

<https://doi.org/10.17048/AM.2023.161>

<https://videotorium.hu/hu/recordings/51305>

Csernai Zoltán: A Big Data adatelemző módszerek alkalmazásának lehetőségei az oktatásban

Csernai Zoltán

Eszterházy Károly Katolikus Egyetem, Digitális Technológia Intézet, Humáninformatika Tanszék
csernai.zoltan@uni-eszterhazy.hu

Absztrakt: A jelenleg is zajló 4. ipari forradalomban a világhálón elérhető digitális megoldások globális elterjedésének hatására ezek az információs rendszerek óriási, nagy mennyiségű adattömeget, úgynevezett Big Data-t generálnak és tárolnak, amelyek feldolgozása és hasznos információvá való alakítása kihívást és egyben lehetőséget jelent a társadalom valamennyire alrendszere számára.

Big Data alatt olyan rendkívül nagy adathalmazokat értünk, amelyek az emberi viselkedéssel és interakciókkal kapcsolatos mintákat, trendeket és összefüggéseket tárnak fel (Yeoman, 2019). Kutatási célom, hogy feltérképezzem Big Data adatelemző módszereket és az ezzel kapcsolatos korábbi kutatásokat, valamint megvizsgáljam, hogyan történik ezeknek a módszereknek az alkalmazása az oktatás területén. A kutatás során a PRISMA-protokoll és a számítógépes tartalomelemzés (induktív kódolás) módszerét alkalmazom.

A kutatásom következő fázisában arra keresem a választ, hogy az online tanulást támogató portálok kurzusainak az adatai milyen módon kerülnek felhasználásra és elemzésre, valamint az oktatói körben milyen lehetséges további igények merülnek fel az online oktatás során.

Kulcsszavak: Big Data, PRISMA-protokoll, számítógépes tartalomelemzés

"A Kulturális és Innovációs Minisztérium ÚNKP-23-3 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült."

1. Bevezetés

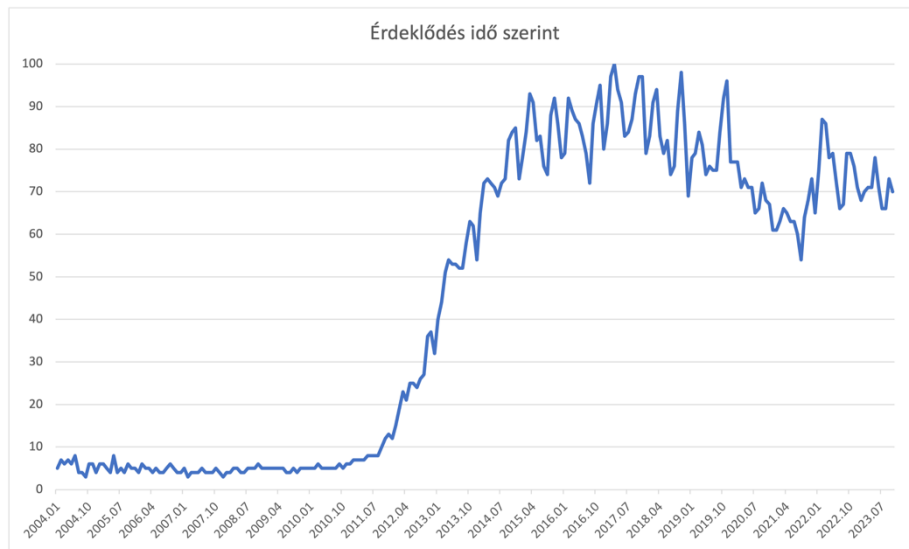
Az elmúlt tíz évben a közösségi médián, a streaming tartalmakon, az online vásárláson és egyéb tevékenységeken keresztül történő digitális elkötelezettség nagymértékben növekedett. Miközben a világot a COVID-19 világjárvány, gazdasági hullámvölgyek sújtották, a társadalomban egy dolog állandó volt, mégpedig az új digitális eszközök növekvő használata személyes és üzleti szükségleteink támogatására.

