

Dunaikavics

A Dunaújvárosi Egyetem online folyóirata 2025. XIII. évfolyam II. szám

Műszaki-, Informatikai és Társadalomtudományok



Válogatás a Dunaújvárosi Egyetemen 2024-ben tartott Tudományos Hét *Informatikatudományi* szekció előadásaiból.

Az írások szerzői:

ADAMCSIK JÁNOS

BOGNÁR LÁSZLÓ

FARKAS IMRE

JOÓS ANTAL

HADARICS KÁLMÁN

HADARICSNÉ DUDÁS NÓRA ERZSÉBET

KOCSÓ EDINA



Dunakavics

A Dunaújvárosi Egyetem online folyóirata 2025. XIII. évfolyam II. szám

Műszaki-, Informatikai és Társadalomtudományok

MEGJELENIK ÉVENTE 12 ALKALOMMAL

SZERKESZTŐBIZOTTSÁG

András István, Bacsa-Bán Anetta, Balázs László, Kovács-Bokor Éva,
Nagy Bálint, Németh István, Rajcsányi-Molnár Mónika.

Felelős szerkesztő Németh István

Szerkesztők: Falus Orsolya, Halmai Nóra, Kőkuti Tamás, Varga Anita

Tördelés Duma Attila

Szerkesztőség és a kiadó címe 2400 Dunaújváros, Táncsics M. u. 1/a.

Kiadja DUE Press, a Dunaújvárosi Egyetem kiadója

Felelős kiadó Dr. habil András István, rektor

<http://dunakavics.uniduna.hu/>

ISSN 2064-5007

Tartalom

ADAMCSIK JÁNOS

Hol a helyem a világban? – Morfondírozik a lassan 40 éves Excel bácsi

5

LÁSZLÓ BOGNÁR-ANTAL JOÓS

The Impact of Allowing ChatGPT on Student Scores and Test Completion Time in a Mid-Term Math Exam

11

BOGNÁR LÁSZLÓ

Így szoktuk elrontani az oktatási felméréseinket már a kérdésfeltevésnél

25

FARKAS IMRE

Műveltséggel kapcsolatos kérdőívek élete a közösségi felületen és az így szerzett tapasztalatok

37

HADARICS KÁLMÁN

Linux-szerver automatikus telepítése on-premise környezetben

41

HADARICSNÉ DUDÁS NÓRA ERZSÉBET

Adatbázis-trendek és technológiák: Hagyományos és modern megközelítések

49

KOCSÓ EDINA

A tanuláselemzés fejlődése és nemzetközi alkalmazása az oktatásban

59

Galéria (Duma Bálint fotói)

68

Hol a helyem a világban? – Morfondírozik a lassan 40 éves Excel bácsi

Összefoglalás: A munkahelyeken, függetlenül attól, hogy valaki kis cégnél dolgozik vagy multinál, talán a legtöbbet használt program a Microsoft Excel, emellett a legnépszerűbb is. Szerettem volna egy kis összefoglalót készíteni arról, hogy honnan indult és hová érkezett és hol tart most, hogy már 39 éves; az ereje teljében van vagy már vágja a centit, közeledve a nyugdíjhoz. Azt láthatjuk, hogy nagy fejlődésen ment keresztül és képes volt a folyamatos kihívásokra reagálni, valamint a kor technikai vívmányait integrálni. Az oktatások során szerzett tapasztalataim alapján úgy látom, hogy erősen indokolt olyan irányba fejlődnie, hogy a kevésbé gyakorlott felhasználók (zömében ez a jellemző) is alkalmazni tudják az egyre inkább bővülő lehetőségeit és ne kihívás legyen számukra a program alkalmazása, hanem igazi segítség. **Kulcsszavak:** Excel, Microsoft, függvények, VBA, programozás, automatizálás, Copilot.

Abstract: In workplaces, whether one works for a small company or a multinational corporation, Microsoft Excel is perhaps the most used and also the most popular program. I would like to provide a brief summary of its journey from where it started, where it has arrived now that it's 39 years old and still going strong or possibly trimming down as it approaches retirement. We can see that it has undergone significant development, consistently responding to challenges and integrating technological advancements of its time. Based on my training experiences, I strongly believe it is justified to further develop in a direction where less experienced users (which is often the case) can effectively utilize its expanding capabilities. The goal should be to make using this program not as a challenge but a genuine aid for them. **Keywords:** Excel, Microsoft, functions, VBA, programming, automation, Copilot.

* Dunaiújvárosi Egyetem,
Informatikai Intézet
Email: adamcsik@uniduna.hu

Hogyan született meg az Excel? Ej, ki emlékszik már erre!

Az Excel első verzióját 1985-ben adta ki a Microsoft az Apple Macintosh számítógépek számára, így 2024-ben már 39 éves. Sokan kötik Simonyi Károly nevéhez ezt a programot is, de ő inkább a Word atyjának tekinthető, de elvülhetetlen szerepe volt az Excel fejlesztésében is, hiszen az Excel elődjének tekinthető Multiplan táblázatkezelőben játszott szerepet. Érdekesség még, hogy a Microsoftnál csak 1987-ben jelent meg Windows 2.0-ás verzióban a továbbfejlesztett Excel 2.0.

Fejlődik az Excel? Fiatalabb már nem leszek, de okosabb még lehetek!

Az Excel a Microsoft Office csomag talán legnépszerűbb programja lett, ma már szinte a legtöbb cégnél mindennapos eszköz. A népszerűsége a felhasználási lehetőségek mellett talán annak is köszönhető, hogy szinte minden nyelvre lefordították, a magyar változat 2000-ben jelent meg, ahol nem csak a felületeket magyarították, hanem a függvényeket is lefordították magyarra. Mindezt természetesen úgy oldották meg, hogy a táblázatok multilingual környezetben is működőképesek.

A népszerűség másik oka talán az lehetett, hogy folyamatosan fejlesztették és fejlesztik, támogatják a felhasználási lehetőségeket és a funkcionális alkalmazást is. Tekintsük át a legfontosabb területek fejlődését!

Függvények

Az Excel 2.0-ban összesen 10 alapvető függvény volt elérhető, a mostani verziójában már több, mint 600 függvény található. Természetesen ennyi függvényt senki nem használ, egy átlagos felhasználó maximum 10–20 függvényt alkalmaz, de nem biztos, hogy mindenki ugyanazt a 10–20 függvényt. A funkcionalitás bővítése érdekében mindenképpen szükség volt a függvénytár duzzasztására. E függvények közül nem mindegyik érhető el magyarul, de folyamatosan bővítik a magyarra lokalizált változatot.

A helyi változatok elkészítésénél nem csak a fordítást kell megoldani, hanem néha a helyi sajátosságokhoz kell alkalmazni a függvényt. Tipikus példa erre a dátumkezelő függvények, ahol minden nyelven máshogy nevezik a hét napjait. Ezért itt igyekeztek már a függvényeket is úgy megoldani, hogy könnyen beépíthető legyen a helyi megnevezések használata. Pl. a HÉT.NAPJA-függvény csak sorszámot ad vissza, amit a HA-függvénnyel lehet magyar megnevezésekre konvertálni.

A dátumból pedig számformátumokkal tudjuk kinyerni a hét napjai megnevezését vagy használhatjuk a SZÖVEG(A1; „nnnn”) függvényt, ami áttételesen ugyanezt valósítja meg.

Van még mindig olyan függvény, amit várnának a magyarok is, például az angol verzióban létezik a BAHTTEXT-függvény, ami egy pénznemre vonatkozó számot szövegesen jelenít meg (1.121.538 – egy-millió-százhuszonegyezer-ötszázharmincnyc):

Ezt a magyar verzióban csak programozással lehet megvalósítani vagy bonyolult segédtablák felhasználásával. Ha valaki ismeri a VBA-t, akkor persze tud olyan függvényt készíteni, ami ugyanúgy használható, mint a többi Excel-függvény, csak mögötte egy program működik.

Érdekessége még a magyar nyelvű fordításnak, hogy a COUNTIF-függvényt először DARABTELI-függvénynek fordították le, majd a 2010-es verzióban áttértek a DARABHA elnevezésre.

Ezt az indokolta, hogy közben megjelentek az alapfüggvényekhez kapcsolt HA-funkciók, így a SZUMHA vagy az ÁTLAGHA és az egységesítés miatt célszerű volt ezt átnevezni.

A függvények bővítésének van egy olyan vonulata is, hogy figyelik a felhasználói igényeket, nehézségeket és ezekre próbálnak reagálni. Ilyen fejlesztés volt az XKERES-függvény megalkotása, mert nagyon sok felhasználónak problémát jelentett az FKERES korlátjainak leküzdése az INDEX- és a HOL.VAN- függvényekkel.

Funkciók

Folyamatosan azt tapasztalhatjuk, hogy funkcionálisan bővül az Excel, igazodva a felhasználói igényekhez. Ilyen folyamatos bővítést láthatunk a Querty esetében is, ami először csak bővítményként (add-in) került be, majd része lett az Excelnek, sőt ma már a Power Query segíti a külső forrásokból való adatátvételt. A Pivot-tábla is először bővítményként jelent meg az Excelben, majd beépítették és továbbfejlesztették, így ma már a Power Pivot is elérhető, de bővítményként, tehát ha alkalmazni szeretnénk, akkor meg kell jeleníteni.

Ezek mellett számos kiegészítővel rendelkezik még az Excel, szinte minden területre találunk megoldást. Például egy nyilvántartórendszerben gyakran szükséges lehet QR-kódok alkalmazása, erre is találunk kiegészítőt a QR4Office-t. De a Power Query és Power BI már beépítve tartalmazza ezt a funkciót.

Ha már szóba került a Power BI (olyan üzleti intelligencia eszköz, amely lehetővé teszi az adatok elemzését, vizualizálását és megosztását a vállalatokon belül), ami az Excelből építkezve került kifejlesztésre és bár nem része az Excelnek, de szoros kapcsolatban áll vele, és az Excel-felhasználók számára integrált funkcióként érhető el.

Programozás

Az Excel 5.0 verziójában jelent meg 1993-ban Visual Basic for Applications (VBA) bevezetése, ezzel lehetővé vált a felhasználók számára, hogy makrókat készítsenek és automatizáljanak folyamatokat.

Itt jelent meg a makróörögzítés lehetősége is, amivel kitárták a kaput azon felhasználók számára is, akik nem ismerik a programozás fortélyait. Ez a mai napig része az Excelnek és nagy segítséget nyújt a fejlesztőknek is, mivel sok esetben a hosszadalmas kódolás helyett alkalmazhatják ezt a funkciót.

Emellett integrálásra kerül az Office JavaScript API, ami lehetővé teszi a fejlesztők számára, hogy testreszabják az Office-alkalmazásokat, automatizálják a feladatokat és integrálják az adatokat más rendszerekkel.

