

A BIG DATA ELEMZÉSEK PERSPEKTÍVÁI A TÁRSADALOMTUDOMÁNYI KUTATÁSOKBAN

DR. HABIL. ALBERT-LŐRINCZ CSANÁD¹

Absztrakt

A tanulmány a Big Data elemzések alkalmazásának perspektíváit vizsgálja a társadalomtudományok területén. A felhasználó akaratlanul is az online tevékenysége során olyan adatokat generál, amelyek alkalmasak a valós érdeklődési köreit, értékrendjét, preferenciáit nemcsak leírni, hanem ezen túlmenően a várható emberi viselkedés is bizonyos mértékig előrelátható. Az adatelemzés ezen módszere a közeli évtizedekben bizonyítottan versenyelőnyhöz juttatott egyes szereplőket különböző szektorokban, mint például az informatikában, kereskedelemben és a pénzügyekben. Azonban a társadalomtudományi kutatásokban csak mérsékelten terjedt el, annak ellenére, hogy a módszerben rejlő lehetőségek szinte korlátlanok, eddig nem kutatott és nem is dokumentált dimenziókhoz férhetünk hozzá. Az adatelemzéssel az eredmények gyorsan tesztelhetővé, pontosabbá és megbízhatóbbá tehetőek, de a hagyományosnak számító kvantitatív és kvalitatív módszereket nem helyettesíti teljes mértékben.

Kulcsszavak: az adatelemzés perspektívái, a humántudományok számszerűsítése, kulcsszóelemzés, kutatómódszertan, viselkedés-kutatás

Bevezetés

A társadalomtudományi kutatások során az egyéneket bevonó hagyományos kvantitatív adatgyűjtés több problémát is felvet, valamint bizonyos korlátok közé szorítja az eredmények értelmezhetőségét. Hagományos adatgyűjtés alatt azt értjük, hogy a kutató igyekszik kialakítani egy reprezentatív mintát, majd lekérdezi az alanyokat, vagy meglévő adatbázisokkal másodlagos adatelemzést végez. Mindkét eljárás lényege ugyanaz, a válaszok az egyének által felvállalt többé-kevésbé szubjektív véleményét fogják tükrözni. Az egyik felmerülő probléma, hogy a mintavételi eljárás valóban reprezentatív-e, a másik, hogy a felvállalt válaszok mennyire őszinték. Eddigi kutatómunkánk során többször szembesültünk olyan visszajelzésekkel, hogy a kutatásba bevont alanyaink az anonimitás garantálása mellett sem vállalták azon véleményüket, ami például eltér a társadalmi elvárásoktól vagy kritikával illeti a munkaadóját. Tehát a kvantitatív adatelemzés legfőbb korlátja, hogy mennyire megbízhatóak az adatok.

Az egyének vélekedése és viselkedése között diszkrépancia tapasztalható, sőt önmagukkal sem őszinték, amint Stephenz-Davidowitz 2017-ben megjelent (2019-ben magyar fordításban is), a *Mindenki hazudik* című könyvében kifejti. A szerző rámutat arra, hogy online környezetben, főként az internetes keresései során, az egyén akaratlanul is felfedi a valós érdeklődési köreit, értékrendjét,

¹ egyetemi docens, Partiumi Keresztény Egyetem, Humántudományok Tanszék, alcs@partium.ro

preferenciáit és különféle dilemmáit. Sokat elárulhatnak valakiről a keresési szokásai, a kulcsszavak, amiket használ, az egyes honlapokon eltöltött idő, az online rendelései, nem is beszélve a szociális médiában folytatott tevékenységéről. A felhasználói szokások a célzott marketinget hatékonyabbá teszik és a felhasználó kereskedelmi profilozásához is alkalmazhatóak. Mivel értékesíthető adatokról van szó, ezeket az online szolgáltatók eltárolják és a Big Data egy kis szeletét fogják alkotni. Mellesleg ebben a helyzetben az Európai Adatvédelmi Rendelet hatástalan (Zarsky 2017; Tesfay et al 2018; McDermott 2017), de erre még a későbbiekben visszatérünk.