A "Data Never Sleeps" 10. éves infografikája⁴ azt mutatja, hogy mennyi adat keletkezik percenként az interneten. Megállapítható, hogy az adatok mennyisége és sokfélesége folyamatosan gyorsul, és a lassulásnak nincsen semmi jele.

A Google Trends⁵ mérőalkalmazás segítségével azt figyelhetjük meg, hogy a világ Big Data iránti érdeklődése hogyan alakult 2004-től napjainkig (1. ábra).

⁴ <https://web-assets.domo.com/miyagi/images/product/product-feature-22-data-never-sleeps-10.png>

⁵ <https://trends.google.com/trends/>



*1. ábra A Big Data iránti érdeklődés növekedése az idők során
A kép forrása: Saját forrás*

A "Big Data" kifejezés keresése az interneten 2011 után kezdett emelkedni, azonban a 2021-es évben egy csökkenés figyelhető meg, aminek valószínűsíthető oka a COVID-19 világjárvány hatása a világra.

Az Y tengelyen található számok a keresési érdeklődést jelzik a grafikon legmagasabb pontjához képest az adott régióban és időszakban. A 100-as érték a kifejezés legnagyobb népszerűsége; az 50-es érték azt jelzi, hogy a kifejezés feleannyira népszerű. A 0-as érték pedig azt jelenti, hogy nem áll rendelkezésre elég adat a kifejezéshez.

A tanulmány először azt vizsgálja, hogy a "Big Data" kifejezésnek milyen definíciós kísérletei találhatók meg a szakirodalmakban. Ezt követően a Big Data szisztematikus irodalmi áttekintésére kerül sor a PRISMA-protokoll és a számítógépes tartalom elemzés (induktív kódolás) módszerét alkalmazva. A tanulmány következő részében a Big Data feldolgozásának folyamata és a legnépszerűbb adatelemzési technológiák kerülnek áttekintésre. A tanulmány végén az adatvezérelt intelligens osztályterem egy gyakorlati példa arra, hogy a Big Data hogyan alkalmazható az oktatás területén.

1.1A Big Data definíciós kísérletei

A Big Data kifejezés szinte mindenütt jelen van és nem csak kizárólag az információs technológia (IT) területén, hanem más tudományágban is elterjedt, mint például a szociológia, az orvostudomány, a biológia, a közgazdaságtan és a menedzsment. Manapság a "Big Data" kifejezésnek nincsen hivatalos definíciója, az emberek többféle és ellentmondásos jelentéssel használják.

A továbbiakban a Big Data leggyakoribb meghatározásai kerülnek ismertetésre.

„Olyan adathalmazok, amelyek mérete meghaladja a tipikus adatbázis-szoftverek rögzítési, tárolási, kezelési és elemzési képességeit.” (Manyika és mtsai., 2011)

„A Big Data négy jellemzője a mennyiség, a sebesség, a változatosság és az érték.” (Dijcks, 2013)

„A Big Data fogalma magában foglalja a nagy és összetett adathalmazok tárolását és elemzését, újszerű technikák alkalmazásával.” (Ward & Barker, 2013)

„A Big Data egy olyan információs eszköz, amelyet a mennyiség, a sebesség és a változatosság jellemez, valamint különleges technológiák és elemzési módszerek szükségesek az értéké alakításhoz.” (De Mauro és mtsai., 2016)

„A Big Data együttese magába foglalja a korábban soha nem látott mértékű és változatos forrásból érkező adatok rögzítését, feldolgozását, elemzését, megosztását, illetve az eredmények vizualizálását.” (Szűts & Yoo, 2016)

„Big Data alatt olyan rendkívül nagy adathalmazokat értünk, amelyek az emberi viselkedéssel és interakciókkal kapcsolatos mintákat, trendeket és összefüggéseket tárják fel.” (Yeoman, 2019)

Úgy gondolom, hogy célszerű lenne egy konszenzusos meghatározást létrehozni, bár ez még jelenleg nem tűnik egy egyszerű feladatnak.