Ez kimondottan a fejlesztők számára készült, hogy saját kiegészítőikkel bővíthessék az Excelt. 2023-tól pedig már hivatalosan is használható a Python in Excel. Ez lehetővé teszi a Python-kód futtatását közvetlenül az Excel munkafüzetekben. Ennek segítségével a Python adatkezelési képességeit használhatjuk az Excel népszerűségét kihasználva. Az Office 365 megjelenésével lehetővé vált az Excel online használata is, de ez nem támogatja a VBA-kódokat, makrókat, azokat csak a desktop verzióban tudjuk futtatni.

Az Excel Online-ban a Power Automate tesztli lehetővé, hogy különböző műveleteket automatizáljunk a felhőben. Emellett elérhető a Office Scripts funkció, amely lehetővé teszi a JavaScript-alapú szkriptek írását az Excel Online-ban. Ezzel lényegében ki lehet dobni az eddig használt VBA-kódjainkat.

Mesterséges intelligencia (AI)

A ChatGPT 2023. márciusi nyilvános indulásától kezdve, természetesen ismeri az Excelt is, így rendkívül jól használható a mindennapi munka során. Sokkal gyorsabban tud a felmerült kérdésekre választ adni, mint a keresők vagy az Excel beépített help-je. Ráadásul ismeri a VBA-t is, így egyszerű feladatok megoldására szinte azonnal használható kódokat képes generálni. Ezen felül feltölthetünk Excel-fájlokat is, ezekben képes elemezni és feldolgozni is. Képes akár diagramokat vagy statisztikaiösszegzéseket készíteni, de bonyolultabb számításokat is végezni. A Microsoft az Office 365 előfizetőinek 2024-től tette elérhetővé a Copilot rendszerét. Ez segít az adatok kezelésében és elemzésében, javaslatokat tud adni munkafolyamatokra, hogy gyorsabbá és hatékonyabbá tegye a munkát.

Leszázálekolás vagy megújulás?

Mivel széles körben használják az Excelt és mint ahogy említettem, talán a legnépszerűbb programja az Office-nak, így előreláthatóan nem küldik még nyugdíjba és várhatóan nem is tudják leváltani még jó darabig az ERP-rendszerek, azok rugalmatlansága és költségigényes alkalmazása miatt. Különösen a KKV- és a magánszektorban hosszú távon még erőteljes segítséget jelent az üzleti elemzésekben.

A tapasztalatok azt mutatják, hogy népszerűsége ellenére kevesen használják ki a lehetőségeit, még mindig ott van az „ágyúval lövünk verébre”-hatás az alkalmazásánál. Várhatóan ebben nagy előrelépés lesz az AI, a Copilot integrálása, ami segíteni fog az adatok gyorsabb elemzésében, akár szöveges válaszok generálásával, és támogatja majd az automatizálási lépéseket, hogy a napi feladatokat, mint például az e-mailek előkészítése vagy a megbeszélésekre való felkészülés, része legyen a feladat megoldásának.

Hivatkozások

<https://www.controllingportal.hu/nyugdijba-kuldi-vagy-felturbozza-az-ex-celt-az-ai/>

https://hvg.hu/tudomany/20230317_microsoft_office_365_irodai_programok_mesterseges_intelligencia_asszisztens_copilot_word_excel_power-point_outlook_teams

<https://www.vg.hu/karrier/2024/01/meg-excel-gomb-mesterseges-intelligencia>



The Impact of Allowing ChatGPT on Student Scores and Test Completion Time in a Mid-Term Math Exam

Abstract This study explores the impact of ChatGPT usage on student performance in a mid-term mathematics exam, focusing on two key research questions. First, whether there is a significant difference in the average total scores between students using ChatGPT and those not using it. The findings indicate that students who used ChatGPT scored significantly lower on average than their non-GPT-using peers, with fewer high-achieving students and a higher proportion failing to meet the minimum performance threshold. Second, the study examines whether there is a significant difference in the average time taken to complete the test between the two groups. Surprisingly, no notable difference in completion time was observed, challenging the assumption that ChatGPT users would either complete the exam faster or spend time comparing their answers with those generated by GPT. These results highlight the need for further investigation into the role of AI tools like ChatGPT in education, particularly their effectiveness in enhancing learning outcomes in mathematics.

Keywords: ChatGPT; AI in education; student performance; mathematics exam; test completion time; AI tools; mid-term exam; learning outcomes; educational technology; AI-assisted learning.

Absztrakt: Ez a tanulmány a ChatGPT használatának hatását vizsgálja a diákok teljesítményére egy félévközi matematika vizsgán, két fő kutatási kérdésre összpontosítva. Először is, hogy van-e szignifikáns különbség a ChatGPT-t használó és nem használó diákok átlagos összpontszámai között. Az eredmények azt mutatják, hogy a ChatGPT-t használó diákok átlagosan jelentősen alacsonyabb pontszámot értek el, mint a ChatGPT-t nem használó társaik, kevesebb volt a kiemelkedő teljesítményt nyújtó diákok száma, és nagyobb arányban nem érték el a minimális teljesítmény küszöböt. Másodszor, a tanulmány azt vizsgálja, hogy a két csoport között van-e szignifikáns külön-

* University of Dunajváros,
Institute of Computer Engineering,
Department of Mathematics
Email: bognarl@uniduna.hu

** University of Dunajváros,
Institute of Computer Engineering,
Department of Mathematics
Email: joosa@uniduna.hu

[1] Gulwani, S.–Polozov, O.–Singh, R. (2017): Program Synthesis. *Foundations and Trends® in Programming Languages*, 4., (1–2.), pp. 1–119.

[2] Miller, L. A. (1981): Natural Language Programming: Styles, Strategies, and Contrasts. *IBM Systems Journal*, 20., (2.), pp. 184–215.

[3] Dohmke, T. (2022): *GitHub Copilot Is Generally Available to All Developers*. Retrieved from <https://github.blog/2022-06-21-github-copilot-is-generally-available-to-all-developers/>

[4] OpenAI. (2022): *Introducing ChatGPT*. Retrieved from <https://openai.com/blog/chatgpt>

[5] Hu, K. (2023): *ChatGPT Sets Record for Fastest-Growing User Base – Analyst Note*. Reuters. Retrieved from <https://www.reuters.com/technology/chatgpt-sets-recordfastest-growing-user-base-analyst-note-2023-02-01/>

[6] Mehdi, Y. (2023): *Reinventing Search with a New AI- powered Microsoft Bing and Edge, Your Copilot for the Web*. Retrieved from <https://blogs.microsoft.com/blog/2023/02/07/rein-venting-search-with-a-new-ai-powered-microsoft-bing-and-edgeyour-copilot-for-the-web/>

[7] Pichai, S. (2023): *An Important next Step on Our AI Journey*. Retrieved from <https://blog.google/technology/ai/bard-google-ai-search-updates/>

[8] Buolamwini, J.–Gebru, T. (2018): Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: *Conference on Fairness, Accountability and Transparency*, pp. 77–91. PMLR.

[9] Liang, P. P.–Wu, C.–Morency, L.–P.–Salakhutdinov, R. (2021): Towards understanding and mitigating social biases in language models. In: *International Conference on Machine Learning*, pp. 6565–6576.

[10] Bender, E. M.–Gebru, T.–McMillan-Major, A.–Shmitchell, S. (2021): On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610–623.

ség a teszt kitöltéséhez szükséges átlagos idő tekintetében. Meglepő módon a teszt kitöltési ideje nem különbözött számottevően, ami megkérdőjelezi azt a feltételezést, hogy a ChatGPT-felhasználók vagy gyorsabban teljesítik a vizsgát, vagy időt töltenek a GPT által generált válaszok összehasonlításával. Ezek az eredmények rávilágítanak arra, hogy tovább kell vizsgálni a ChatGPT-hez hasonló mesterséges intelligencia eszközök szerepét az oktatásban, különösen a matematika tanulási eredményeinek javításában való hatékonyságukat.

Kulcsszavak: ChatGPT; mesterséges intelligencia az oktatásban; tanulói teljesítmény; matematika vizsga; teszt kitöltési ideje; AI-eszközök; félévközi vizsga; tanulási eredmények; oktatási technológia; mesterséges intelligenciával támogatott tanulás.

Introduction

For decades, research areas such as neural networks, programming synthesis [1], and natural language programming [2] have been advancing steadily, but it is only recently that these technologies have entered the public spotlight through major commercial releases. In June 2022, GitHub Copilot, an AI-powered code generation tool, was launched after a year of private beta testing [3]: Shortly thereafter, in November 2022, OpenAI introduced ChatGPT [4], which gained 100 million users within just two months, setting a record for the fastest-growing app [5]: By early 2023, both Microsoft and Google integrated conversational AI like ChatGPT into their web search platforms [6; 7]: The rapid adoption of AI tools has sparked a range of concerns, including bias [8; 9], ethics [10], misinformation

[11], privacy [12], energy consumption [13], and the consolidation of corporate power [14]: In the education sector, particularly, educators are questioning the impact of AI tools on student learning and performance, wondering whether these tools enhance or undermine educational outcomes [15]:

In this study, we examine the effects of ChatGPT on student performance during a mid-term mathematics exam by focusing on two primary research questions:

Research Question 1: Is there a significant difference between the average total scores of students using ChatGPT and those not using it?

Research Question 2: Is there a significant difference in the time taken to complete the test between the two groups?

By addressing these questions, this research aims to offer insights into the role of AI tools like ChatGPT in academic assessments and their broader implications for educational environments.

Conditions of the experiment

The experiment was conducted with foreign students of *Mathematics 1* and *Engineering Mathematics 1* at the University of Dunaújváros, in Hungary. In the following we will refer to these subjects as Mathematics 1 only. In Mathematics 1, students write two tests in the Moodle Learning Management System during the semester. The final grade is determined by the sum of the scores achieved on these two mid-term tests. In this paper, we look only at student results in the first test. In this test, each student was asked 5 questions, one from each of the 5 sub-topics. The questions differed only in that they contained random parameters generated by the Moodle system. A total of 140 students submitted valid tests, 22 of which self-reported using ChatGPT. Hence, for these 140 tests, a total of $140 \times 5 = 700$ different questions were asked, and of these, $22 \times 5 = 110$ questions were solved using ChatGPT.

Learning material for the test covers the basic topics of linear algebra, which includes an introduction to matrices, matrix operations, calculating

[11] Kreps, S.–McCain, R. M. –Brundage, M. (2022): All the News That's Fit to Fabricate: AI-generated Text as a Tool of Media Misinformation. *Journal of Experimental Political Science*, 9., (1), pp. 104–117.

[12] Butterick, M. (2022): *GitHub Copilot Investigation*. Joseph Saveri Law Firm & Matthew Butterick. Retrieved from <https://githubcopilotinvestigation.com/>

[13] Strubell, E.–Ganesh, A.–McCallum, A. (2019): *Energy and Policy Considerations for Deep Learning in NLP*. arXiv preprint arXiv:1906.02243.

