DOI: 10.38146/BSZ.2022.4.7

Szövegbányászat a dark neten: rendészettudományi alkalmazások[1]

Text mining on the darknet: use cases in law enforcement studies

Szigeti Ákos
doktorandusz
Nemzeti Közszolgálati Egyetem,
Rendészettudományi Doktori Iskola
szigeti.akos@uni-nke.hu

Absztrakt

Cél: A kibertér felhasználói tömérdek elemezhető szöveges adatot hoznak létre, ahogy a látható (surface) weben, úgy a láthatatlan (deep) weben, és azon belül az anonimitásra épülő dark net platformjain is. A szövegbányászat különböző elemzési eljárásai lehetőséget kínálnak e nagy adatmennyiség (big data) automatizált elemzésére, amit számos kutató kiaknáz. Tanulmányom célja a rendészettudomány szempontjából releváns jó gyakorlatok, alkalmazási példák áttekintése, bemutatása.

Módszertan: A szövegbányászat társadalomtudományban való elterjedésének újszerűsége miatt kutatásom során a kurrens szakirodalom feldolgozására specializálódott, úgynevezett state-of-the-art szakirodalomelemzést alkalmaztam, melynek célja az adott kutatási terület új perspektíváinak bemutatása.

Megállapítások: A nemzetközi szakirodalomban megjelennek például a legális-illegális tartalmak klasszifikációját célzó nyelvmodellek, melyek megerősítik a dark net kettős felhasználhatóságáról szóló elméletet. Az illegális kereskedelmi tevékenységet (is) végző dark netes kriptomarketek élete jellemzően a rendvédelmi szervek beavatkozásával, bezárással végződik, ahogy történt az a Silk Road nevű kriptomarkettel is 2013-ban. A bezárásokat követő felhasználói aktivitás trendjeit elemző topikmodellezési eljárások segítséget nyújthatnak e rendészeti beavatkozások értékelésében.

1 Az Innovációs és Technológiai Minisztérium Kooperatív Doktori Program Doktori Hallgatói Ösztöndíj Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

Érték: A tanulmány e példák bemutatásával a szövegbányászat mint kutatási módszer rendészettudományi kiaknázásában rejlő lehetőségeket világítja meg.

Kulcsszavak: kutatásmódszertan, szövegbányászat, darknet, kiberbűnözés, rendészettudomány

Abstract
Aim: The users of cyberspace create an enormous amount of textual data on the surface web and on the deep web as well, including the anonymity based darknet platforms. The various automated analytical methods of text mining allow us to analyse these big data sources, which opportunity is already exploited by several researchers. The aim of my study was to review the use cases which are relevant from the aspect law enforcement studies.
Methodology: As text mining is relatively new in social sciences, I applied state-of-the-art review methodology, which is specialized in reviewing current literature to offer new perspectives in a field of research.
Findings: In the international literature, we can find examples for classifying legal and illegal content by statistical language models, strengthening the theory of darknet's dual-usage. Darknet markets usually end up being closed by law enforcement agencies, just like it happened in the case of the Silk Road market, back in 2013. Analysing the trends of user activity after the closure of specific darknet markets can help in evaluating the interventions of law enforcement.
Value: By presenting these examples, this study shed light on the exploitable opportunities of text mining as a research method in law enforcement studies.

Keywords: research methodology, text mining, darknet, cybercrime, law enforcement studies

Bevezető

A digitális tér szolgáltatói és platformjai gyűjtik és elemzik a látogatottsági mu-tatókat, a kattintások számát vagy éppen a tranzakciók alakulását. E könnyen számszerűsíthető adatok mellett a kibertér felhasználói hatalmas mennyiségű szöveges adatot is létrehoznak: a közösségi média oldalain, fórumokon, blogo-kon és egyéb platformokon osztják meg gondolataikat hozzászólások és posztok formájában. A nagy adatmennyiség algoritmusok segítségével történő feldol-gozásával foglalkozó adattudomány úgynevezett természetesnyelv-feldolgozás