1.2 A Big Data szisztematikus irodalmi áttekintése (PRISMA-protokoll) és számítógépes tartalomelemzése (MAXQDA)

A PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) módszertani ajánlás az információkeresés számszerű nyomon követését teszi lehetővé, melyet az alábbi folyamatábrával lehet ábrázolni (2. ábra).



2. ábra Szisztematikus irodalmi áttekintés (PRISMA-protokoll)

A kép forrása: Kamarási V., & Mogyorósy G. (2015). Szisztematikus irodalmi áttekintések módszertana és jelentősége. Segítség a diagnosztikus és terápiás döntésekhez. Orvosi Hetilap, 156(38), 1523–1531. <https://doi.org/10.1556/650.2015.30255>

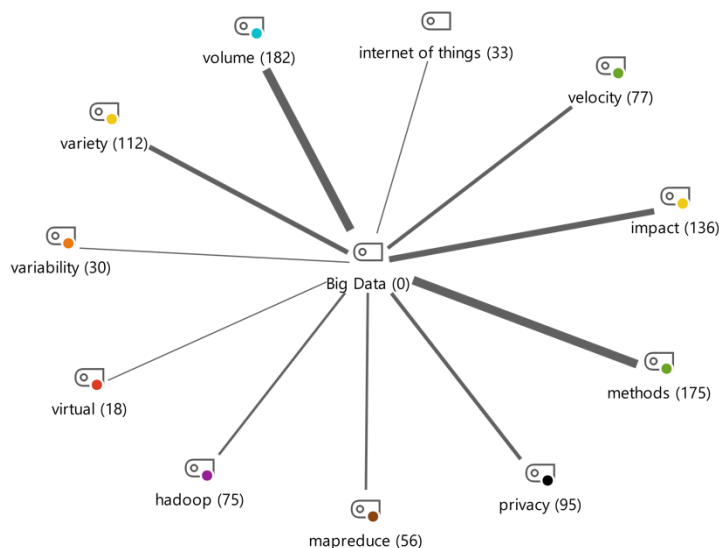
Az azonosításkor az elektronikus adatbázisokban, illetve az egyéb forrásokban (weboldalak, hivatkozások) kulcsszavak alapján (lásd. "Big Data", "Big Data" AND "education") történik a keresés. Mivel az egyes közlemények több elektronikus adatbázisban is szerepelhetnek, ezért fontos a duplikációk eltávolítása.

A szűrés során a beválogatott és a kizárt közleményeket (pl. ha nem áll rendelkezésre a teljes szöveg) határozzuk meg.

A megfelelelőség során azok a tanulmányok kerülnek kizárásra, melyek nem megfelelők a kutatás szempontjából. Fontos, hogy a kizárás ténye szöveges formában is kerüljön indoklásra.

A beválasztás az utolsó lépés, amikor a beválogatott tanulmányok végső számát határozzuk meg.

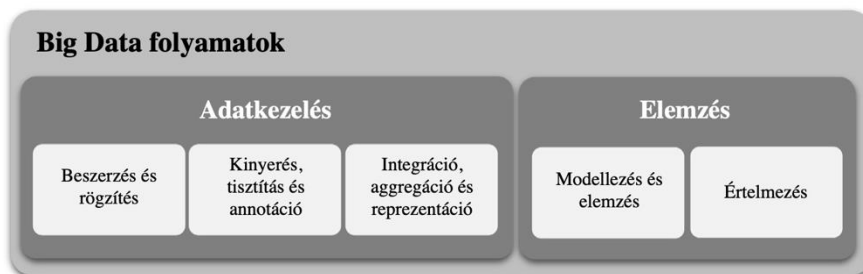
A szisztematikus irodalmi áttekintés után a MAXQDA számítógépes tartalomelemző szoftverben a Maxmaps funkció segítségével egy vizuális megjelenítés készült (3. ábra).



3. ábra Big Data vizuális megjelenítése (MAXQDA Maxmaps)
 A kép forrása: Saját forrás

2. A Big Data feldolgozásának folyamata

A Big Data feldolgozásának folyamata öt szakaszra bontható. Ez az öt szakasz alkotja a két fő alfolyamatot: az adatkezelést és az elemzést. (4. ábra).