[14] Xiang, C. (2023): *OpenAI Is Now Everything It Promised Not to Be: Corporate, Closed-Source, and For-Profit*. Retrieved from <https://www.vice.com/en/article/5d3naz/openai-is-now-everything-it-promised-not-to-be-corporate-closed-source-and-for-profit>

[15] Johnson, A. (2023): *ChatGPT In Schools: Here's Where It's Banned—And How It Could Potentially Help Students*. Retrieved from <https://www.forbes.com/sites/ariannajohnson/2023/01/18/chatgpt-in-schools-heres-where-its-banned-and-how-it-could-potentially-help-students/>

determinant and inverse, operations with vectors such as scalar multiplication, vector multiplication and mixed multiplication, calculating the angles of vectors, and addition, subtraction, multiplication and division of complex numbers given in algebraic form. More specifically:

- The first question (Q1) was on addition, subtraction, multiplication, transposition, determinant, adjoint and inverse calculus with matrices.
- The second question (Q2) was on addition and subtraction of vectors, multiplication by scalar, linear independence of vectors, base of vectors, rank of matrix, scalar multiplication of vectors and solution of linear system of equations.
- The third question (Q3) was about mixed product of vectors, scalar product of vectors, equation of a plane, equation of a line, length of a vector, product of vectors and closed angle of vectors.
- The fourth question (Q4) was taken from the same set as the first question.
- The fifth question (Q5) was taken from the topics: real and imaginary parts of complex numbers, sum of complex numbers, difference of complex numbers, multiplication of complex numbers and absolute value of complex numbers.

Students have 45 minutes to complete the test. At the end of the 45 minutes, any test that has not been submitted by the student will be automatically submitted. The student will immediately see the result of their test and the correct answers. Teaching was face-to-face, but the test was online. Students were only told on the morning of the test that they could use ChatGPT for the test. The test included a self-report question about whether they had used ChatGPT during the test. There were no penalties or rewards for using ChatGPT. There was no use of ChatGPT at all in the mathematics lessons.

The mathematical problem-solving capabilities of ChatGPT and its pitfalls were not demonstrated. Students were allowed to use not only ChatGPT, but also other similar tools.

Comparison of Total scores of ChatGPT users and non-users

Figure 1. shows the distribution of the students' Total scores. The Total score is the sum of the scores for the 5 questions of the test. The mean Total score was 16.57, the standard deviation was 5.018. It seems that most students passed the mid-term test, with many scoring the maximum 20 points. Only 14 students (10%) failed to reach the minimum performance of 60%, i.e., 12 points. However, if we split the students by GPT use, we get a more nuanced picture.

Figure 1. The distribution of students' Total scores

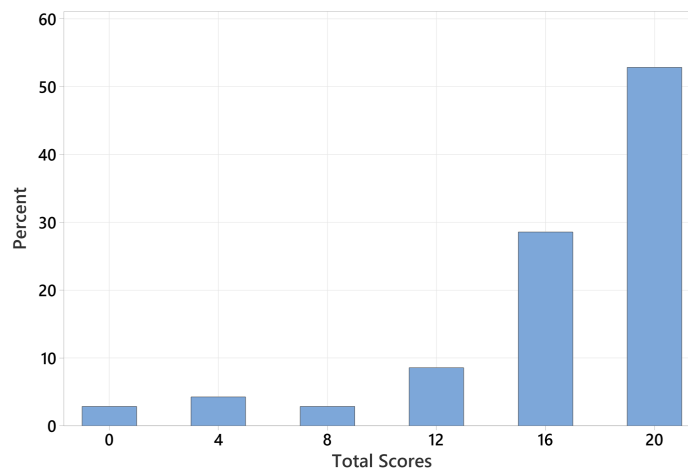
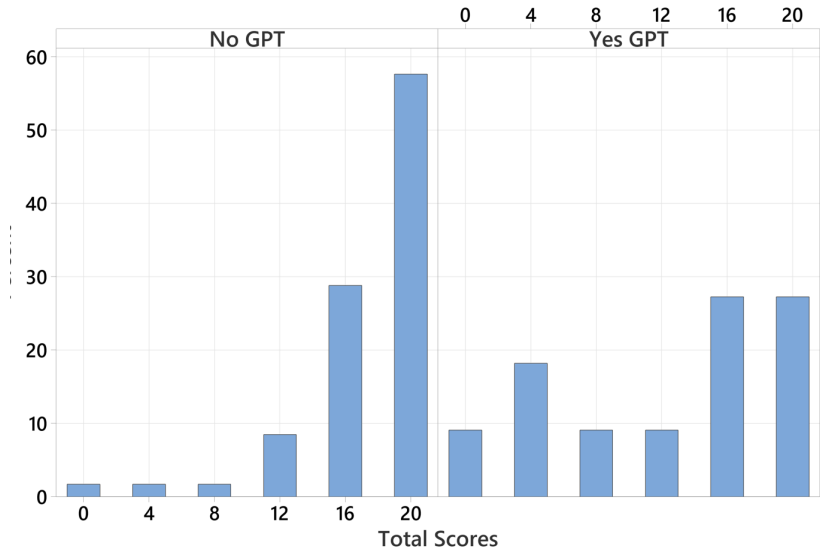


Figure 2. shows the distribution of Total scores for students who do not use ChatGPT and students who use ChatGPT separately. The mean Total score for students not using ChatGPT was 17.356 with a standard deviation of 4.110, and the mean Total score for students using ChatGPT was 12.36 with a standard deviation of 7.26. On the one hand, it is striking how much lower the average score of those who used ChatGPT was, and on the other hand, the distribution of scores in the two cases varies considerably. Among the ChatGPT users the highest scorers are few, and while 36.3% of them failed to reach the minimum 12 points, only 5.1% of non-users failed.

To check that this trend is not only apparent in these samples, but that a similar result would be obtained from the whole population, we performed a two-sample t-test for considering the difference of the means of the Total scores. Let μ_1 be the population mean of the questions Total score when GPT was not used and let μ_2 be the population mean of the questions Total score when GPT was used. Let the null hypothesis be $\mu_1 - \mu_2 = 0$ and the alternative hypothesis be $\mu_1 - \mu_2 > 0$. Now the observed t-value is 4.04, and the P-value is less than 0.0001. This means that the difference between the mean Total scores is highly significant.

Figure 2. The distribution of students' Total scores for students not using ChatGPT and students using ChatGPT



As the distributions of the two samples are quite different, and thus the application of the t-test is questionable, we performed a Mood's Median Test as well. Here the observed Chi-square-value is 3.43, and the P-value is 0.064.

This means that the difference between the medians is also significant at this level. Both tests support our finding that, for the given experimental set-up and conditions, there is a significant difference between the mean of the Total scores of students who use GPT and those who do not use GPT. Those using ChatGPT score significantly lower on average.

It is quite surprising. One would think that the GPT would be an extra help in solving tasks and therefore GPT users would have a higher score. Further research questions will address this phenomenon, among others. What are the causes and factors behind:

Question 4. How do students who used ChatGPT differ from other students? Maybe they are the ones who studied less for the test, and this is reflected in the results?

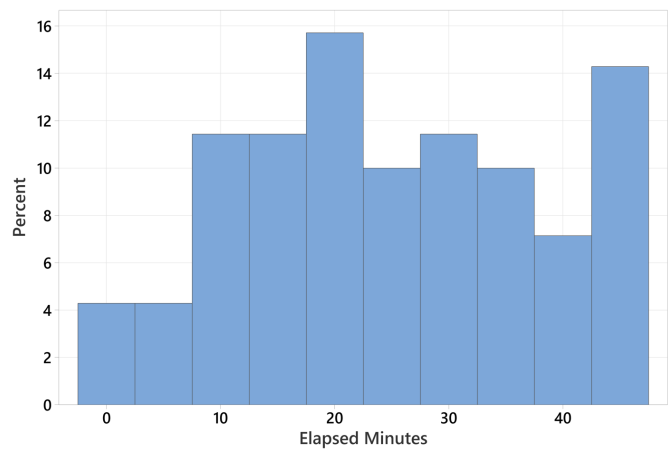
Question 5. Even if they studied less, why didn't the use of ChatGPT make up for it? How correct are the answers given by ChatGPTs?

Question 6. How consistent are the answers given by ChatGPTs? Do all GPTs give the same answer to the test questions in all cases?

Comparison of time used by GPT users and non-users

Figure 3. shows the distribution in minutes of the time students spent solving the test. The average number of minutes was 24.64 with a standard deviation of 12.9 minutes.

Figure 3. The distribution of elapsed minutes

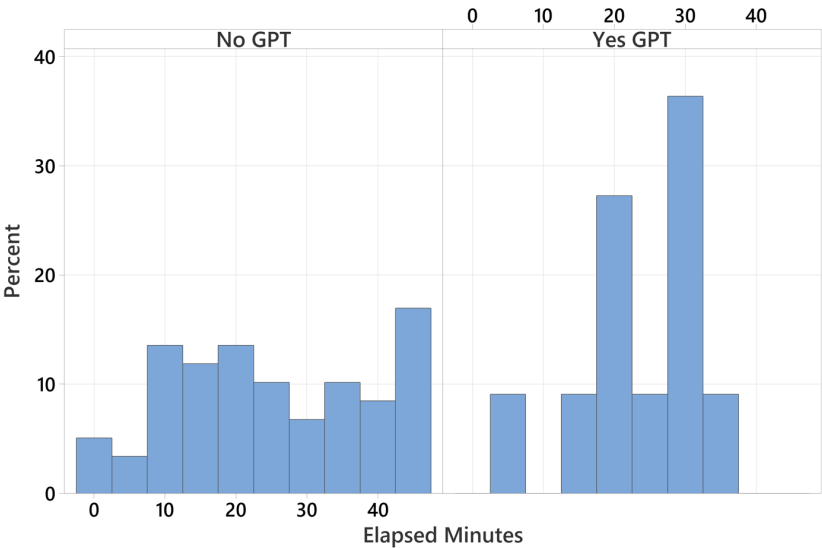


This average value, this distribution, suggests that the test was not difficult in this sense, with only a few students taking the maximum 45 minutes allowed. About half of the students used between 15 and 35 minutes.

Figure 4. shows the relative frequency histogram of elapsed minutes for students not using ChatGPT and students using ChatGPT. The mean of elapsed minutes of students not using ChatGPT was 24.88 with standard deviation 13.6 and mean 23.36 standard deviation 8.18 of students using ChatGPT. The average times used by the two groups are very close. Although the distribution of the data shows some differences in the two samples, the normality test for both samples confirmed that they could be from a normally distributed population, so there is reason to believe that these average times would not be significantly different in the populations.

We performed a two-sample t-test to prove this. For the null hypothesis $\mu_1 - \mu_2 = 0$, and the alternative hypothesis $\mu_1 - \mu_2 \neq 0$ we obtained the observed t-value of 0.50 and the P-value of 0.622. These values support our finding that the average durations used by the two groups are not significantly different.