Adott tehát a kérdés, hogy a kereskedelmi profil megalkotására alkalmas adatok milyen egyéb területeken hasznosíthatóak és hozzá tudnak-e járulni az egyéni és társadalmi viselkedés kutatásához, megértéséhez. A kérdés megválaszolásához nézzük meg a Big Data fogalomkörét. Bár jelenleg még nincs általánosan elfogadott meghatározása, a fogalom az 1990-es évek elején jelent meg az informatikai forradalommal összefüggésben. A lényege abban rejlik, hogy „eddig nem látott nagyságrendű adatok tárolását, feldolgozását, a rejtett és váratlan összefüggések megtalálását feltételezi” és segítségével „olyan következtetéseket lehet levonni, melyek kisebb mennyiségű adat feldolgozása során nem tűnnének fel” (Szűcs & Yoo 2016 9-10 o.). A szakirodalom egyetért a Big Data legfőbb jellemzőit illetően, amelyet az 5V felosztásként ismerünk: volume, velocity, variety, veracity és value – mennyiség, sebesség, változatosság, valódiság és érték (Naqshbandi 2021).

Kétségtelen, hogy Big Data új perspektívákat nyitott számos kutatási terület számára. Ezen adatok a társadalmi, politikai és gazdasági tendenciák kiértékelésére is alkalmasak (McNeely & Hahm 2014).

Milyen ágazatok profitálhatnak a Big Data-ból és mit forradalmasíthat?

A Big Data lehetőségeinek kiaknázása többek között hozzájárult a Google sikeréhez², ami nem meglepő, hiszen a cég egyben úttörője is az adattechnológiának. Azok a szektorok, ahol rendelkezésre állt a megfelelő pénz és szaktudás, a Big Data kezdetei óta profitálnak az adatok nyújtotta lehetőségekből: a piackutatás, a marketing, a pénzügy, a befektetési alapok stb., vagyis egyszerűen a versenyszféra terén.

Choi és Hal a Google kereső kulcsszó lekérdezések kapcsán 2009-ben publikáltak egy rövid tanulmányt, melyben bizonyították, hogy egy adott iparágban, a lekérdezések korrelálnak a gazdasági aktivitás jelenlegi szintjével, és így a későbbi adatközlések előrejelzésében segítségünkre lehetnek. 2011-ben Hal, aki a Google vezető közgazdásza, egy újabb tanulmányban azt vizsgálta a kulcsszó lekérések kapcsán, hogy hogyan kaphatnak betekintést az emberek érdeklődési körébe, szándékaiba és jövőbeli cselekedeteibe – ami már túlmutat a pusztán gazdasági hasznosíthatóságon. Megjegyeznénk, hogy a felhasználók által használt kulcsszavakkal csak egy parányi részét lehet elérni a Big Data-nak, bár itt, a 2004-es évvel kezdődően, témérdek információhoz lehet hozzáférni. 2012-ben Lohr arról számolt be, hogy a Big Data korában élünk és ez, a következő évtizedben a társadalom minden területét forradalmasítani fogja, minden tudományág egyre inkább adatintenzívebb lesz. Ovadia már 2013-ban felhívja a figyelmet arra a szerepre, amelyet a Big Data a társadalomtudományokban tölthet be. Hesse, Moser és Riley (2015) a Big Data tudássá való átalakításáról értekezett. Stephenz-

² A statista.com szerint 2022 júliusában majdnem 84%-os a keresőmotorok terén a világszeresedése.

Davidowitz (2019), a Google volt adattudósa és a New York Times rovatvezetője még ennél is tovább ment, azt példázva, hogy amit az emberekről eddig gondoltunk, az nagyrészt téves, majd könyvében a Big Data segítségével végzett számos kutatásáról számol be.