4. ábra A Big Data feldolgozásának folyamata

A kép angol nyelvű forrása: Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144. Library, Information Science & Technology Abstracts. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

Az adatkezelés magában foglalja az adatok megszerzéséhez és tárolásához, valamint az adatok elemzésre való előkészítéséhez és visszakereséséhez szükséges folyamatokat és támogató technológiákat.

Az elemzés a nagy mennyiségű adat analitikájára és a belőlük származó információk megszerzésére használt technikákra utal.

A továbbiakban röviden áttekintésre kerülnek a legnépszerűbb adatelemzési technikák.

2.1. Szövegelemzés

A szövegelemzés olyan technikákra utal, melyek szöveges adatokból (pl. e-mailek, blogok, online fórumok) nyerik ki az információkat. A szövegelemzés lehetővé teszi a nagy mennyiségű, ember által generált szöveg értelmes összefoglalókká alakítását, így elősegítve a cégek számára a megfelelő döntéshozatalt.

Az információ kinyerési (IE) technikák strukturálatlan szövegből nyernek ki strukturált adatokat.

A szövegösszefoglalási technikák automatikusan egy vagy több dokumentumból készítenek összefoglalót. Az így kapott összefoglaló az eredeti szöveg(ek) legfontosabb információit közvetíti.

A kérdésválaszoló (QA) technikák a természetes nyelven feltett kérdésekre adnak választ. A QA rendszerekre jó példa az Apple Siri és az IBM Watson. Az Apple Siri egy tudásalapú megközelítést kihasználó rendszer, ezzel szemben az IBM Watson a kérdést szemantikailag elemzi és az alapján generálja ki a válaszokat.

2.2. Audio és video analitika

Az audio analitika elemzi és kinyeri az információkat a strukturálatlan hangadatokból. Ennek az elsődleges alkalmazási területe az ügyfélfelvétel központ (call center).

A hangelemző rendszerek egy élő hívást elemeznek úgy, hogy az ügyfél korábbi és jelenlegi interakciói alapján valós időben fogalmazzanak meg ajánlásokat az ügynököknek.

A video analitika, más néven videótartalom-elemzés (VCA) számos olyan technikát foglal magában, amelyekkel a videofolyamatok megfigyelése, elemzése és értelmes információk kinyerése lehetséges. A videoanalitika másik lehetséges alkalmazása a kiskereskedelemben vásárlói magatartások vizsgálata.

2.3. Közösségi média és előrejelző analitika

A közösségi média analitika a közösségi médiacsatornák strukturált és strukturálatlan adatainak elemzésére vonatkozik. A közösségi médiában az információ két forrását a felhasználók által generált tartalmak (pl. érзések, képek, videók és könyvjelzők) és a hálózati entitások (pl. emberek, szervezetek és termékek) közötti kapcsolatok és kölcsönhatások jelentik.