Figure 4. The distribution of elapsed minutes for students not using ChatGPT and for those using ChatGPT



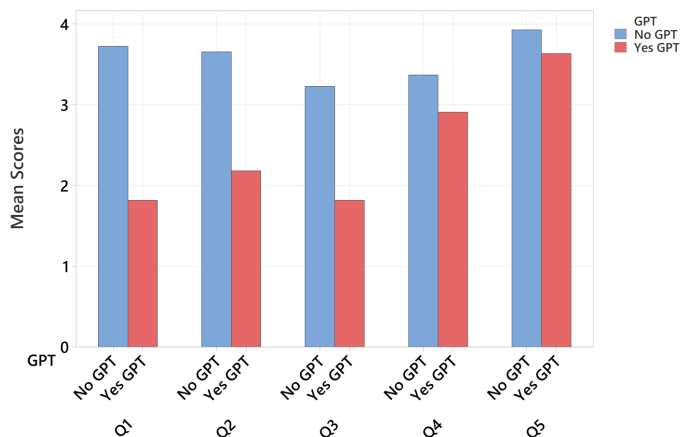
This may also be considered surprising. On the one hand, we might have thought that those who use GPT would finish sooner. On the other hand, we might have thought that someone who uses GPT would get away with comparing the GPT answer with their own answer. This is not the case.

Comparison of scores for non-GPT users and GPT users for questions on different topics

Figure 5. shows the mean scores of the questions on different topics for those who do not use ChatGPT and those who use GPT. It seems that those without GPT have a more balanced performance on different questions, on different topics. For each question, the average score obtained was between 3.2 and 3.9 points.

The same cannot be said for those using GPT, where the results were quite extreme. Question 1, which included calculating the inverse of matrices, had the lowest score of 1.8, while the highest score of 3.6 was for question 5 on complex numbers. It is particularly surprising that their average scores for the first and fourth questions are so different (1.8 and 2.9): Both questions were on the same topic, different questions from the same question bank. If both questions were based on ChatGPT answers, the question arises again whether ChatGPT gives the correct answer in all cases.

Figure 5. The average scores for questions on different topics for those who do not use ChatGPT and those who use GPT



An analysis of variance (ANOVA) was carried out to examine the significance of differences in the responses to the different topics. A two-factor interaction model was used. The „Question” factor has 5 levels (Q1; Q2; Q3; Q4; Q5) and the „GPT Use” factor has two levels (Yes GPT; No GPT), so that in total $5 \times 2 = 10$ factor level combinations were used to examine the differences in average response scores. Since not only the two factors but also the interaction term was found to be significant in the model, it is worth comparing the average scores for the 10 factor level combinations.

Table 1. shows the grouping using the Tukey Method and 95% confidence.

Table 1. Comparison of the average scores for the different questions

	Question*GPT Use	Mean	Grouping		
1.	Q5 No GPT	3.92	A		
2.	Q1 No GPT	3.72	A		
3.	Q2 No GPT	3.65	A		
4.	Q5 Yes GPT	3.63	A	B	
5.	Q4 No GPT	3.36	A	B	
6.	Q3 No GPT	3.22	A	B	
7.	Q4 Yes GPT	2.90	A	B	C
8.	Q2 Yes GPT	2.18		B	C
9.	Q3 Yes GPT	1.81			C
10.	Q1 Yes GPT	1.81			C

According to this notation, if we want to compare two factor level combinations, those that have at least one grouping symbol (A, B, ...) that appears in both are not significantly different. Thus, there is no significant difference between the cases in the first 7 rows of the table, because the group A symbol appears in each case. The case in row 8 (Q2-Yes GPT) is different from the cases in the first three rows. The cases in rows 9 and 10 are significantly different from the cases in the first six rows. This clustering also shows that, for those using GPT, the mean scores for questions 1 and 4 are significantly different, even though these questions are from the same topic and the same question bank. The rows in the table are arranged in descending order of average scores. Question 5 on complex numbers was among the easy questions for both GPT users and non-users. Among the GPT users, the lowest scores were for questions 1 and 3, i.e. the inverse of matrices and different types of multiplication of vectors were the most challenging.

Student learning activities

The students took the test in Moodle, where their learning activity was tracked. By activity, we mean the number of clicks made by the student. So, we could measure the number of times a student clicked on the file containing the course material or completed the practice test.

Figure 6. shows student activity by received scores. The first column shows that the average activity of students with a score of 0 was close to 50 clicks. In addition, the last column shows that the average activity of students with a score of 20 was close to 270 clicks. Obviously, we would expect that students with higher scores would have been more engaged in learning the course material.

Figure 6. The mean number of activities per students vs. Total scores

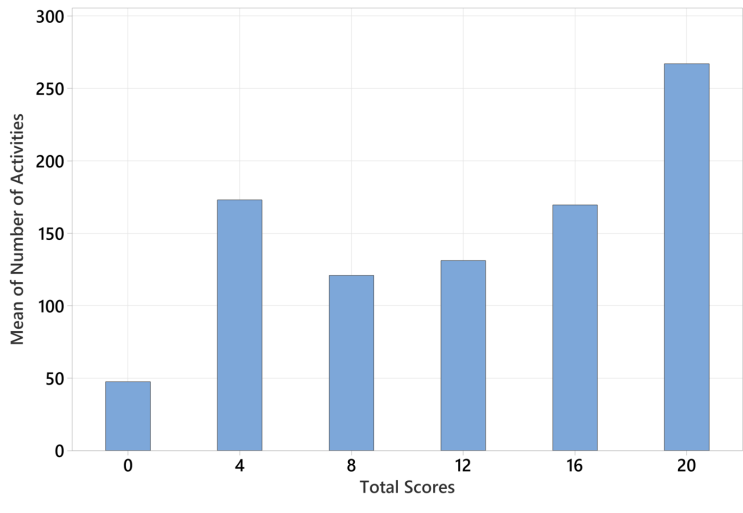


Figure 7. The mean number of activities per student vs. GPT usage

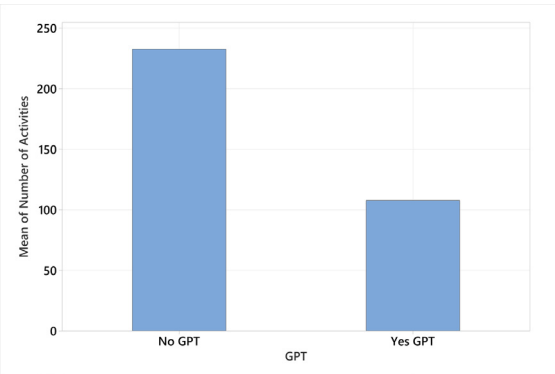


Figure 7. presents a comparison of the activity levels of students using GPT and those not using it. To determine whether there is a statistically significant difference between their activities, we conducted a two-sample two-sided t-test.

Let μ_1 be the population mean of clicks of students not using GPT and μ_2 be the population mean of clicks of students using GPT. Let the null hypothesis be- $\mu_1 - \mu_2 = 0$ and the alternative hypothesis be $\mu_1 - \mu_2 \neq 0$. Now the observed t-value is 2.33, and the P-value is 0.023. This means that the difference between the mean click counts is statistically significant.

Figure 8. shows the activity of students using GPT and students not using GPT by scores received. The average number of clicks for students not using GPT was extremely high at 4 scores. The activity of students not using GPT is generally higher. We can speculate that the less prepared students used a chatbot.

Figure 8. The mean number of activities per student vs. Total scores and GPT usage

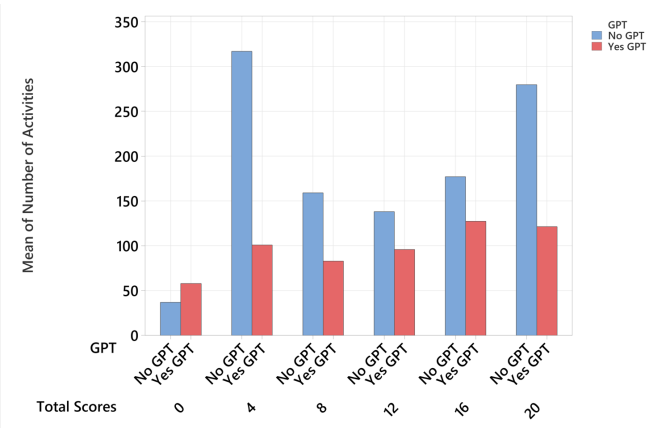
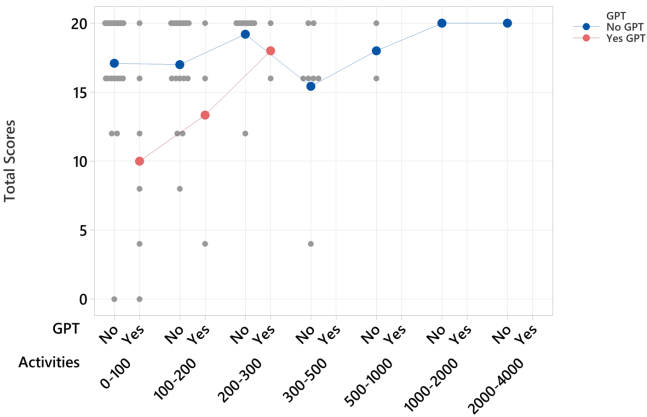


Figure 9. shows the individual value plot of the total scores and the trends. Again, it is obvious that the student who has practiced more will get a higher score.

This can be seen in the trend of the students who use GPT and in the trend of the students who do not use GPT.

Figure 9. The Total scores vs. the mean of the number of activities and GPT usage



Conclusions

In this study examining the impact of using ChatGPT in a mid-year math exam, several noteworthy findings emerged, shedding light on the complexities and nuances associated with integrating AI tools into educational settings.

Performance Discrepancy:

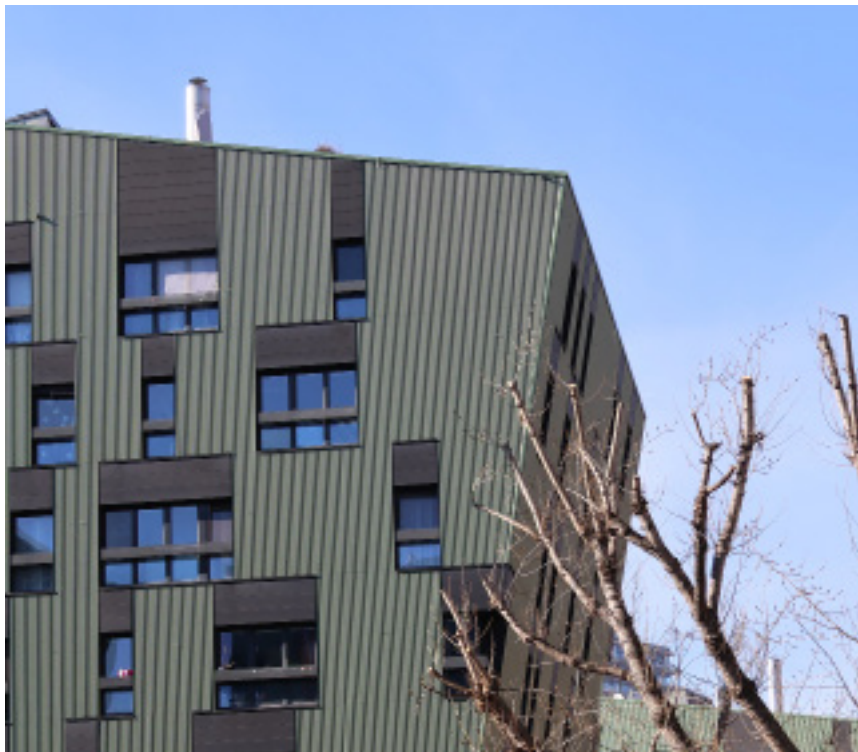
Contrary to expectations, students using ChatGPT demonstrated significantly lower average total scores compared to their non-GPT counterparts. The distribution of scores among ChatGPT users exhibited fewer high achievers, and a notable proportion failed to reach the minimum performance threshold.

