AI HELPT RECHERCHEUR ZOEKEN NAAR DIGITALE SPOREN

Sneller en dieper speuren met textmining en link analysis

AI WORDT AL MET SUCCES TOEGEPAST IN E-DISCOVERY ALS MACHINELEARNINGTECHNIEK TEN BEHOEVE VAN PREDICTIVE CODING OM SLIMMER E-MAILS EN DOCUMENTEN TE FILTEREN EN TE CLUSTEREN. AI WORDT OOK AL INGEZET DOOR OPSPORINGSINSTANTIES BIJ HET AUTOMATISCH ZOEKEN NAAR AFBEELDINGEN MET WAPENS, DRUGS EN BLOOT EN NAAR CHATBERICHTEN DIE BLIJK GEVEN VAN SEKSUELE INTENTIES. DIT ZIJN VOORAL AI-TOEPASSINGEN WAARBIJ DE COMPUTER DIGITALE INHOUD CLASSIFICEERT. HANS HENSELER SCHETST WAT WE NOG MEER KUNNEN VERWACHTEN VAN AI IN DE ZOEKTOCHT NAAR DIGITALE SPOREN.

door Hans Henseler beeld Shutterstock en Unsplash

ONZE SMARTPHONE EN HET EXPLOSIEF GROEIENDE INTERNET OF THINGS bevatten nog veel meer andere digitale sporen die een schat aan informatie bevatten voor forensisch onderzoek. Ook zijn deze sporen persoonlijker dan onze onderlinge communicatie, omdat ze niet alleen ons bewuste gedrag maar in toenemende mate ook ons onbewuste gedrag laten zien. De hoeveelheid informatie groeit navenant, en succesvolle zoekstrategieën gaan verder dan het lezen van e-mails, documenten en chats of het bekijken van foto's en video's.

LINK ANALYSIS

Het bedenken en toetsen van scenario's is van steeds groter belang in opsporingsonderzoeken. Voor een onderzoek kan het soms belangrijker zijn om te weten met wie is gecommuniceerd, waar iemand is geweest, wat die persoon heeft gedaan en wanneer, dan om te weten wat is gecommuniceerd. Juist hier zal AI ook kunnen helpen door onderzoekers te ondersteunen bij het leggen van verbanden, ook wel aangeduid als link analysis. Het analyseren van een sociaal netwerk aan de hand van e-mails wordt al langer toegepast, maar met AI kan link analysis op veel meer soorten data worden uitgevoerd.
Digital forensics op smartphones en computers kan inzicht geven in de verschillende telefoonnummers, user accounts op sociale media en e-mailadressen van een persoon. Het kan interessant zijn om te weten op welke apparaten deze identiteiten nog meer terug te vinden zijn en welke activiteit er te vinden is rondom een specifieke gebeurtenis. Is er gezocht op bepaalde zoektermen? Was de verdachte of het slachtoffer aan het wandelen? Wat was de locatie van de smartphone of auto? Wie waren er nog meer aanwezig op dat moment? Het zijn vragen die kunnen helpen bij het bedenken wat er gebeurd kan zijn (zie kader).

TEXTMINING

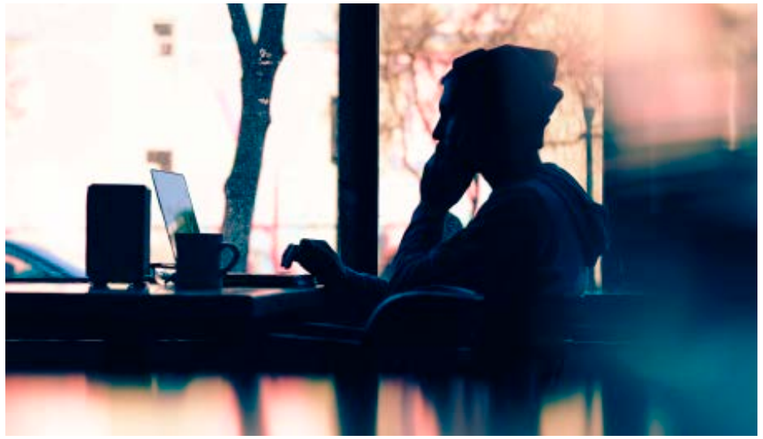
Meer dan 90% van de informatie waarover we kunnen beschikken, is ongestructureerd. Denk aan documenten, e-mail- en chatberichten, webpagina's enzovoorts. AI kan onderzoekers helpen bij het vinden van patronen in een wir-