(angolul natural language processing, NLP) nevű területe éppen az ilyen strukturálatlan szöveges adatok kezelésére szakosodott. A szövegbányászat természetesnyelv-feldolgozási algoritmust alkalmazva alakítja át a strukturálatlan adathalmazt elemezhető szövegkorpusszá, vagyis a szöveg nagyobb, strukturált halmazává, amit utána a számos eljárás közül választva elemezhetünk (Németh, Katona & Kmetty, 2020). Ilyen például a pozitív-negatív tartalmak elkülönítésére képes szentiment-analízis (vagy más néven véleménybányászat), a gépi tanuláson alapuló és az elemzett dokumentumokban jelenlévő témák azonosítását célzó topikmodellezés, ahogy ide tartoznak a különböző, szemantikus hálók elemzésén (a szavak számának, együttes előfordulásának és távolságának mérésén) alapuló, úgynevezett szóbeágyazási modellek is (Veltri, 2020). Bár e módszerek fejlesztésében korunk technológiai vállalatai éllovasnak tekinthetők[2], az akadémiai szféra szerepe sem elhanyagolható[3], a szövegbányászat alkalmazása pedig mára a nagy adatmennyiség feldolgozására szakosodott, úgynevezett számítógépes társadalomtudomány (angolul computational social science) területének meghatározó elemévé vált. A szövegbányászat társadalomtudományi alkalmazásáról már több magyar nyelvű tanulmány is született (Novák, Siklósi & Prószéky, 2018; Németh, Katona & Kmetty, 2020; Katona & Németh, 2021), ugyanakkor a hazai rendészettudomány területén máig nem jelent meg a módszer. Tanulmányomban amellett érvelek, hogy ahogy számos más társadalomtudományi kutatási kérdés megválaszolásában, úgy a bűnözésre adott intézményi és társadalmi reakciók (és azok hatásosságának) vizsgálatában is segítséget nyújthatnak a különböző automatizált szöveganalitikai eljárások. Érvelésem alátámasztása érdekében a dark neten és a dark netről szóló fórumokon megvalósított, rendészettudományi szempontból releváns, szövegbányászatot alkalmazó kutatások eredményeit tekintem át; ahol lehetséges, ott érvényességüket más kutatási módszerekkel gyűjtött eredményeken keresztül vizsgálva. Amellett, hogy a dark net maga is egy új kutatási terület, melyről a társadalomtudomány még viszonylag kevés ismerettel rendelkezik, vizsgálatom aktualitását tovább erősíti, hogy a dark netes kriptomarketeken zajló kábítószer-kereskedelem (mely a dark net-kutatások jelentős részének kutatási tárgyát képezi) a COVID−19 világjárvány okozta kijárási korlátozások következtében még inkább felerősödött (EMCDDA & Europol, 2020).

2 Úttörő munkát végeztek például a szentiment-analízis fejlesztése terén az IBM (Nasukawa & Yi, 2003), vagy a szóbeágyazási modellek létrehozása során a Google munkatársai (Mikolov, Sutskever, Chen, Corrado & Dean, 2013).

3 A GloVe nevű szóbeágyazási modellt például a Stanford Egyetem kutatói fejlesztették ki (Pennington, Socher & Manning, 2014).

A kettős felhasználhatóságú dark net

A dark netet (vagy más néven dark webet[4]), ahogy a neve is jelzi, a globális internet sötét oldalaként tartjuk számon. A dark nethez tartozó oldalakat a hagyományos keresőmotorok (mint például a Google vagy a Bing!) nem tudják elérni, így azok a deep web, vagyis az internet nem indexált részét képezik. Azonban míg a deep web többi tartalmát jellemzően egy szolgáltatóhoz (például e-mail kliensbe vagy vállalati intranetbe) történő egyszerű bejelentkezés után elérhetjük, a dark net tartalmainak eléréséhez egy anonim jelenlétet lehetővé tevő böngészőre is szükségünk van. Ilyen például a TOR (The Onion Router), melynek fejlesztését eredetileg az Amerikai Tengerészet kezdeményezte. Céljuk az internet – és az azon keresztül zajló kommunikációjuk – lehető legmagasabb szintű titkosítása volt, amit az internetes forgalom különböző szervereken való átvezetésével értek el, úgy, hogy az információt lépésenként titkosították. A TOR Project weboldalán kvázi misszióként fogalmazzák meg, hogy „*az internet felhasználóinak privát hozzáférésük kell legyen a cenzúrázatlan internethez*" (a szerző saját fordítása, URL1).