Time Usage:

Surprisingly, the average time taken by students using ChatGPT and those who did not, was not significantly different. The assumption that ChatGPT users might finish sooner or engage in comparison with GPT-generated answers was not supported by the data.

Student Learning Activities:

Student activity, measured by the number of clicks, revealed a positive correlation between engagement and higher scores. Notably, students who scored lower were generally less active, indicating a potential link between engagement in learning activities and academic performance. Despite the lower learning activity among ChatGPT users, the question still remains as to why both ChatGPT and Copilot failed to compensate for this lack of engagement.



Így szoktuk elrontani az oktatási felméréseinket már a kérdésfeltevésnél

Összefoglalás: A kérdőíves felmérések a felsőoktatás szívverései – legalábbis a mi számunkra, akik a statisztikától várjuk, hogy válaszoljon az élet minden nagy kérdésére. Igyekszem bemutatni, hogyan ronthatjuk el már az első lépéseknél a kutatási folyamatot, amikor a kérdéseinket megfogalmazzuk. Amikor nem nézzük meg, hogy a kérdéseink relevánsak-e, megbízhatóak-e, hogy a válaszok mérhetőek-e, és gyermeketegnek tartjuk az olyan módszereket, mint „face validity”, „content validity”, „pilot study”. Persze ezekre a vizsgálatokra tényleg csak akkor van szükség, ha feladjuk az eredmények fantáziadús értelmezhetőségének örömét, az új, soha nem látott tézisek felfedezésének katarzist. Összeszedtem a leggyakoribb hibákat – és egy-két jótanácsot is adok.

Kulcsszavak: Kérdőíves felmérések; felsőoktatás; statisztika; kérdésfeltevés; relevancia; mérhetőség; megbízhatóság; face validity; content validity; pilot study.

* Dunaiújvárosi Egyetem, Informatikai Intézet, Matematika és Számítástudományi Tanszék
Email: bognarl@uniduna.hu

Abstract: Surveys are the lifeblood of higher education research – at least for those of us who like to imagine that statistics can answer life's biggest questions. This paper takes a somewhat tongue-in-cheek look at how we often manage to derail our research process from the very start: designing survey questions. We skip over mundane details like whether our questions are relevant, measurable, or reliable, and scoff at “face validity,” “content validity,” or even pilot studies as unnecessary hurdles. After all, why settle for clarity and accuracy when we can bask in the joy of wildly creative interpretations or the thrill of conjuring groundbreaking (if implausible) new theses? With that spirit in mind, I’ve gathered the most frequent mistakes in survey design – along with a few tips, should anyone be curious about avoiding them.

Keywords: Survey research; higher education; statistics; question design; relevance; measurability; reliability; face validity; content validity; pilot study.

Bevezetés

A Dunaújvárosi Egyetemen – és úgy általában a felsőoktatásban – egyre több kérdőíves felmérést végzünk a hallgatók vagy éppen az oktatók körében. Egy statisztikus szemszögéből nézve azonban elég gyakran csúsznak hibák az elemzésekbe. Ahogy mondani szokás, előbb-utóbb biztosan elrontjuk, de azért az nagyon nem mindegy, hogy előbb vagy utóbb.

Sajnos a legnagyobb esélyünk arra van, hogy már az elején elhibázzuk. Ezt két területen tudjuk a legkönnyebben megtenni, a kérdések megfogalmazásában, és abban, hogy hogyan választjuk ki a válaszadókat. Aztán persze nem feledkezünk meg a biztos befutóról, a statisztikai elemzésekről, és az azokból levont következtetésekről. Három kis cikket szentelek ezeknek a problémáknak. Ez a cikk a kérdések megfogalmazásáról szól.

Milyen kérdéseket tegyünk fel, és hogyan?

NE A KÉRDŐÍV-KÉRDÉSEKKEL KEZDJÜK

A KUTATÁSI KÉRDÉSEK

Mielőtt elkezdenénk a konkrét kérdőív-kérdéseket (*survey questions, survey items*) megfogalmazni, elengedhetetlen tisztázni a kutatási kérdéseinket (*research questions*), hiszen ezek irányítják az egész folyamatot. Ha a kutatási kérdések zavarosak, az egész felmérés félresiklik. Íme, egy jó kutatási kérdés feltevésének legfontosabb szempontjai:

Célzott és világos: A kutatási kérdés határozza meg pontosan, hogy mit szeretnénk megvizsgálni, így lesz a kutatás célja világos és egyértelmű. Ez biztosítja, hogy releváns kérdéseket tegyünk fel, hogy releváns adatokat gyűjtsünk.

Mérhető: A kérdés olyan formában legyen megfogalmazva, hogy lehetővé tegye a mérhető és elemezhető eredményeket.

Specifikus, de nem túl szűk: A kutatási kérdés legyen elég specifikus ahhoz, hogy konkrét válaszokat adjon, de ne legyen annyira szűk, hogy ne lehessen általánosítani az eredményeket.

Újdonságot hordozó: A kutatási kérdésnek valami újat kell hozzátennie a meglévő tudáshoz, különösen, ha a témában már számos kutatás készült.

Érthető a kutatásban részt vevők számára: A kutatási kérdésnek világosnak és érthetőnek kell lennie a célcsoport számára, hogy a kutatásban részt vevők pontosan megértsék, mit vizsgálunk.

Példák a rossz és a jó kutatási kérdésekre:

1. Példa

Rossz: „Hogyan befolyásolja az online oktatási platformok használata a hallgatók teljesítményét a matematika tantárgyakban?” Ez a kérdés túl általános, mivel nem határozza meg, hogyan mérjük a „teljesítményt”, és milyen konkrét eredményeket várunk.

Jó: „Milyen változást okozott az online oktatási platformok használata a hallgatók vizsgaeredményeiben a matematika tantárgyak esetében a 2022/2023-as tanévben?” Ez a kérdés konkrét eredményre fókuszál (vizsgaeredmények), meghatározott időszakot és tantárgyat céloz, valamint mérhető változást vizsgál.

2. Példa

Rossz: „Jobbak-e a hallgatók vizsgaeredményei az online oktatásban, mint a hagyományos oktatásban?” Ez a kérdés túl általános, nem specifikálja a vizsgálni kívánt tantárgyat vagy célcsoportot, és nem ad támpontot a „jobbak” pontos értelmezésére.

Jó: „Milyen különbségek vannak az online és a hagyományos oktatásban részt vevő hallgatók vizsgaeredményei között az informatika alapszakos hallgatók körében?”

Ez a kérdés specifikusan összehasonlít két oktatási módszert a vizsgaeredmények alapján, egy adott célcsoportot megjelölve (informatika alapszakos hallgatók).

3. Példa

Rossz: „Milyen hatással van a valós idejű visszajelzés a távoktatásban a hallgatókra?” Itt a kérdés nem specifikus, nem világos, milyen „hatást” vizsgálunk, és nem határozza meg, mely szempontok alapján mérjük a visszajelzés eredményeit (pl. elégedettség, teljesítmény stb.).

Jó: „Hogyan változott a hallgatók elégedettségi szintje a valós idejű visszajelzés hatására a távoktatásban az előző félévhez képest?” A kérdés pontosítja, hogy milyen szintű változást keresünk (elégedettségi szint), és meghatározza az összehasonlítás alapját (előző félév).

4. Példa

Rossz: „Milyen hatása van a tanulási szokásokra az éjszakai tanulásnak?” Ez a kérdés túl általános, nem határozza meg, hogyan mérjük a „hatást” és mit értünk „tanulási szokások” alatt.

Jó: „Milyen változást tapasztaltak a hallgatók tanulási hatékonyságában a rendszeres éjszakai tanulás következtében a vizsgaidőszakban?” Ez a kérdés konkrétabb, mérhető eredményre fókuszál („tanulási hatékonyság”), és meghatározza a vizsgálat időszakát („vizsgaidőszak”).

A KÉRDŐÍV KÉRDÉSEI

Most, hogy tisztáztuk a kutatási kérdések szerepét és fontosságát, nézzük a helyes kérdőív-kérdésfeltevés szempontjait:

Kapcsolódás a kutatási kérdéshez: A felmérés minden egyes kérdésének világosan kapcsolódnia kell valamelyik kutatási kérdéshez. Ha egy kérdés nem segíti a kutatási cél megvalósítását, akkor felesleges.

Világosság és egyértelműség: A kérdések legyenek egyszerűek és könnyen érthetők. Minden válaszadónak ugyanúgy kell értelmeznie őket, hogy az eredmények összehasonlíthatók legyenek.

A kérdés céljának meghatározása: Minden kérdésnek legyen egyértelmű célja: mit szeretnénk megtudni vele? Például, ha egy adott viselkedés gyakoriságát mérjük, ne kérdezzünk véleményt, hanem konkrét számokat.

Ne legyen irányított: A kérdések ne sugallják a helyes választ. A semleges megfogalmazás elengedhetetlen.

A kérdések sorrendje: A kérdéseket úgy kell elrendezni, hogy logikusan kövessék egymást, és ne befolyásolják a válaszadókat. A könnyebb, általános kérdésekkel kezdünk, majd fokozatosan térünk rá a komplexebbrekre.

Teljeskörű válaszlehetőségek: Gondoskodjunk arról, hogy a válaszlehetőségek lefedjék a lehetséges válaszokat. Kerüljük a hiányos válaszlehetőségeket, és adjunk lehetőséget „egyéb” választ is adni, ha szükséges.

Előítéletek és kettős kérdések elkerülése: Két témát egy kérdésbe sűríteni, vagy előítéletes kifejezéseket használni félrevezető és zavart okoz.

Elemzés szempontjából értelmezhető válaszok: A kérdéseket úgy kell kialakítani, hogy a válaszok egyértelműen és könnyen elemezhetők legyenek.

Ne legyen túl fárasztó: Ha túl sok a kérdés vagy a kérdések/válaszok túl komplexek, túl sokáig tart elolvasni és értelmezni őket, akkor a kitöltők belefáradhatnak a folyamatba. Vagy abbahagyják a kérdőív kitöltését, vagy megbízhatatlan válaszokat adnak. Ha a kérdőív élvezetes, az növeli a kitöltés arányát és javítja az adatok minőségét is.