PROBLEMEN MET HET KOPPELEN VAN DIGITAAL BEWIJS UIT VERSCHILLENDE TOOLS

CASE-STANDAARD

CASE is een opensourcestandaard waarin een ontologie is ontwikkeld die gebruikt kan worden voor het beschrijven van digitaal bewijs afkomstig uit verschillende domeinen zoals incident response, contraterrorisme, strafrechtelijk onderzoek, forensisch onderzoek en het verwerven van inlichtingen. Met CASE kunnen onderzoeken in verschillende jurisdicties beter gecoördineerd worden, zodat sneller dezelfde criminele personen en organisaties in beeld komen en er een volledige beeld ontstaat over de volledige omvang van hun criminele activiteiten. Het ontwikkelen van een standaard is een goed begin, maar kan pas echt succesvol worden indien die wordt overgenomen door commerciële bedrijven die digitaal forensische tools ontwikkelen. Daarom werd op 26 en 27 maart door Interpol in Den Haag een technische workshop georganiseerd, waarin de initiatiefnemers van CASE samen met marktleiders van digitaal forensische tools over de voor- en nadelen van CASE hebben gediscussieerd. In een panel is in detail gesproken over de meest voorkomende soorten digitale sporen en over het toekomstperspectief van CASE, waarbij de bereidheid van de industrie om CASE te omarmen een belangrijke rol zal spelen.

Meer dan 90% van de informatie waarover we kunnen beschikken, is ongestructureerd



war van digitale sporen. Met behulp van textmining kunnen verbanden ontdekt worden tussen entiteiten (personen, locaties, bedrijven enzovoorts) en gebeurtenissen waarover wordt gecommuniceerd in ongestructureerde data, zoals documenten, e-mail- en chatberichten. Textmining in ongestructureerde data zoals tekst is foutgevoelig, maar met voorkennis uit gestructureerde digitale sporen kan de kwaliteit van textmining flink verbeterd worden. Met behulp van een visualisatie van sporen in een netwerk, op een kaart of op een tijdlijn kan een forensisch onderzoeker veel sneller patronen in digitale sporen ontdekken en scenario's ontwikkelen of toetsen.
De analyse van gestructureerde informatie mag dan eenvoudiger lijken dan de analyse van ongestructureerde informatie, maar is zeker niet zonder problemen. Verschillende standaarden leiden tot problemen met het koppelen van digitaal bewijs dat afkomstig is vanuit verschillende tools, en met de uitwisseling van digitaal bewijs tussen verschillende organisaties en landen. Om dit groeiende probleem aan te pakken, is CASE ontwikkeld (zie kader). Internationale samenwerking tussen overheidsorganisaties, commerciële ontwikkelaars van digitale forensische tools en onderzoeks- en kennisinstituten moet ertoe bijdragen dat er overeenstemming is over standaarden voor het uitwisselen van digitaal bewijs.
AI en krachtige visualisaties helpen de rechercheur om met veel grotere nauwkeurigheid nieuwe patronen en verbanden te ontdekken, waarbij ook gezocht kan worden in ongestructureerd bewijsmateriaal, zoals chat en e-mail. Een tijdlijn met gebeurtenissen in smartphones en computers, en een geografische weergave van gps-gegevens helpen bij het identificeren van interessante gebeurtenissen die belangrijk zijn voor het construeren of toetsen van een scenario.
Minstens even belangrijk is dat AI zal kunnen helpen bij het aandragen van alternatieve scenario's die mogelijk door rechercheurs over het hoofd worden gezien, en ze daarmee kan behoeden tegen tunnelvisie.

DE 7 W-VRAGEN

In een opsporingsonderzoek is bijna altijd de centrale vraag wie in verband kan worden gebracht met het misdrijf. Naast de wie-vraag zijn er nog een aantal standaardvragen die helpen bij de opsporing van een strafbaar feit.
Tezamen worden ze ook wel de zeven w-vragen genoemd:
1. Wie kan in verband worden gebracht met het misdrijf?
2. Wat is er gebeurd?
3. Waar is het misdrijf gepleegd en waar kunnen mogelijk sporen gevonden worden?
4. Met welke middelen is het misdrijf gepleegd?
5. Op welke wijze is het misdrijf gepleegd?
6. Wanneer is het misdrijf gepleegd?
7. Waarom is het misdrijf gepleegd?