Az elmúlt években egyre több publikáció jelent meg a Big Data különböző területeken történő hasznosításáról, beleértve a gazdaságtant, az orvostudomány, közegészségügyet, szociológiát, pszichológiát, tehát az alcímben megfogalmazott kérdésre egyértelműen az a válasz, hogy mindenki, pontosabban minden tudományág profitálhat a Big Data nyújtotta lehetőségekből.

Az alcím második tézise, hogy pontosan mit forradalmasít a Big Data-n alapuló adatelemzés, nem teljesen egyértelmű. A versenyszférában ismert, rendszeresen alkalmazott adatelemzési technikákról van szó, évtizedes tapasztalattal rendelkező elemzők dolgoznak a tőkeerős, innovatív vállalatoknak. Az általuk alkalmazott elemzési módszertan és az algoritmusaik viszont üzleti titoknak minősülnek. Tehát az adatok forradalma jelenleg a versenyhelyzet szülte folyamatos innovációt jelent.

A társadalomtudományok területén azonban más a helyzet, bár léteznek Big Data elemzésen alapuló kutatások, ezek száma viszonylag alacsony a hagyományosnak tekintett módszerekkel végzett kutatásokkal szemben. Ennek valószínűsíthető okai, hogy ezen területek finanszírozottsága sokkal alacsonyabb a versenyszférához képest, a humántudományi szakemberek matematikai és informatikai kompetenciái számukra többnyire terület-idegenek, és ezekhez hozzáadódik a kidolgozott és elfogadott módszertani sablonok hiánya.

A társadalomtudományok szerintünk lépéshátrányban vannak a versenyszférához képest az innováció tekintetében. Bár a humántudományok számszerűsítése már a 19. század közepétől növekvő jelentőséggel bír (Runkel 1962; Desrosières 1996), ez a folyamat a személyi számítógépek, internet és elérhető, felhasználóbarát statisztikai programok elterjedésével tovább fokozódott. Annak ellenére, hogy a neves szaklapok és a doktori iskolák fokozottan elvárják az eredmények kvantifikációját, mégsem számítunk adatforradalomra a humántudományok területén. Az intenzív finanszírozás és a több egymástól távolálló területen is képzett szakemberek viszonylagos hiánya mellett további érvként felhoznánk a Big Data elemzések korlátait, melyekből kifolyólag nem váltják ki a hagyományos, bevált módszereket. Az adatelemzések vitathatatlan előnye, hogy új perspektívát nyitnak és lehetővé teszik a hipotézisek gyors tesztelését és olyan kutatásokat, amikre eddig nem volt mód. Arra számítunk, hogy a Big Data lehetővé teszi az eredmények megbízhatóságának növelését, pontosabbá tételét.

Hogyan történik az adatelemzés?

A Big Data elemzés, angolul „analytics” vagy rövidítve „BDA” egy olyan komplex folyamat, mely magába foglalja az információk feltárását, mint például rejtett minták, összefüggések, piaci trendek és felhasználói preferenciák. Ezen információk kinyerése alkalmazásokon keresztül történik, melyek prediktív modelleket, statisztikai algoritmusokat és az analitikai rendszerek által működtetett „mi lenne, ha” elemzéseket feltételeznek. Az adatokat ún. adattárházakban (data warehouse) vagy adattengerekben (data lake) gyűjtik és tárolják. Ezt követi az adatok letisztítása különböző szkriptkészítő eszközökkel vagy adatminőség-javító szoftverek segítségével: kiszűrjük az esetleges hibákat vagy inkonzisztenciákat, mint például az ismétlődések, formázási hibák, és végül rendszerezik az adatokat. Az utolsó lépés az

adatok szoftveres elemzése adatbányászattal, prediktív elemzéssel, adatmegjelenítéssel, machine learninggel stb. (Chai et al s.a.). A fenti folyamat a kutató vagy adatelemző részéről matematikai, statisztikai és informatikai ismereteket feltételez, tehát az egyéni kutató helyett a több szakemberből álló kutatói csoportok hatékonyabbak.