Példák a rossz és a jó kérdőív kérdésekre:

1. Példa

A kutatási kérdés: „Milyen hatása van az online oktatási platformoknak a hallgatók vizsgaeredményeire a matematika tantárgyban?”

Rossz: „Mennyire találja hasznosnak a tanárok által tartott konzultációkat egy online kurzusnál?” Ez a kérdés nem kapcsolódik közvetlenül a kutatási kérdéshez, mivel a konzultációk hasznosságát méri, nem pedig az online platformok hatását a vizsgaeredményekre. Eltér a fő fókuszról, és nem segíti a kutatási cél elérését.

Jó: „Mennyire segítette az online oktatási platform használata a vizsgára való felkészülést a matematika tantárgyban?” Ez a kérdés közvetlenül a kutatási kérdésre vonatkozik, és mérhető választ ad arról, hogyan befolyásolta az online platform használata a hallgatók felkészülését a vizsgára, ami összhangban van a kutatás céljával.

2. Példa

Rossz: „Mennyire szereti az online tananyagokat?”

Válaszlehetőségek: 1 (egyáltalán nem szereti) – 5 (nagyon szereti). Az alapvető probléma az, hogy a „szereti” túl szubjektív és többféleképpen értelmezhető. Néhány válaszadó az érzelmi élményt, míg mások a tananyag hasznosságát vagy könnyű használhatóságát értékelhetik, ami torzíthatja az eredményeket. Nem biztosít megbízható, mérhető adatokat, amelyek összehasonlíthatók lennének más válaszokkal.

Jó: „Mennyire érzi hatékonyak az online tananyagokat a tanulási folyamatában?”

Válaszlehetőségek: 1 (egyáltalán nem hatékony) – 5 (nagyon hatékony). Ez a kérdés specifikus, jól mérhető, skálázott választ biztosít, így az eredmények könnyen elemezhetők.

3. Példa

Rossz: „Milyen gyakran használja az online oktatási platformokat?”

Válaszlehetőségek: „Soha”, „Ritkán”, „Gyakran”, „Mindig”. A kérdés nem ad meg időintervallumot (pl. naponta vagy hetente), így a válaszok nem lesznek összehasonlíthatók, és nehéz lesz az adatok elemzése.

Jó: „Milyen gyakran használja az online oktatási platformokat egy hét alatt?”

A válaszlehetőségek: „Soha”, „1–2 alkalommal hetente”, „3–4 alkalommal hetente”, „5–6 alkalommal hetente”, „Napi rendszerességgel (7 alkalom vagy több hetente)”.

Ez a változat pontosabb képet ad arról, hogy a válaszadók milyen rendszerességgel használják a platformokat, így az eredmények jobban összehasonlíthatók és értelmezhetők. A kérdés világosan meghatározza az időintervallumot (egy hét), és lehetőséget biztosít az egyértelmű válaszokhoz.

4. Példa

Rossz: „Mennyire elégedett a visszajelzésekkel?”

Válaszlehetőségek: „Egyáltalán nem elégedett”, „Kissé elégedett”, „Semleges”, „Elégedett”, „Nagyon elégedett”. Ez a kérdés túl általános, mivel nem tisztázza, hogy pontosan milyen aspektusra vonatkozik az elégedettség (pl. gyorsaság, minőség, hasznosság). Így a válaszok nem lesznek egyértelműen értelmezhetők, és az elemzés nehezítetté válik.

Jó: „Mennyire elégedett a tanároktól kapott visszajelzések gyorsaságával?”

Válaszlehetőségek: 1 (nagyon elégedetlen) – 5 (nagyon elégedett). Ez a kérdés konkrétan egy aspektust mér (visszajelzések gyorsasága), és jól használható skálát biztosít az elemzéshez.

5. Példa

Rossz: „Mely tényezők befolyásolták leginkább az online tananyagok hatékonyságát az Ön tanulási folyamatában? (Válasszon ki akár több lehetőséget, és értékelje mindegyik fontosságát 1-től 5-ig terjedő skálán!)”

Válaszlehetőségek:

- A tananyag minősége (Értékelés: 1 – egyáltalán nem fontos, 5 – nagyon fontos)
- Az oktatási platform kezelhetősége (Értékelés: 1 – egyáltalán nem fontos, 5 – nagyon fontos)
- A tanárok visszajelzése (Értékelés: 1 – egyáltalán nem fontos, 5 – nagyon fontos)

- Az időbeosztás rugalmassága (Értékelés: 1 – egyáltalán nem fontos, 5 – nagyon fontos)
- A tananyag elérhetősége (Értékelés: 1 – egyáltalán nem fontos, 5 – nagyon fontos)
- A tanulók motivációja (Értékelés: 1 – egyáltalán nem fontos, 5 – nagyon fontos)
- A technikai problémák (Értékelés: 1 – egyáltalán nem fontos, 5 – nagyon fontos)
- Az internetkapcsolat sebessége (Értékelés: 1 – egyáltalán nem fontos, 5 – nagyon fontos)
- Egyéb (kérem, részletezze) (Értékelés: 1 – egyáltalán nem fontos, 5 – nagyon fontos)

Ez a multi choice kérdés túl sok opciót ad meg, és a válaszadók akár több tényezőt is választhatnak, ami nagymértékben megnehezíti az elemzést. Az ilyen típusú kérdéseknél az adatok rendszerezése és értelmezése szinte lehetetlen feladat, mivel a különböző válaszkombinációk száma nagyon magas lehet, ami megnehezíti a világos következtetések levonását. Ez még akkor is igaz, ha nem engedjük kiválasztani, hogy melyik válaszlehetőséget pontozzák, hanem az összeset pontozniuk kell.

Jó: „Az alábbi tényezők közül melyik befolyásolta leginkább az online tananyagok hatékonyságát az Ön tanulási folyamatában? (Válasszon egyet!)”

Válaszlehetőségek:

- A tananyag minősége.
- Az oktatási platform kezelhetősége.
- A tanárok visszajelzése.
- Az időbeosztás rugalmassága.
- A tananyag elérhetősége.
- A technikai problémák.
- Az internetkapcsolat sebessége.

Ez a változat egyszerűbbé teszi az elemzést, mivel a válaszadóknak csak egy tényezőt kell kiválasztaniuk, amelyet legfontosabbnak tartanak. A válaszok világosan összehasonlíthatók, mivel csak egy változó van minden válaszadónál, így könnyebb az egyes tényezők relatív fontosságát megállapítani. A válaszlehetőségek kódolása és az eredmények összehasonlítása egyszerűbb, mivel egyértelmű válaszokat kapunk, és nem kell összetett kombinációkat kezelni. Ezáltal az adatok könnyebben elemezhetők és a statisztikai feldolgozás is egyszerűbbé válik, miközben az eredmények következetesebbek és átláthatóbbak lesznek.

6. Példa

Rossz: „Milyen gyakran használja az egyetemi könyvtárat és az online adatbázisokat?”

Válaszlehetőségek: „Soha”, „Ritkán”, „Gyakran”, „Mindig” Ez a kérdés két különböző dolgot kérdez egyszerre, így nehéz meghatározni, hogy a válaszadó melyikre reagál (a könyvtárra vagy az adatbázisokra?). Ez zavart és hibás adatokat eredményezhet.

Jó: „Milyen gyakran használja az egyetemi könyvtárat egy héten belül?”

Válaszlehetőségek: „Soha”, „1–2 alkalommal hetente”, „3–4 alkalommal hetente”, „5–6 alkalommal hetente”, „Napi rendszerességgel.”

És külön kérdés az online adatbázisokra:

„Milyen gyakran használja az egyetemi online adatbázisokat egy héten belül?” Ez a változat világosan elkülöníti a két különböző erőforrást, így a válaszok konkrétak és jobban elemezhetők.

7. Példa

Rossz: „Mennyire találja hasznosnak a csoportmunkát a tanulmányai során?”

Válaszlehetőségek: „Egyáltalán nem”, „Kissé”, „Nagyon”, „Extrém módon” A válaszlehetőségek nem egyenletesen skálázottak, mivel a „Kissé” és „Nagyon” közötti átmenet nincs jól meghatározva, ami miatt a válaszadók számára nehéz lehet pontosan elhelyezni a válaszaikat. Ráadásul a „csoportmunka” különféle kontextusokban értelmezhető, ami szintén pontatlansághoz vezethet.

Jó: „Mennyire találja hasznosnak a csoportmunkát az egyetemi órákon?”

Válaszlehetőségek: „Egyáltalán nem hasznos”, „Kissé hasznos”, „Mérsékeltén hasznos”, „Nagyon hasznos”, „Rendkívül hasznos” A válaszlehetőségek fokozatosan skálázottak, lehetőséget adva a válaszadónak, hogy árnyaltabb véleményt adjanak. A kérdés pontosítja, hogy melyik környezetben (egyetemi órákon) mérjük a csoportmunkát, így jobban segít a kérdésre vonatkozó konkrét adatgyűjtésben.

8. Példa

Rossz: „Mennyire érzi, hogy a tanárok által nyújtott visszajelzések segítenek a fejlődésében?”

Válaszlehetőségek:

- „A visszajelzések ritkán hasznosak, mivel gyakran túl általánosak vagy nem elég részletesek ahhoz, hogy konkrét iránymutatást adjanak.”
- „A visszajelzések néha érthetők, de gyakran homályosak és nem mindig nyújtanak elegendő útmutatást a további fejlődéshez.”
- „A visszajelzések általában hasznosak, de néha nem elég részletesek ahhoz, hogy pontosan tudjam, mit kell tennem.”

– „A visszajelzések mindig teljesen érthetőek, részletesek, és minden esetben segítettek abban, hogy pontosan tudjam, miben kell fejlődnöm és hogyan tegyem meg.”

Ez a kérdés túlságosan részletes válaszokat ad, amelyek olvasása és értelmezése sok időt igényelhet a válaszadótól, ami fáradtságot és frusztrációt okozhat.

Jó: „Mennyire segítik a tanári visszajelzések a fejlődését?”

Válaszlehetőségek:

– „Egyáltalán nem segítenek”

– „Ritkán segítenek”

– „Többnyire segítenek”

– „Nagyon segítenek”

Ez a verzió rövidebb, világosabb válaszlehetőségeket biztosít, így a válaszadás gyorsabb és könnyebb, anélkül, hogy a válaszadó kifáradna.

HOGYAN LEHET ELŐZETESEN MEGBIZONYOSODNUNK A KÉRDŐÍVÜNK MEGBÍZHATÓSÁGÁRÓL ÉS ÉRVÉNYESSÉGÉRŐL?