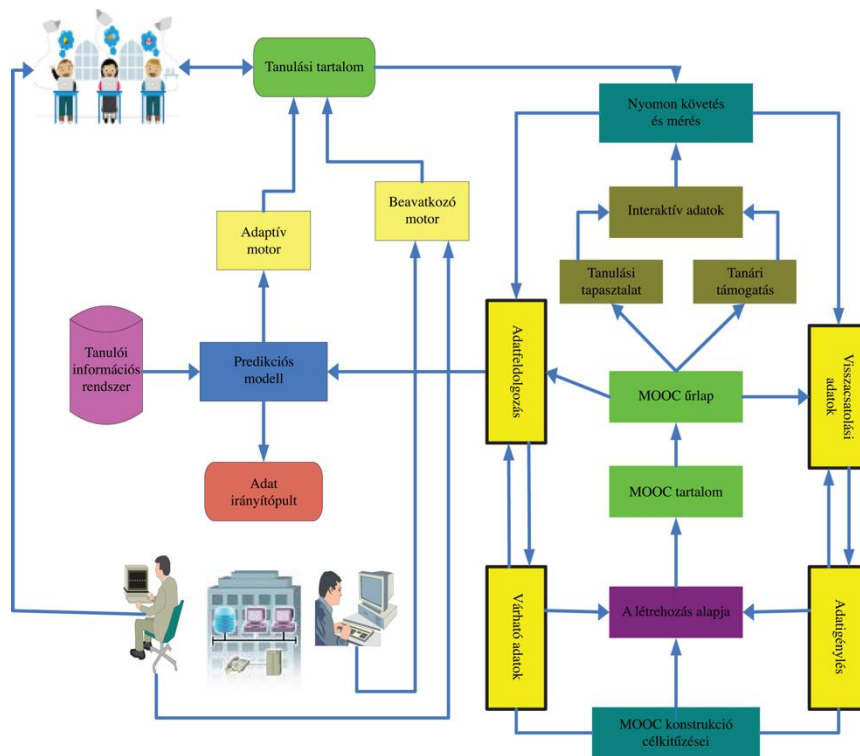
Az előrejelző analitika számos olyan technikát foglal magában, melyek a múltbeli és a jelenlegi adatok alapján előre diktálják a jövőbeli eredményeket. Az előrejelző analitika lényege, hogy az adatokban mintákat tárjon fel és összefüggéseket ragadjon meg.

3. Big Data alkalmazása az oktatás területén

3.1. Adatvezérelt intelligens osztályterem

A Beijing Normal University által kifejlesztett intelligens oktatási platform célja, hogy összegyűjtse a diákok tudásának, képességeinek, pszichológiájának és viselkedésének reprezentációs információit, illetve mélyreható adatbányászatot és modellezést végezzen. Ezzel a folyamattal nemcsak a diákok főiskolai felvételi vizsgájához tudományos referenciarendszert biztosítani, hanem segít a tanároknak a pontos tanításban és a diákoknak a tanulási problémák időben történő megoldásában.

Az intelligens osztályterem adatáramlási diagramja az 5. ábrán látható.



5. ábra Adatvezérelt intelligens osztályterem

A kép angol nyelvű forrása: Zhang, R., Zhao, W., & Wang, Y. (2020). Big data analytics for intelligent online education. *Journal of Intelligent & Fuzzy Systems*, 1–11. Academic Search Complete.

A Beijing Normal University intelligens osztálytermi környezete pontosabb és átfogóbb támogatást nyújt a tanárok tanítási és a tanulók tanulási döntéshozatalában, ezen kívül elősegíti az adatvezérelt pontos tanítást és tanulást.

A Big Data-nak köszönhetően a jövőben az is elképzelhető, hogy az adatvezérelt tanítás fokozatosan a fő oktatási paradigmává fog válni.

4. A kutatás folytatása

A kutatásom következő fázisában további kutatások szakirodalmi feltérképezését tervezem, ismertetem a Big Data technológiához kapcsolódó két kutatási terület, az oktatási adatbányászat és a tanulási analitika folyamatát, leggyakoribb módszereit (prediktív modellezés, klaszterelemzés, kapcsolatbányászat), illetve megvizsgálom egy online tanulást támogató portál kurzusának aktivitásait Big Data adatelemző módszerek, eszközök és platformok segítségével.

Irodalomjegyzék

Ahmadi, M., Dileepan, P., Wheatley, K. K. (2016). A SWOT analysis of big data. *Journal of Education for Business* 91(5), 1–294. Academic Search Complete

Ahmed, W., Ameen, K. (2017). Defining big data and measuring its associated trends in the field of information and library management. *Library Hi Tech News* 34(9), 21–24. Library, Information Science & Technology Abstracts. <https://doi.org/10.1108/LHTN-05-2017-0035>

Antal, P., Stókáné, P. M. (2015). Mobil eszközök alkalmazása iskolai környezetben. *Sárospataki Pedagógiai Füzetek különszám*, 193-211.