Van ugyanis a dark netnek egy „kevésbé sötét" oldala is: például egyes nem demokratikus országokban politikai aktivisták és újságírók is használják a platformot, hírek, információk és véleményük cenzúramentes megosztására (UNODC, 2021). Bár a dark net valós biztonsági kockázatot jelent, az, illetve annak megfogható része, a platformjai és az elérését biztosító böngésző valójában csupán egy technológiai eszköz, melyet sokan, sokféle célból használnak (Mirea, Wang & Jung, 2019). Jardine (2015) szerint a dark net rendészeti eszközökkel való kezelése nyilvánvalóan nem a legideálisabb megoldás, sokkal jobb lenne, ha az emberek egyszerűen nem használnák a TOR böngészőt illegális dolgokra, hiszen így nem fenyegetné a dark net oldalakat az illegalitás miatti betiltás. Ezzel megmaradna a lehetőség az elnyomó országokra jellemző cenzúra és megfigyelés megkerülésére, melyet tehát a TOR böngésző által nyújtott online anonimitás biztosít. Ez az idealisztikus elképzelés aligha lehet korunk valósága, így Jardine is inkább amellett érvel, hogy az anonimitás és a TOR böngésző korlátozása helyett a fókusznak inkább a hálózaton történő konkrét események rendészeti kezelésén kellene lennie, ezzel minimalizálva a dark net okozta társadalmi károkat, és egyúttal maximalizálva a platformnak köszönhető pozitív hatásokat (Jardine, 2015). A fent hivatkozott szerzők érvelését egy triviális hasonlattal szemléltethetjük: csak azért, mert az utcán illegális kábítószer-kereskedelem is zajlik, még nem célszerű az utcai közlekedést betiltani.

4 A dark web kifejezést jellemzően a dark net szinonimájaként használja a szakirodalom (Jardine, 2015).

A dark net eléréshez szükséges technológia tehát önmagában még nem „sötét” (sőt, valójában még csak nem is szürke, hiszen a TOR böngésző használata teljesen legális), ugyanakkor tudjuk, hogy a dark net illegális tartalmak (például gyermekpornográfia) feltöltésére, illegális szolgáltatások (így akár bérgyilkosság) hirdetésére, illetve az úgynevezett kriptomarketeken illegális áruk (például tiltott gyógyszerek, kábítószerek, fegyverek stb.) kereskedelmére is lehetőséget ad. E kriptomarketek működése kifejezetten hasonlít más online kereskedelemre szakosodott platformok működéséhez (például az Amazonéhoz vagy az eBay-éhez), a kulcsfontosságú különbség az, hogy a kriptomarketek elérése teljes mértékben anonim, a hatóságok által nem visszakereshető módon történik (EMCDDA & Europol, 2020).