A kérdőív megbízhatóságáról (*reliability*) és érvényességéről (*validity*) előzetesen megbizonyosodni kulcsfontosságú, hiszen a rosszul kialakított kérdések torz eredményekhez vezetnek. Ma már a tudományos publikációk bírálói is rendszeresen számonkérlik, hogy hogyan ellenőriztük a kérdőívünk minőségét. Itt van néhány szempont és módszer, amelyek segíthetnek biztosítani a kérdőív jó minőségét, de legalábbis meg szokta nyugtatni a bírálókat.

Ezek a módszerek első olvasásra akár gyermeknek, felesleges túlbonyolításnak is tűnhetnek, de nem így van. Sokat segíthetnek, olyan hibákat, félreérthetőségeket tisztázhatnak, amelyeket a kérdések összeállítója már azért nem vesz észre, mert neki bizonyos dolgok már nyilvánvalónak tűnnek, de egy válaszadónak nem.

Face validity (Első benyomás alapú érvényesség)

A **Face validity** azt vizsgálja, hogy a kérdések első ránézésre megfelelően mérik-e azt, amit mérni szeretnénk. A gyakorlati megvalósítása a következő lépéseken alapul:

- *Szakértői felülvizsgálat*: Szakértők (akik a vizsgált területhez értenek) átnézik a kérdőívet, és véleményt alkotnak arról, hogy a kérdések relevánsak és pontosak-e a kutatási cél szempontjából.
- *Laikus tesztelés*: Kérjünk meg néhány laikust, hogy tekintsék át a kérdéseket, és mondják el, szerintük azok világosak-e, és tükrözik-e a kívánt célt. Ez segít felmérni, hogy a kérdések érthetőek-e azok számára, akik a kérdőívet kitöltik majd.
- *Több körben történő finomítás*: A kapott visszajelzések alapján módosítsuk a kérdéseket, majd ismét végezzük el a tesztet, hogy lássuk, javult-e a megértés és a célhoz illeszkedés.

Content validity (Tartalmi érvényesség)

A Content Validity azt méri, hogy a kérdőív lefedi-e a teljes témát, amit a kutatási kérdések érintenek. Ennek gyakorlati megvalósítása a következőképpen történhet:

- *Szakértői panel összehívása*: Több szakértőt bevonva egy panelt alakítunk, amely áttekinti a kérdőívet. A szakértők minden kérdést külön értékelnek, és megvizsgálják, hogy azok lefedik-e a kutatási téma minden lényeges aspektusát.
- *Tartalmi mátrix készítése*: Készítsünk egy mátrixot, amelyben felsoroljuk a kutatási kérdéseket és a kérdőív kérdéseit. A mátrixban megjelöljük, hogy mely kérdések melyik kutatási aspektust fedik le, így biztosítva, hogy minden fontos terület érintve legyen.
- *Hiányosságok feltárása*: Ha bármely kutatási aspektus alul reprezentált, további kérdéseket kell hozzáadni, vagy átfogalmazni a meglévőket, hogy azok jobban illeszkedjenek a teljes tartalmi körhöz.

Pilot study (Próbavizsgálat)

A **Pilot study** egy kisebb mintán végzett teszt, amely feltárja a kérdőív hibáit és hiányosságait a megbízhatóság és az érvényesség szempontjából. Ennek gyakorlati megvalósítása:

- *Kisebbségi minta kiválasztása*: Válasszunk ki egy kisebb csoportot (pl. 10–20 fő), amely reményeink szerint reprezentálja a teljes célcsoportot, akik számára a kérdőívet terveztük.
- *Kérdőív kitöltetése*: Kérjük meg, hogy töltsék ki a kérdőívet. Fontos, hogy a résztvevők a kitöltés során jegyzeteket készíthessenek, ha valamit nehezen értenek vagy nem egyértelmű számukra.

- *Visszajelzések gyűjtése*: A kitöltés után kérdezzük meg a résztvevőket, hogy mely kérdéseket találták nehéznek vagy zavarosnak. Ez segít feltárni a kérdőív szerkezeti vagy megfogalmazási problémáit.
- *Javítás és újrapróbálás*: Az első próbavizsgálat eredményei alapján finomítsuk a kérdéseket, és ha szükséges, végezzünk egy újabb tesztelést, amíg a kérdőív megfelelően működik.

A kérdőív finomhangolása interjúkkal

A kérdőív kitöltését követően végzett **interjúk** segítenek megérteni, hogyan értették meg a válaszadók a kérdéseket. Ennek gyakorlati megvalósítása:

- *Problémás kérdések azonosítása*: Az interjúk során kiderülhet, ha egy kérdést másképp értelmeztek, mint ahogy azt mi terveztük. Ilyen esetekben átgondoljuk és módosítjuk a kérdést, hogy a jövőbeni válaszadók egyértelműen tudják értelmezni.

Konklúzió

A kérdőíves felmérések hatékonysága nagymértékben függ a kérdések helyes megfogalmazásától. Ahogy a cikk bemutatta, a kutatási kérdések pontossága és a kérdőív kérdéseinek világossága, specifikussága kulcsfontosságú a releváns és mérhető adatok gyűjtéséhez. A rosszul megfogalmazott kérdések, amelyek homályosak vagy túl általánosak, torz eredményekhez vezethetnek, és megnehezítik az elemzést. Ezen túlmenően a kérdőív validitásának és megbízhatóságának előzetes ellenőrzése – például szakértői felülvizsgálatok és próbavizsgálatok révén – biztosítja, hogy a kérdőív alkalmas legyen a kutatási kérdések megválaszolására. A megfelelő kérdésfeltevés alapjaiban határozza meg a kutatási folyamat sikerét, és segít megelőzni a gyakori hibákat, amelyek negatívan befolyásolhatják az eredményeket.

Irodalom

- Bryman, A. (2016): *Social research methods* (5th Ed.). Oxford: Oxford University Press.
- Cohen, L.–Manion, L.–Morrison, K. (2018): *Research methods in education* (8th Ed.). London: Routledge.
- Creswell, J. W. (2014): *Research design: Qualitative, quantitative, and mixed methods approaches* (4th Ed.). London: SAGE Publications.
- Csapó B. (2004): *Tudás és iskola*. Budapest: Műszaki Könyvkiadó.
- Dörnyei, Z. (2007): *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Falus I. (1998): *Bevezetés a pedagógiai kutatás módszereibe*. Budapest: Műszaki Könyvkiadó.

- Field, A. (2017): *Discovering statistics using IBM SPSS statistics* (5th Ed.). London: SAGE Publications.
- Molnár G. (2016): *Az oktatási technológia alapjai*. Budapest: Typotex Kiadó.
- Pataki Gy.–Halász G. (2010): *Kérdőíves kutatások tervezése és lebonyolítása az oktatáskutatásban*. Budapest: Oktatás-kutató Intézet.
- Varga J.–Kárpáti A. (2009): *A pedagógiai értékelés elmélete és gyakorlata*. Budapest: Műszaki Könyvkiadó.



Műveltséggel kapcsolatos kérdőívek élete a közösségi felületen és az így szerzett tapasztalatok

Összefoglalás: A közösségi felületeken bármely témában minden információ a nap 24 órájában támad minket. Jelen cikkben a közösségi médiában közreadott űrlapok életének bemutatása történik saját tapasztalatok alapján. Tárgyalásra kerül a kérdőíves kutatási módszer lényege, valamint a vizsgálat alá vont terület bemutatása. A szemléltetésre kerülő kérdőívek kitöltöttségének alakulása, és az abban történt változások okainak tárgyalása. A cikk végén a szerző kísérletet tesz a két kitöltés közötti különbségek magyarázatára, és az ebből nyert hasznos tapasztalatok ismertetésére, melyeket a legközelebbi hasonló kutatásnál érdemes betartani.

Kulcsszavak: Kitöltési szám, hivatkozás, like.

Abstract: Information on any topic bombards us on social media platforms 24 hours a day. This article presents an analysis of the life cycle of forms shared on social media, based on personal experiences. It discusses the essence of the questionnaire-based research method and provides an overview of the examined area. Furthermore, it explores the completion trends of the surveyed questionnaires and the reasons behind observed changes. In conclusion, the author attempts to explain the differences between two instances of data collection and outlines valuable insights that should be considered in future similar research endeavors.

Keywords: Fill number, reference, like.

Bevezetés

A 21. század embere napjainkban átlagosan minimum 5–6 órát tölt naponta a képernyő előtt, mely szám az évek során folyamatosan növekedett, és emelkedni is fog [1]. Az internet, televízió és egyéb digitális tartalmak a

* *Dunaiújvárosi Egyetem, Informatikai Intézet, Számítógéprendszerek és Irányítástechnika Tanszék; Eszterházy Károly Katolikus Egyetem, Neveléstudományi Doktori Iskola*

E-mail: farkasi@uniduna.hu

[1] Tessényi, J. (2024): Egyetemisták médiahasználati szokásainak vizsgálata=Examination of university students' media usage habits. *GRADUS*, 11., (1.).

[2] Alexa, C. (2020): Hangadók az ifjúság szociális életében. *Kulturális Szemle*, 95.

[2] Alexa, C. (2020): Hangadók az ifjúság szociális életében. *Kulturális Szemle*, 95.

[3] Molnár, D. (2010): Empirikus kutatási módszerek a szervezet-fejlesztésben. *Humán Innovációs Szemle*, 1., (1–2.), pp. 61–72.

különböző infokommunikációs eszközökön keresztül történő felhasználása mind azt segíti elő, hogy minél több időt töltsön a felhasználó magányosan az eszközzel kettesben. A gyorsan változó világ, a gyors információfogyasztást segíti elő.

Ez az állapot az emberekben kifejlesztette azt a képességet, hogy pár másodperc alatt (esetleg annak töredéke alatt) el tudja dönteni mi az, ami érdekli, és mi az, ami nem. A legtöbb időt a közösségi médiában töltik el kortársaink, hiszen ott tartják a kapcsolatot az ismerősökkel, rokonokkal. Ott szereznek be minden új pletykát, ezzel helyettesítve az utcai kispadokat. Ezen felületeknek komoly személyiségformáló hatása is van [2].

Jelen cikk két kérdőív életét mutatja be, melyek megjelenése között másfél év telt el. Az úrlapok témája mindkét esetben a műveltség és növelésének lehetséges módjai települési szinten. A cikk szerzője azt szemlélteti, hogy milyen tapasztalatokat szerzett ezen két kérdőív kitöltési ideje alatt, és milyen életet éltek a kitöltők a közösségi médiában (facebook-on). Levonható-e következtetés a kitöltöttség szempontjából a megjelenés napja között? Az aktivitásra milyen módon tud hatni a kérdőív készítője? A többszöri megosztásnak, felhívásnak van-e serkentő hatása?

A kérdőíves kutatás módszere

A kérdőív segítségével nem mélyreható összefüggések feltárása volt a cél (erre nem is alkalmas). A legnagyobb előnye a kérdőív-készítésnek, hogy segítségével rövid idő leforgása alatt nagy adatmennyiség gyűjthető össze, ami elősegíti a kutatás objektivitását, reprezentativitását és az eredmények általánosíthatóságát.