Az adatelemzés egyszerű formái

Létezik egy olyan szűk szelete a Big Data-nak, amit minimális informatikai és statisztikai ismeretekkel az egyéni kutató is hasznosítani tud. Ez a módszer nem más, mint a Google kulcsszó lekérdezések elemzése, egy olyan módszer, amit elsősorban a weboldal keresőoptimalizálásával (search engine optimization - SEO) foglalkozó szakemberek is alkalmaznak. Nielsen (2013) kissé sarkított meglátása szerint a Google kulcsszó lekérdezője nem más, mint az emberi szándékok adatbázisa, amely egy sokkal nagyobb vízió kis része: egy adatbázis, amely a világ összes tudását tartalmazza. A szerző szerint kétségtelen, hogy ezen adatok nagyban hozzájárulhatnak a világ megértéséhez.

A módszer lényege, hogy a Google Trends felületen (trends.google.com) megjeleníthető, hogy egy tetszőleges kulcsszóra, vagy kifejezésre milyen gyakran kerestek rá. A felület a keresések relatív számát adja meg egy normalizált értéktartományon belül, vagyis a teljes keresési mennyiségéhez viszonyítva egy adott pillanatban milyen gyakran kerestek rá az adott kulcsszóra. A beállítások között kiválasztható egy adott régió, ország, település, a keresési periódus (2004-től az elmúlt óráig), kategória és internetes kereső. Lehetőség van két vagy több kulcsszó összehasonlítására. Az így kapott ingyenes információ kiindulási alapnak tekinthető, de még nem elegendő a kutatáshoz. Például a kiugró keresési csúcsok összevethetőek a közelmúlt eseményeivel, aminek már konkrét információ értéke lehet a kutatás szempontjából, de a ritkábban lehívott kulcsszavakra kapott relatív értékek nem értelmezhetőek. Az adatok kiegészítéseként szükségünk lesz az abszolút keresési mennyiségekre is, azonban ezeket a Google Trends nem jeleníti meg. Az abszolút keresési mennyiségekhez többféleképpen is hozzáférhetünk, fizetése ellenében használhatjuk a Google Ads felületét (<https://ads.google.com/>), vagy „third party” alkalmazások és kiterjesztések közül is válogathatunk. Stephenz-Davidowitz (2019) a már korábban hivatkozott könyvében számos példán keresztül szemlélteti, hogy az emberek szándékai, preferenciái és valószínűsíthető viselkedései hogyan mutathatóak ki ezzel a módszerrel, illetve a kulcsszó alapú kutatások során milyen korlátokra lehet számítani.

Ki rendelkezhet a Big Data-val és mit kell betartaniuk a kutatóknak?

Az „adattulajdonlás” kifejezés az elmúlt néhány évben felkeltette a kutatók figyelmét. A kérdés azért releváns, mert meghatározza, hogy ki és milyen mélységig férhet hozzá, vizsgálhatja, kutathatja, vagyis milyen mértékben nyilvánosak az adatok. A tulajdonlást két szempontból vizsgáljuk: az egyén szempontjából és vállalati szempontból. A két szempont közti kapocs abban rejlik, hogy a Big Data egyének, felhasználók millióinak online tevékenységéből, viselkedési mintázataiból és okos eszközeik által képzett adataiból épül fel, amelyeket nevezhetünk online attribútumoknak. Amint megosztjuk a különböző online szolgáltatókkal, vállalatokkal, eltárolják ezen adatokat és saját céljaik megfelelően hasznosítják.