A dark netes kábítószer-kereskedelem problematikája napjainkban talán aktuálisabb, mint korábban: a European Monitoring Centre for Drugs and Drug Addiction (EMCDDA, a Kábítószer és a Kábítószer-függőség Európai Megfigyelőközpontja) és a European Union Agency for Law Enforcement Cooperation (Europol, a Bűnüldözési Együttműködés Európai Uniós Ügynöksége) 2020. májusi jelentése szerint a járványügyi helyzet miatti kijárási korlátozások eredményeképpen nőtt a dark net illegális kábítószer-kereskedelemben betöltött szerepe. Bár a személyes találkozásokra építő kábítószer-kereskedők a kijárási korlátozások fennállásának idején kevesebbet vásároltak a kriptomarketeken, az egyéni fogyasztók rendeléseinek száma jelentősen nőtt (EMCDDA & Europol, 2020). Érdemes megjegyezni, hogy a kripotomarketeken zajló kábítószer-kereskedelemnek lehet pozitív hatása is. Mivel az itt zajló tranzakciók máig jellemzően kereskedő és fogyasztó (és nem kereskedő és kereskedő) között zajlanak (UNODC, 2021), a kriptomarketek forgalmának növekedése hatással lehet a kábítószer-kereskedelemhez kapcsolódó erőszakos bűnözés alakulására is: a kriptomarket egy közvetlen csatorna lehet a termesztők/előállítók és a fogyasztók között, ezzel szűkítve az utcai kereskedelem terét, és megkímélve a fogyasztókat a veszélyes helyek felkeresésétől (Martin, 2014; UNODC, 2021). Megjegyzendő ugyanakkor, hogy a személyes találkozást a veszélyessége miatt nem kockáztató potenciális vásárlók számára éppen a kriptomarketek adnak lehetőséget a kábítószerek vásárlására, ezzel bővítve a potenciális vásárlók és így a kábítószer-fogyasztók körét, egyúttal növelve a napjainkban tapasztalható kábítószer-krízist (Pergolizzi, LeQuang, Taylor, Raffa & NEMA Research Group, 2017). E hatást gyengítheti, hogy az online kábítószer-vásárlás esetén a vevőnek valamilyen címet meg kell adnia, ami magában hordozza a vevő zsarolhatóságát és a rendészeti szereplők általi beazonosíthatóságát (EMCDDA & Europol, 2020).

A kriptomarketeket célzó rendészeti beavatkozások tervezése előtt tehát sok szempontot érdemes figyelembe venni, melyek súlyozása nem könnyű feladat.

Mindenesetre úgy tűnik, hogy a dark net egyes platformjain (kriptomarketeken, fórumokon, hírfolyamokon) megjelenő tartalmakat célzottan érdemes értékelni és kezelni, amihez a platformok minél szélesebb körű és minél mélyebb megismerésére van szükség. Tekintve, hogy e platformokon jellemzően nagy mennyiségű szöveges adat jön létre (gondoljunk például a fórumok hozzászólásaira vagy a kriptomarketek eladóira és az ő termékeikre adott visszajelzésekre, értékelésekre), a rendészettudomány hatékony eszközei lehetnek a különböző szövegbányászati megoldások.

Szövegbányászat alkalmazása a dark neten

A szövegbányászat dark neten való alkalmazására a nemzetközi szakirodalomban számos példát találhatunk. A kurrens szakirodalom feldolgozására specializálódott úgynevezett state-of-the-art szakirodalomelemzésem (Grant & Booth, 2009) során e példákból emelem ki a rendészettudomány szempontjából releváns alkalmazási eseteket. Avarikioti és szerzőtársai (2018) például a dark net minél több oldalának megismerése céljából, webcrawler (keresőrobot) segítségével gyűjtötték össze több, mint 34 ezer dark netes weboldal tartalmát, melyek közül végül több, mint 7500 oldalon találtak elemezhető adatot. Elemzésükből megtudhatjuk például, hogy ezen oldalak körülbelül kétharmada angol nyelvű, 11%-uk orosz, és további 24 nyelvet detektáltak az előbbieknél alacsonyabb arányban. A felügyelt tanulás alapú (tanulóhalmazzal rendelkező) úgynevezett Support Vector Machine (tartóvektor-gép, SVM) osztályozó algoritmus segítségével végzett tartalmi klasszifikációjuk alapján a kábítószerre utaló tartalmak aránya például csupán 4,3%, míg a legnagyobb arányban a blog (13,5%) és a szoftver (12,3%) kategóriába estek a tartalmak. A korábban említett kettős felhasználhatóság elméletét erősíti, hogy eredményeik szerint az oldalak 6,2%-án szerepeltek aktivizmushoz kapcsolódó tartalmak. Elemzésük szerint összeségében a tartalmak közel kétharmada (62,6%) volt legális. Ugyanakkor amennyiben az adott szolgáltatóhoz tartozó oldalakat, például egy kriptomarket esetén az összes aloldalt (az egyes áruk oldalait) összevonva kezeljük, akkor elmondhatjuk, hogy a szolgáltatók több, mint fele (56,4%) illegális tartalmat vagy szolgáltatást (is) nyújt (Avarikioti, Brunner, Kiayias, Wattenhofer & Zindros, 2018). Choshen és szerzőtársai (2019) szintén az illegalitás klasszifikációját tűzték ki célul: olyan nyelvmodellt hoztak létre, mely a dark neten fellelhető szövegek nyelvezete alapján képes eldönteni, hogy azok legális vagy illegális tartalmakra utalnak. Tanulmányukkal nem az Avarikioti és munkatársai (2018) által végzett feltáró kutatást ismételték meg, céljuk sokkal inkább a módszertan fejlesztése volt. Eredményeik alapján

