cedeño, Alberto; Stein, Benno; Rosso, Paolo (2011), "Overview of the 3rd International Competition on Plagiarism

- Detection", Notebook Papers of CLEF 2011 LABs and Workshops, 19–22
September, Amsterdam, Netherlands (PDF)
34. Stein, Benno; Lipka, Nedim; Prettenhofer, Peter (2011), "Intrinsic Plagiarism Analysis" (PDF), *Language Resources and Evaluation*, 45 (1): 63–82, doi:10.1007/s10579-010-9115-y, ISSN 1574-020X
35. Potthast, Martin; Barrón-Cedeño, Alberto; Stein, Benno; Rosso, Paolo (2011), "Cross-Language Plagiarism Detection" (PDF), *Language Resources and Evaluation*, 45 (1): 45–62, doi:10.1007/s10579-009-9114-z, ISSN 1574-020X
36. Weber-Wulff, Debora (June 2008), "On the Utility of Plagiarism Detection Software", In *Proceedings of the 3rd International Plagiarism Conference*, Newcastle Upon Tyne (PDF)
37. "Plagiarism Prevention and Detection - On-line Resources on Source Code Plagiarism" Archived 15 November 2012 at the Wayback Machine.. Higher Education Academy, University of Ulster.
38. Roy, Chanchal Kumar; Cordy, James R. (26 September 2007). "A Survey on Software Clone Detection Research". School of Computing, Queen's University, Canada.

Literatūra

- Carroll, J. (2002). *A handbook for deterring plagiarism in higher education*. Oxford: The Oxford Centre for Staff and Learning Development, Oxford Brookes University. (96 p.), ISBN 1873576560
- Zeidman, B. (2011). *The Software IP Detective's Handbook*. Prentice Hall. (480 p.), ISBN 0137035330