Egyéni szempontból, Al-Khouri (2012) szerint a saját adatok a szó szoros értelmében csak a személyes online tevékenységhez kapcsolódó attribútumokat tartalmazzák. Ezek olyan információk, amelyek a megosztás pillanatáig sajátunk tekinthetők, viszont amikor kifejezett szándékkal vagy implicit módon megosztjuk őket, a tulajdonjogot is átruházzuk a szolgáltatóra. Minden egyes megosztással az adatok tulajdonosának a száma nő és új adatok generálódnak, új tulajdonjogokat létrehozva. Például amikor egy bejelentkezett Google fiók felhasználó a weben keresést indít, felhasználói keresési minták generálódnak. Noha a keresési információk nem tartoznak a Google-hoz, a keresési mintákban gyűjtött adatok már igen. Ez egy olyan probléma, amely a hivatkozott szerző szerint túlmutat a kereskedelem, az etika és az erkölcs határain, és adatvédelmi problémákhoz és a magánélet védelméhez vezet. A jogalkotónak elkerülhetetlenül újra kell terveznie az adatvédelmi törvényeket nemzeti és nemzetközi szinten (Al-Khouri 2012). A Big Data tehát személyes adatokat is tartalmaz, ezek nem megfelelő kezelése sérelmet okozhatnak, ezért a kutatók felelőssége az adatok legális és etikus felhasználása (Wiltshire & Alvanides 2022).

A normatív beavatkozás szükségességére Al-Khouri már 2012-ben felhívta a figyelmet, válaszként az Európai Unió 2016-ban kidolgozta az új adatvédelmi rendeletét (Európai Parlament és Tanács 2016), mely 2018 május 25-én lépett hatályba. A személyes adatokkal kapcsolatos egyéni kontroll megerősítése volt az uniós jogalkotó egyik fő célja (van Ooijen 2019), viszont a gyakorlatban az online szolgáltatók és vállalatok a tényleges adatvédelem helyett inkább áttértek „*agree or quit*” (Houser & Voss 2018 104 o.) modellre, mi szerint a felhasználók vagy elfogadják – megfelelő tájékoztatás mellett – a saját üzleti és adatvédelmi feltételeiket, vagy lemondanak az online szolgáltatás vagy platform használatáról. Megjegyeznénk, hogy a lemondásra nincs alternatíva, mivel jellemzően ezen online vállalatok monopolhelyzetben vannak, így az adatgyűjtés és értékesítés továbbra sem áll le, csak a felhasználók „rákényszerült” beleegyezésével történik. A jelenlegi adatvédelmem nem más, mint egy kattintással elfogadott megállapodás, ahol a felhasználónak nincs érdemi mozgáster, „*take-it-or-leave-it*” alapon működik (Compagnucci 2020).

Vállalati szempontból a kiindulási pont, hogy kik gyűjtik és tárolják az adatokat, mivel mindkét folyamat költségigényes. Innen következik, hogy egy hatékony adatinfrastruktúra csak néhány nagy profitorientált vállalatnál áll rendelkezésre, ún. domináns technológiai platformoknak, mint például a Google és a Facebook (Nielsen 2013).

A kutathatóság szempontjából az volna az ideális, ha az adathalmaz a nyilvánosság számára is elérhető lenne. Nielsen (2013) amellett érvel, hogy előnyös lenne, ha az adathalmazt vagy infrastruktúrát non-profit szervezetek vagy esetleg magánszemélyek „laza” hálózata működtetnék, mint a felfedezés és innováció platformját. Azonban egy ilyen infrastruktúra egyetlen darabja jelentős tőkeköltséggel járna, amely egy nagy számítógépcsoport üzemeltetését feltételezi. A valóság az, hogy a fent említett domináns technológiai platformok magántulajdonban vannak, a platform tulajdonosa pedig megválaszthatja, hogy kivel és milyen mélységig osztja meg az adatokat, hogy versenyelőnyhöz jusson az emberi tudás feltérképezése által. A fent hivatkozott szerző eszmefuttatásában kifejezi azon reményét, hogy a jövőben a legjobb adatinfrastruktúra bárki számára elérhető lesz a világon egy olyan hatékony platformként, mely lehetőséget biztosít a kísérletezéshez és felfedezéshez.