nem csupán a különböző szavak előfordulása, de a szövegek szintaktikai jellemzői alapján is lehetséges a legális és az illegális szövegek elkülönítése, így a dark net platformjai – véleményük szerint – nagyszerű tesztkörnyezetnek bizonyulnak az ilyen osztályozást célul kitűző nyelvmodellek fejlesztéséhez (Choshen, Eldad, Hershcovich, Sulem & Abend, 2019). Ezek az eredmények túlmutatnak a dark net kutatásán, hiszen a legális és az illegális tartalmak ilyen módon létrehozott klasszifikációs eljárásai vélhetően más környezetben is alkalmazhatók, hasznosíthatók lehetnének. Tavabi és munkatársai (2019) a deep web és a dark web (melyeket összefoglalóan d2web-nek neveztek el) 80 fórumának 482 ezer hozzászólásán végeztek gépi tanuláson (machine learning) alapuló, úgynevezett Latent Dirichlet Allocation (látens Dirichlet allokáció, LDA) topikmodellezést a dokumentumokban jelen lévő rejtett mintázatok feltárása érdekében. Az LDA a generatív modellek csoportjába tartozó valószínűségi modell, ami egy szövegkorpusz dokumentumait rögzített számú téma összességeként reprezentálja, a témákat a korpusz szavainak eloszlása alapján azonosítva (Blei, 2003). Ezt a modellt Tavabi és szerzőtársai (2019) a fórumok aktivitásának dinamikáját vizsgálandó, úgynevezett Beta Process HMM (BP-HMM) modellezéssel egészítették ki. Ebben a megközelítésben az egyes fórumok többváltozós idősorként jelennek meg, ahol a változók az LDA által talált témák. Ezek a változók kerülnek be a BP-HMM modellbe, ami megvizsgálja a fórumokon megjelenő témák közötti eltéréseket, és segít nyomon követni a különböző fórumokon folytatott beszélgetéseket, s azonosítani az esetleges rendellenes viselkedést vagy fontos eseményeket. Az így kapott eredményekből kiemelendő például az AlphaBay és a Hansa kriptomarketek Federal Bureau of Investigation (FBI, Szövetségi Nyomozó Iroda) általi leállításának nyomon követése. Mivel a két kriptomarketet éppen a kutatás időablakában zárták be (a később részletesebben bemutatandó Bayonet Művelet keretében), a kutatók az adatokból következtetni tudtak arra, hogy a két oldal felhasználói a Dream Market nevű kriptomarketre költöztek át (Tavabi et al., 2019).

Bár nem szövegbányászati módszer segítségével, de szintén nagy adatmennyiség feldolgozásával jutottak hasonló eredményre ElBahrawy és szerzőtársai (2020). A szerzők közel 39 millió felhasználó több, mint 130 millió BitCoin kriptovaluta-tranzakcióját vizsgálták, és közülük több, mint 8 millió felhasználó esetében találtak kripotmarkettel zajló interakciót. E tranzakciók vizsgálatán keresztül megállapították, hogy egy kriptomarket bezárása után a felhasználók szinte kivétel nélkül egy másik kriptomarketre migrálnak át (ElBahrawy et al., 2020). Mindez kísértetiesen hasonlít a bűnözés rendőrségi beavatkozást követő térbeli áthelyeződésének offline térben már régóta vizsgált jelenségére (Guerette & Bowers, 2009). Az ilyen és ehhez hasonló kutatások eredményei

lehetőséget adnak arra, hogy felmérjük egy adott rendészeti beavatkozás hozadékaként előálló előnyöket és hátrányokat, és végső soron értékeljük a beavatkozás sikerességét (Riloff, Wiebe & Phillips, 2005).