Digitaal bewijs kan uitstekend helpen bij het beantwoorden van de zeven w-vragen. De wie-vraag kan vaak beantwoord worden door te achterhalen welke gebruiker schuilgaat achter een e-mailadres, user account of telefoonnummer. Communicatie in sms, chat en e-mail geeft inzicht in wat er is gebeurd. Telecomgegevens, gps-locaties en wifinetwerken kunnen iets zeggen over de locatie waar een telefoon was. Foto's en video kunnen inzicht geven in de manier waarop en met welk middel een misdrijf is gepleegd. Datum en tijd van een bestand of spoor zeggen iets over wanneer iets is aangemaakt, gewijzigd of gezien. Computers en smartphones houden gedetailleerd bij wanneer apps en gebruiker actief zijn geweest en welke bestanden zijn geraadpleegd. Afgezien van bewuste uitingen van een verdachte in e-mails en chatberichten kan bijvoorbeeld de zoekgeschiedenis uit een browser of specifieke apps ook inzicht geven in motief en of voorbedachte rade terwijl de gebruiker zich daar niet van bewust was.

SYMPOSIUM E-DISCOVERY

Op donderdagmiddag 11 april werd het symposium E-Discovery georganiseerd bij Hogeschool Leiden. Deze 10de editie was gewijd aan het zoeken naar digitale sporen met behulp van AI. Zes verschillende experts zijn ingegaan op nieuwe toepassingen van AI in e-discovery en digitaal forensisch onderzoek. Onderwerpen die aan de orde kwamen, zijn: sentiment mining in e-mails om het onderbuikgevoel van een organisatie te meten; fraudeonderzoekers die gebruikmaken van datascience en slimme patroonherkenning die bedrijfse-mails herkent die mogelijk te maken hebben met fraude; en de ontwikkeling en inzet van textmining voor het zoeken in open bronnen om overheidsdiensten in Europe beter te helpen bij het uitvoeren van hun taken. Predictive coding wordt in de Nederlandse e-discoverypraktijk nog niet vaak toegepast. Toch zijn de voordelen duidelijk, en volgens experts is predictive coding volwassen en is het niet meer nodig zoekwoorden te bedenken en daarmee te filteren. Tenslotte is er stilgestaan bij de kansen en de risico's die artificial intelligence met zich meebrengt als oplossing voor digitaal forensisch onderzoek. Op de website van het symposium zijn de presentaties en videoregistraties van de sprekers te vinden: https://www.hsleiden.nl/digital-forensics/agenda/symposium/e-discovery-2019.
Van 24-26 april zullen tijdens de jaarlijkse Europese DFRWS (Digital Forensics Research Workshop-conferentie) bijna 200 onderzoekers uit meer dan 20 landen bijeenkomen om de nieuwste ontdekkingen op digitaal forensisch onderzoeksgebied uit te wisselen. Beide evenementen laten zien hoe snel de tools en technieken ontwikkelen waarmee gedetailleerde digitale sporen kunnen worden gevonden in de nieuwste smartphones, besturingssystemen en IoT-apparaten.
Op 17 juni zal tijdens de ICAIL 2019 (International Conference on AI and Law) een workshop worden gegeven over AI en Intelligent Assistance. Tot nu toe werd AI vooral gezien als een hulpmiddel bij Technology Assisted Review (TAR).
Deze workshop probeert te verkennen of AI ook een stap verder kan gaan dan de ondersteuning bij review. Als de samenwerking tussen mens en AI nog verder verbeterd kan worden, kunnen we dan wellicht spreken over Technology Assisted Discovery?

REACTIES EN BIJDRAGEN
Voor reacties en nieuwe bijdragen van IT-experts:
Tanja de Vrede
020-2356415
t.d.vrede@agconnect.nl

AUTEUR

HANS HENSELER is lector Digital Forensics & E-Discovery bij de specialisatie Forensische ICT aan Hogeschool Leiden. Hij is tevens director Digital Evidence Review bij Magnet Forensics uit Canada, dat software ontwikkelt voor opsporingsinstanties (henseler.h@hsleiden.nl).