Bár Nielsen megállapításait közel egy évtizede fogalmazta meg, ami az információs technológiák világában hosszú idő, véleményünk szerint ma is érvényesek. Azonban az adatok tulajdonlásának kérdése ma sem tekinthető lezártnak, például Fadler és Legner (2020) szerint újra kell értelmeznünk az eddigi koncepciókat, hogy megfeleltessük a Big Data elemzések (BDA) során felmerülő kihívásoknak. Összehasonlító tanulmányuk során három adattulajdonlási típust azonosítanak, amelyek a gyakorlatban is előfordulnak: adat, adatplatform és adattermék. Szerintük a jogok mellett a felelőségek is tisztázásra szorulnak.

A fenti felosztásra épített analógia szerint, ha Big Data elemzést végez egy kutató, maga az adat és a felhasznált adatplatform nyilván nem a sajátja. Viszont, ha rendelkezik hozzáférési jogosultsággal, az elemzést követő adattermék, rendszerint tanulmány már a saját szellemi tulajdonát fogja képezni. Az adatok forrását és platformot azonban meg kell neveznie, az esetleges személyes adatokat el kell fednie és az általános kutatói etikát szintén be kell tartania.

Konklúziók

A nagy áttörés, amelyet az adatforradalom jelentett a Big Tech cégek és a pénzügyi világ számára, csak lassan szivárog át a nem intenzíven finanszírozott területekre, mint például a társadalomtudományok. Jelenleg viszonylag alacsony ezen a területen a Big Data-t is felhasználó kutatások száma, a módszertana szintén kidolgozatlan. Ebből kifolyólag nagy hangsúlyt kap az egyéni kreativitás és intuíció. Azon kutatóknak, akik ebbe a típusú elemzésbe belevágnak, több területen is átfogó ismeretekkel kell rendelkezniük, valamint hozzáférést kell szerezniük az ingyenesen és/vagy fizetés ellenében elérhető információkhoz.

A módszerben rejlő lehetőségek szinte korlátlanok, eddig nem kutatott (és nem is dokumentált) dimenziókhöz férhetünk hozzá. Az adatelemzéssel az eredmények gyorsan tesztelhetővé, pontosabbá és megbízhatóbbá tehetőek. A hagyományosnak tekintett módszerek, a kisebb/nagyobb mintás felmérések és a minőségi elemzések továbbra is fontosak, az általánosan elvárt kutatói etika továbbra is érvényes.

Hivatkozások

Al-Khouri A. M. (2012): Data Ownership: Who Owns 'My Data'? *International Journal of Management & Information Technology* 2(1), 1-8. o. DOI: 10.24297/ijmit.v2i1.1406

Chai W., Labbe M., Stedman C. (s.a.): Big data analytics. *Techtarget*. <https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics>

Choi H., Hal V. (2009): *Predicting the Present with Google Trends*. Google Inc. https://static.googleusercontent.com/media/www.google.com/en//googleblogs/pdfs/google_predicting_the_present.pdf

Compagnucci M. C. (2020): *Big Data, Databases and "Ownership" Rights in the Cloud*. Springer Singapore, DOI:10.1007/978-981-15-0349-8

Desrosières A. (2016): *The quantification of the social sciences: an historical comparison*. Springer Cham. DOI:10.1007/978-3-319-44000-2_15

Európai Parlament és Tanács. (2016): Az Európai Parlament és a Tanács (EU) 2016/679 számú Rendelete a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről (általános adatvédelmi rendelet). *Az Európai Unió Hivatalos Közlönye* 59, pp. 1-88. <https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:32016R0679&from=EN>

Fadler M., Legner C. (2020): *Who owns data in the enterprise? Rethinking data ownership in times of Big Data and analytics*. Twenty-Eighth European Conference on Information Systems (ECIS2020). https://serval.unil.ch/resource/serval:BIB_E6BE23FE6102.P001/REF

Hal V. (2011): *Predicting the Present*. Google Inc., <https://www.thinkwithgoogle.com/marketing-strategies/search/predicting-the-present>

Hesse B. W., Moser R.P., Riley W. T. (2015): From Big Data to Knowledge in the Social Sciences. *The ANNALS of the American Academy of Political and Social Science* 659(1), 16-32. o. DOI:[10.1177/0002716215570007](https://doi.org/10.1177/0002716215570007)

Lohr S. (2012): The age of Big Data. *The New York Times*. <https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>

McDermott Y. (2017): Conceptualising the right to data protection in an era of Big Data. *Big Data & Society*, 4(1), 1-7. o. <https://doi.org/10.1177/2053951716686994>

McNeely C. L., Jong-on H. (2014): The Big (Data) Bang: Policy, Prospects, and Challenges. *Review of Policy Research*, 31(4), 304-310. o.