A dark net marketekkel kapcsolatban már a surface weben (az internet indexált, bárki által elérhető felületén) is végeztek szöveganalitika elemzést. Porter (2018) munkájában a Reddit közösségi oldal DarkNetMarkets nevű (azóta betiltott), úgynevezett subredditjében (alfórumában) végzett LDA topikmodellezést, szintén az Alphabay és a Hansa bezárását magába foglaló időablakban. Eredményeiből kiderül, hogy a hozzászólók korábbi laza, nyugodt hangvétele a bezárásokat követően egy aggódó, bizonytalan és biztonságorientált hangvételre váltott, és nőtt az igény az olyan titkosított kommunikációt lehetővé tevő csatornák használatára, mint amilyen a PGP (Pretty Good Privacy) (Porter, 2018). Ezek az eredmények arra utalnak, hogy a beavatkozás jelentős hatással volt a kriptomarketek felhasználóira is – kérdés azonban, hogy a növekvő bizonytalanság valójában hány felhasználót térített el a vásárlástól. Bradley és Stringhini (2019) nem szövegbányászat segítségével, hanem kvalitatív kutatásban, a DarkNetMarkets és a dnmuk nevű Reddit alfórumok tartalomelemzésével vizsgálták két beavatkozás hatásosságának eltéréseit. Az egyik vizsgálatba bevont akció a Hyperion Művelet (Operation Hyperion) volt, melyet különböző rendészeti szervek Five Eyes Law Enforcement Group (FVEY, Öt Szem Rendészeti Csoport[5]) nevű nemzetközi együttműködése valósított meg 2016 novemberében, és keretében elsősorban a svéd rendőrség hallgatott ki gyanúsítottakat: eladókat és vevőket egyaránt (EMCDDA & Europol, 2020). Majd a holland rendőrség létrehozott egy dark netes weboldalt, melynek – vélhetően elrettentő célzattal – a *„Dark net market felhasználók nyomozás alatt”* („Darknet market users under investigation”) nevet adták (Bradley & Stringhini, 2019). A korábban már említett Bayonet Művelet (Operation Bayonet) nevű beavatkozás alig egy évvel később, 2017 júliusában történt: az FBI bezárta az egyik legnagyobb dark net marketet, az Alphabay-t, majd ezt követően a holland rendőrség azonosította a Hansa market szervereit. Az Alphabay felhasználói jelentős részének Hansára való átmigrálását, és így adataiknak hatóság általi begyűjtését követően, egy hónappal később bezárták a Hansát is (EMCDDA & Europol, 2020). Bradley és Stringhini (2019) kutatási eredményei szerint a Bayonet Művelet hatása sokkal jelentősebb volt, mint a Hyperioné. Főleg a Hansa bezárása előtti adatgyűjtés keltett aggodalmat a felhasználókban, és az, hogy ekkor több felhasználó a pénzét is elveszítette. Ez utóbbi beavatkozást követően többen utaltak arra, hogy szünetet tartanak vagy teljesen befejezik a kriptomarketes kereskedelmi tevékenységüket (Bradley & Stringhini, 2019).