ESZTERHÁZY KÁROLY KATOLIKUS EGYETEM
INFORMATIKA KAR • DIGITÁLIS TECHNOLÓGIA INTÉZET
AGRIA MÉDIA KONFERENCIA 2023

Dobos, A., Antal, P., Vojtkó, A. (1999). Landscape analysis in concept of nature conservation in the Hór valley with GIS methods. In: Lehoczky, L., Kalmár, L. (szerk.) Proceedings of International Conference of PhD Students. Miskolc, University of Miskolc Innovation and Technology Transfer Centre, 47-57.

De Mauro, A., Greco, M., Grimaldi, M. (2015). What is Big Data? A Consensual Definition and a Review of Key Research Topics. AIP Conference Proceedings, 1644(1), 97–104. Academic Search Complete

De Mauro, A., Greco, M., Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review* 65(3), 122–135. Library, Information Science & Technology Abstracts.
<https://doi.org/10.1108/LR-06-2015-0061>

Dijcks, J.-P. (2011). Oracle: Big data for the enterprise [White paper]. Oracle Corporation

Franke, B., Plante, J.-F., Roscher, R., Lee, E.-S. A., Smyth, C., Hatefi, A., Chen, F., Gil, E., Schwing, A., Selvitella, A., Hoffman, M. M., Grosse, R., Hendricks, D., Reid, N. (2016). Statistical inference, learning and models in big data. *International Statistical Review. Revue Internationale de Statistique* 84(3), 371. MathSciNet via EBSCOhost.

Gandomi, A., Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* 35(2), 137–144. Library, Information Science & Technology Abstracts.
<https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

Kacfeh Emani, C., Cullot, N., Nicolle, C. (2015). Understandable big data: A survey. *Computer Science Review* 17(17), 70. MathSciNet via EBSCOhost.

Kamarási V., Mogyorósy G. (2015). Szisztematikus irodalmi áttekintések módszertana és jelentősége. Segítség a diagnosztikus és terápiás döntésekhez. *Orvosi Hetilap* 156(38), 1523–1531.
<https://doi.org/10.1556/650.2015.30255>

Kvaszingerne, P. Cs. (2018). Overview of the accessibility of webpages, related research, advantages, guidelines and recommendation. *Teaching Mathematics And Computer Science* 16:2, 233-262.

Madhavan, K., Richey, M. C. (2016). Problems in Big Data Analytics in Learning. *Journal of Engineering Education* 105(1), 6–14. Academic Search Complete

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity.

Nunes, S., Oliveira, T. A., Oliveira, A. (2020). Statistics and big data: Different perspectives. AIP Conference Proceedings, 2293(1), 420108–1. Academic Search Complete

Racsko, R., Kis-Tóth, L. (2019). A technológia szerepe a 21. századi tanár kompetenciájának fejlesztésében. *Katolikus Pedagógia: Katolikus Pedagógiai Tanszéki Folyóirat / Nemzetközi Neveléstudományi Szakfolyóirat* 8 : 1-2 49-65.

Szűts Z., Yoo J. (2016). Big Data, az információs társadalom új paradigmája. *Információs Társadalom* 16(1), 8. <https://doi.org/10.22503/infars.XVI.2016.1.1>

Tóthné, P. L., Lengyelne, M. T., Kis-Tóth, Lajos (2014). Statisztikai programrendszerek, Eger, EKF Líceum Kiadó

Wang, J., Zhao, B. (2020). Intelligent system for interactive online education based on cloud big data analytics. *Journal of Intelligent & Fuzzy Systems* 1–11. Academic Search Complete

Ward, J. S., Barker, A. (2013). Undefined By Data: A Survey of Big Data Definitions (arXiv:1309.5821). arXiv. <http://arxiv.org/abs/1309.5821>

Yeoman, I. (2019). Big Data. *Journal of Revenue and Pricing Management* 18(1), 1–1.
<https://doi.org/10.1057/s41272-019-00191-9>

Zhang, R., Zhao, W., Wang, Y. (2020). Big data analytics for intelligent online education. *Journal of Intelligent & Fuzzy Systems* 1–11. Academic Search Complete