Naqshbandi G. (2021): "Big Data: The 3 V's of Data." *Wevolver*, 2021:9, www.wevolver.com/article/big-data-the-3-v-s-of-data

Nielsen M. (2013): Who Owns Big Data? In *Ch@nge: 19 Key Essays on How the Internet Is Changing Our Lives*. Madrid: BBVA. <https://www.bbvaopenmind.com/en/articles/who-owns-big-data>

Houser K. A., Voss, W. G. (2018): GDPR: The End of Google and Facebook or a New Paradigm in Data Privacy? *Richmond Journal of Law & Technology*, 25(1), 109. o.

Ovadia Steven (2013): The Role of Big Data in the Social Sciences, *Behavioral & Social Sciences Librarian*, 32(2), 130-134. o. DOI: [10.1080/01639269.2013.787274](https://doi.org/10.1080/01639269.2013.787274)

Runkel P. J. (1962): Quantification in the social sciences. *The Mathematics Teacher*, 55(1). 20-33. o. <http://www.jstor.org/stable/27956514>

Statista.com. (2022): Worldwide desktop market share of leading search engines from January 2010 to July 2022 <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/#:~:text=Google%20in%20the%20global%20market&text=Google%20has%20dominated%20the%20search,share%20as%20of%20June%202021>

- Stephens-Davidowitz S. (2019): *Mindenki hazudik*. Budapest: Athenaeum Kiadó
- Szűts Z. & Yoo J. (2016): Big Data, az információs társadalom új paradigmája. *Információs Társadalom*. 16(1), 8-28. o. DOI: 10.22503/inftars.XVI.2016.1.1
- Tesfay W. B., Hofmann P., Nakamura T. és mtsai. (2018): PrivacyGuide: Towards an Implementation of the EU GDPR on Internet Privacy Policy Evaluation. In: *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics (IWSPA '18)*. Association for Computing Machinery, New York. 15–21. o. DOI: 10.1145/3180445.3180447
- van Ooijen I., Vrabec H. U. (2019): Does the GDPR Enhance Consumers' Control over Personal Data? An Analysis from a Behavioural Perspective. *Journal of Consumer Policy*, 2019(42), 91–107. o. DOI:10.1007/s10603-018-9399-7
- Wiltshire D., Albanides S. (2022): Ensuring the ethical use of big data: lessons from secure data access. *Heliyon*, 8(2), 1-6. o. DOI:10.1016/j.heliyon.2022.e08981.
- Zarsky Tal Z. (2017): Incompatible: The GDPR in the Age of Big Data. *The Seton Hall Law Review*, 47(2), 995-1020. o.

THE PERSPECTIVES OF BIG DATA ANALYSIS IN SOCIAL SCIENCE RESEARCHES

DR. CSANÁD ALBERT-LÓRINCZ

In this study I examine the perspectives of Big Data analysis in the field of social sciences. The data generated by internet users not only reveal their interests, personal values and preferences, but they also can be used for predicting the expected human behavior to a certain extent. In recent decades this method of data analysis has been proven to give some companies a competitive advantage in various sectors, such as IT, commerce and finance. However, the method is still not widely used in social sciences, even though it has almost unlimited possibilities since it can open up dimensions that have not been researched and documented until now. With data analysis, the results can be quickly tested, accuracy and reliability can be increased, but at the same time it cannot completely replace the traditional quantitative and qualitative methods.

Keywords: perspectives of data analysis, quantification of social sciences, keyword analysis, research methodology, behavioral research