5 A szerző saját fordítása.

Konklúzió

Tanulmányomban olyan dark neten megvalósuló, szövegbányászatot alkalmazó kutatásokat mutattam be, melyek a rendészettudomány számára is releváns kutatási eredményeket hoztak. Bár a különböző automatizált szöveganalitikai eljárások relatíve új módszernek tekinthetők a társadalomkutatás tágabb területén is, és a dark net is egy kifejezetten új kutatási terepként jelent meg az elmúlt évtizedben, a dark net szövegbányászat segítségével való vizsgálatára már így is remek példákat találhatunk a nemzetközi szakirodalomban. Az adattudomány területéről kölcsönzött szövegbányászat módszere a dark net feltérképezése mellett lehetőséget ad olyan specifikus kutatási feladatok elvégzésére, mint amilyen például a dark net illegális tartalmainak detektálása. Az ilyen, legális-illegális tartalmak klasszifikációjára képes nyelvmodellek fejlesztése pedig azon túl, hogy megerősíti a dark net kettős felhasználhatóságára vonatkozó elméletet, más környezetben való hasznosíthatóságának köszönhetően a dark net kutatásán túlmutató jelentőséggel is bírhat. További példa a szövegbányászat dark neten való alkalmazására a kriptomarketekkel kapcsolatos rendészeti beavatkozások hatásának vizsgálata. Az e témára fókuszáló, szövegbányászatra épülő kutatások megállapítását, mely szerint egy kriptomarket hatóság általi leállítása után a felhasználók jellemzően másik kriptomarketen folytatják az illegális tevékenységüket, a kriptovaluta-tranzakciók elemzése is megerősíti. Ugyanakkor szövegbányászaton alapuló és kvalitatív tartalomelemzést végző kutatás is képes bemutatni a felhasználók hozzáállásának bezárások utáni változását: a korábbi nyugodt hangulatot aggodalom váltja fel, és – főleg az adatvesztéssel és anyagi kárral járó Hansát érintő beavatkozás esetében – a bezárás egyeseket el is rettenthet további tevékenységüktől. A hosszútávú hatások vizsgálata későbbi kutatások feladata lehet – hiszen a dark netes kriptomarketek kábítószer-kereskedelemben betöltött szerepe nemhogy csökkenni nem látszik, az utóbbi időben inkább növekedést lehetett elkönyvelni, amihez napjainkban a digitalizáció járványügyi korlátozások miatti felgyorsulása is hozzájárult. A fenti példák azt hivatottak illusztrálni, hogy a szövegbányászat a rendészettudomány számára is értékes kutatási eredményeket képes hozni, így érdemes lehet megfontolni a kutatási folyamatokba való beillesztését, más, „tradicionális" társadalomkutatási módszerekkel (például kvalitatív tartalomelemzéssel, interjúkészítéssel stb.) való együttes alkalmazását. Az így kapott kutatási eredmények segítséget nyújthatnak a rendészeti beavatkozások tervezésében, a dark netre vonatkozó kriminálpolitikai irányvonalak kialakításában.

Felhasznált irodalom

Avarikioti, G., Brunner, R., Kiayias, A., Wattenhofer, R. & Zindros, D. (2018). Structure and Content of the Visible Darknet. *Computers and Society, 4*, 1811.01348.

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(1), 993–1022.

Bradley, C. & Stringhini, G. (2019). A Qualitative Evaluation of Two Different Law Enforcement Approaches on Dark Net Markets. *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 453–463. https://doi.org/10.1109/EuroSPW.2019.00057

Choshen, L., Eldad, D., Hershcovich, D., Sulem, E. & Abend, O. (2019). The Language of Legal and Illegal Activity on the Darknet. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4271–4279. https://doi.org/10.18653/v1/P19-1419

ElBahrawy, A., Alessandretti, L., Rusnac, L., Goldsmith, D., Teytelboym, A. & Baronchelli, A. (2020). Collective dynamics of dark web marketplaces. *Scientific Reports*, *10*(1), 18827. https://doi.org/10.1038/s41598-020-74416-y

EMCDDA & Europol (European Monitoring Centre for Drugs and Drug Addiction and Europol). (2020). *EU Drug Markets: Impact of COVID–19*. Publications Office of the European Union, Luxembourg. https://www.emcdda.europa.eu/system/files/publications/6585/TD0417834ENN.pdf

Grant, M. J. & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, *26*, 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Guerette, R. T. & Bowers, K. J. (2009). Assessing the extent of crime displacement and diffusion of benefits: a review of situational crime prevention evaluations. *Criminology*, *47*(4), 1331–1368. https://doi.org/10.1111/j.1745-9125.2009.00177.x

Jardine, E. (2015). The Dark Web Dilemma: Tor, Anonymity and Online Policing. *Global Commission on Internet Governance Paper Series*, *21*. https://doi.org/10.2139/ssrn.2667711

Katona E. & Németh R. (2021). Automatizált szöveganalitika a korrupció kutatásában. *Socio. Hu, 11*(1), 108–124. https://doi.org/10.18030/socio.hu.2021.1.108

Nasukawa, T. & Yi, J. (2003). Sentiment analysis. *Proceedings of the International Conference on Knowledge Capture – K-CAP '03. the international conference.* https://doi.org/10.1145/945645.945658

Németh R., Katona E. & Kmetty Z. (2020). Az automatizált szövegelemzés perspektívája a társadalomtudományokban. *Szociológiai Szemle, 30*(1), 44–62. https://szociologia.hu/dynamic/44_62_oldal.pdf

Novák, A., Siklósi, B. & Prószéky, G. (2018). Segíthetnek-e a szóbeágyazási modellek a társadalomtudósoknak? *Magyar Tudomány, 179*(7), 945–954. https://doi.org/10.1556/2065.179.2018.7.3

Martin, J. (2014). *Drugs on the dark net: How cryptomarkets are transforming the global trade in illicit drugs.* Palgrave Pivot. https://doi.org/10.1057/9781137399052

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z. & Weinberger, K. Q. (Eds.), *Proceedings of the Conference on Advances in Neural Information Processing Systems 26 (NIPS)* (pp. 3136–3144). La Jolla: Neural Information Processing Systems Foundation. https://arxiv.org/pdf/1310.4546.pdf

Mirea, M., Wang, V. & Jung, J. (2019). The not so dark side of the darknet: A qualitative study. *Security Journal*, *32*(2), 102–118. https://doi.org/10.1057/s41284-018-0150-5

Pennington, J., Socher, R. & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP).* https://doi.org/10.3115/v1/d14-1162

Pergolizzi, J. V., LeQuang, J. A., Taylor, R., Raffa, R. B. & NEMA Research Group. (2017). The "Darknet": The new street for street drugs. *Journal of Clinical Pharmacy and Therapeutics*, *42*(6), 790–792. https://doi.org/10.1111/jcpt.12628

Porter, K. (2018). Analyzing the DarkNetMarkets subreddit for evolutions of tools and trends using LDA topic modeling. *Digital Investigation*, *26*, S87–S97. https://doi.org/10.1016/j.diin.2018.04.023

Riloff, E., Wiebe, J. & Phillips, W. (2005). *Exploiting subjectivity classification to improve information extraction*. AAAI-05/1106–1111. https://www.aaai.org/Papers/AAAI/2005/AAAI05-175.pdf

Tavabi, N., Bartley, N., Abeliuk, A., Soni, S., Ferrara, E. & Lerman, K. (2019). Characterizing Activity on the Deep and Dark Web. *Companion Proceedings of The 2019 World Wide Web Conference* (pp. 206–213). IW3C2. https://doi.org/10.1145/3308560.3316502

UNODC (United Nations Office on Drugs and Crime, az Egyesült Nemzetek Szövetségének Kábítószer-ellenőrzési és Bűnmegelőzési Hivatala) (2021). *World Drug Report 2020. 4: Cross Cutting Issues: Evolving Trends and New Challanges.* United Nations. https://wdr.unodc.org/wdr2020/field/WDR20_BOOKLET_4.pdf

Veltri, G. (2020). *Digital social research*. Polity Press.

A cikkben található online hivatkozás

URL1: *TOR Project – History*. https://www.torproject.org/about/history/

A cikk APA szabály szerinti hivatkozása

Szigeti Á. (2022). Szövegbányászat a dark neten: rendészettudományi alkalmazások. *Belügyi Szemle*, *70*(4), 757–767. https://doi.org/10.38146/BSZ.2022.4.7