Ez köszönhető annak, hogy ma a közösségi médiában gyorsan közre lehet adni egy kérdőív linkjét, és ugyanilyen gyorsan begyűjthetőek az eredmények. A legnagyobb hátránya ennek a módszernek, hogy a kutatás mélyebb összefüggéseire, a vizsgálati alanyok egyéni jellegzetességeire nem képes rávilágítani. Egy pillanatkép a vizsgált populációról, adott körülmények között [3].

A kérdőív összeállításánál fontos arra ügyelni, hogy ne legyen túl hosszú és rövid sem az úrlap. Az előre jelzett kitöltési időt egy átlagos felhasználó se lépje túl nagy mértékben, mert akkor félbehagyja a kitöltést. Túl kevés kérdés esetén pedig felmerül a kitöltőben az, hogy mi szükség volt ezt ilyen formában megkérdezni.

A kérdések összeállításánál mindig szükség van pár (a kérdések feldolgozása szempontjából) releváns demográfiai adat begyűjtésére [4]. Ezt követhetik a lényegi kérdések, melyek a kutatás területére koncentrálnak. A vizsgálat szempontjából fontos kérdéseket, más szövegezéssel érdemes többször is feltenni, melynek célja az, hogy a kérdésre adott válaszok megbízhatóságát lehet vele ellenőrizni.

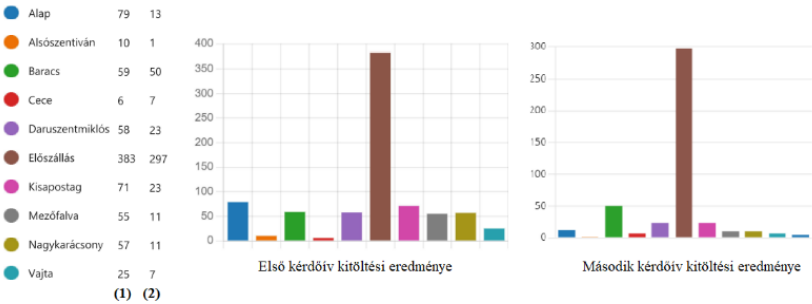
A kitöltés előtt fontos megemlíteni, hogy szigorúan anonim a kitöltés és az adatok csak a kutatáshoz kerülnek felhasználásra.

A kérdőívek által vizsgált terület

A vizsgálatban szereplő két kérdőív 10 község lakosai számára volt elérhető [4]. Ezen községek területét általában úgy szokták nevezni, hogy Dél-Fejér térsége. Ebbe a térségbe tartozik Baracs, Daruszentmiklós, Kisapostag, Mezőfalva, Nagykarcsony, Alsószentiván, Alap, Cece, Vajta és Előszállás települések. Ezen települések összes lakosságszáma 20 962 fő (2022. január elsejei KSH adatok alapján). A kérdőívet a települések polgármesterei osztották meg a községek közösségi oldalain, így jutott el minden településre. A kitöltés-szám természetesen a cikk írójának lakóhelye szerinti településen a legnagyobb.

[4] Farkas, I. (2023): Kérdőív összeállításának kérdései, az eredmények feldolgozásának lehetséges módszerei saját példán keresztül, In: *Informatika Korszerű Technikái 2023 „Tudomány: iránytű az élhető jövőhöz” nemzetközi tudományos konferencia: programfüzet és rövid cikkek*, Dunaújváros: DUE Press, pp. 190–201.

1. ábra. Kitöltések darabszáma



Az 1. ábrán jól látható, hogy Előszállás településen sokkal többen töltötték ki az űrlapot, mert a személyes ismertség ott volt a legnagyobb. A többi településen a kitöltési aktivitás ennek töredéke, de itt is megfigyelhető, hogy néhány helyen a 50–70 a kitöltési szám, míg máshol tíz alatti érték is előfordul.

Ez annak is köszönhető, hogy a település vezetője mennyire aktív a közösségi médiában, mennyire elfogadott személy a községben. A kitöltésből látszik, hogy a második kérdőívet majdnem csak fele anynyian töltötték ki, mint az első. Ezen cikk megírását ez a tény indukálta, hogy van-e ennek valamilyen jól kimutatható oka, vagy csak véletlen.

Az űrlapok élete

A kérdőívek mindkét alkalommal 13 napig voltak kitölthetők. Az első kérdőív megjelenése 2022. augusztus 18-i napon, mely egy csütörtök volt, a második megjelenése 2024. február 2-án egy pénteki napon történt. Az időpontok azért érdekesek, mert a kitöltési eredmény szempontjából meglepő, hogy melyik űrlapot hányan töltötték ki. A nyári időponttal kapcsolatban azt lehetne feltételezni, hogy viszonylag kevés számú kitöltés érkezne, hiszen ebben az időszakban a nyaralások még zajlanak. A februári hétfői időpont pedig azt jelezné előre, hogy itt több kitöltés érkezik, hiszen a munka végeztével a korai esti órákban a megkérdezettek jobban ráérnek és szívesebben tesznek eleget egy ilyen kitöltési kérésnek. Ahogy az 1. ábrán látható volt, ez nem így történt

Következtetés

- A két közreadott kérdőív kitöltésének sikerességével kapcsolatban az alábbi tapasztalatok vonhatók le:
- Pénteken nem szabad kérdőívet közre adni, mert a kitöltési hajlandóság nem megfelelő. Helyette inkább a hét eleji napok preferáltak (hétfő, kedd esetleg szerda).
 - Ha több településről is be kell gyűjteni kitöltéseket, akkor olyan személyeket kell megkérni a települési csoportban való megosztásra, akik rendszeres mozgatói a közösségi életnek (pl. közösségyszervező, művelődésház-vezető), és erre nem feltétlenül a polgármester a legalkalmasabb.
 - Kétszer hasonló témában ne kerüljön ki rövid időn belül kérdőív, mert az unalmassá, érdektelenné válhat.
 - Érdemes elgondolkozni több település esetén egy verseny meghirdetéséről a kitöltési mennyiség szempontjából (nem azonosak az esélyek, hiszen nagyobb lélekszámú településről több kitöltés érkezhets). Az egyenlőséget a laksok száma és a kitöltési szám aránya biztosíthatná. Természetesen ilyenkor gondoskodni kell valamilyen díjról is.
 - A motiváció a kitöltési hajlandóság növelésére. Ismertetni kell, hogy amennyiben kitöltik, milyen előnnyökkel járhat ez a kitöltő számára az életének későbbi szakaszában (pl. tartalmasabb közösségi rendezvények).

Linux-szerver automatikus telepítése on-premise környezetben

Összefoglalás: „Az automatizálás a banális és ismétlődő feladatok visszaszorulását eredményezi.” Rendszermérnökként gyakran találkozunk olyan feladattal, amelyet időről időre el kell végezni és a munkafolyamat előre meghatározott lépésekből áll. Azért, hogy az emberi hibát minimalizáljuk Linux operációs rendszer telepítést és konfigurálását is automatizálhatjuk. Jelen cikk azt mutatja be, hogy egy világszinten jelentős szolgáltató esetében milyen módon valósítható meg egy bérelt on-premise szerver telepítése és előkészítése további konfiguráció-menedzsment feladatok végrehajtására.
Kulcsszavak: Linux, szkript, automatikus telepítés.

Abstract: “Automation is driving the decline of banal and repetitive tasks.” As a systems engineer, we often encounter a task that needs to be performed from time to time, and the workflow consists of predefined steps. To minimize human error, we can automate the installation and configuration of Linux operating systems. This paper describes how to deploy and prepare a leased on-premise server for further configuration management tasks in case of a major global service provider.

Keywords: Linux, script, automatic installation.

* *Dunaiújvárosi Egyetem, Informatikai Intézet*
E-mail: hadarics@uniduna.hu

Bevezetés

Manapság sok olyan infrastruktúra-szolgáltató létezik, amely fizikai szerverek bérletét teszi lehetővé ügyfelek számára. Ezen szolgáltatás igénybevétele esetén a szolgáltató a hálózati infrastruktúrát és a hardver meghibásodások kezelését/cseréjét biztosítja az ügyfeleknek alapértelmezésként. A szerver számítógépre telepített szoftverek konfigurálása, üzemeltetése az ügyfél feladata.

[1] Hetzner Online GmbH – <https://www.hetzner.com>

[2] Installimage – <https://github.com/hetzneronline/installimage>

[3] Hetzner robot – <https://robot.hetzner.com>

[4] Ansible – <https://www.ansible.com>

Ezen publikáció esetében ismertetek egy saját megoldást, amely adott szerver-szolgáltató által biztosított lehetőségek figyelembevételével, automatizált Debian GNU/Linux-telepítés és -konfiguráció menedzsmentjének megoldására mutat be egy gyakorlati megoldást.

A szolgáltató sajátosságai

A Hetzner Online [1] Európa egyik legnagyobb adatközpont üzemeltetője. 1997 óta foglalkozik IT infrastruktúra-szolgáltatásokkal, szerverek elhelyezésével és bérbe adásával. Németországi adatközpontjai mellett Finnországban, USA-ban, Szingapúrban is rendelkezik telephelyekkel. Több százezer működő szervert szolgál ki az infrastruktúrája. Az ügyfeleket elsősorban az innovatív technológiákkal és a vonzó árakkal igyekszik magához csábítani, de kiemelten figyel a fenntarthatóságra is. Németországban 100%-ban a szükséges villamos energiát megújuló energiaforrás (vízenergia) segítségével állítja elő.

A Hetzner Online az ügyfelek részére egy speciális Linux netboot környezetet biztosít bérelt szerverek esetében. Az ún. Rescue-környezet segítségével az ügyfeleket többek között:

- operációs rendszereket telepíthetnek, az ún. Installimage [2] segítségével,
- hardver ellenőrzéseket futtathatnak,
- egy esetlegesen nem működő (nem bootoló) rendszer esetében adatmentést, helyreállítást végezhetnek Linux parancsok segítségével.

A szolgáltató emellett lehetővé teszi az ún. Robot Webservice API [3] használatával ügyféloldali műveletek automatikus elvégzését.

A kliensoldali környezet, és működési feltételek

A címbéli automatizált szervertelepítési megoldással kapcsolatban az alábbiak a célkitűzések:

- Input-adatként használja fel a szolgáltatótól kapott paramétereket.
- Végezze el előre meghatározott módon a szerver meghajtóinak particionálását és a szerver nevének beállítását.
- Kezelje a szerver újraindítását, és újraindítást követően folytatódjon a végrehajtása.
- Telepítse a választott konfigurációmenedzsment-megoldást (Ansible [4]), és tegye lehetővé az újonnan telepített szerver távoli menedzsmentjét.

- Az automatizált telepítést Linux asztali környezetből vagy egy korábban létrehozott szerverről lehet indítani, adott szkript végrehajtásával.

Az előzőek figyelembevételével, a működési feltételek az alábbiak:

- *openssh-client* – OpenSSH-kliens parancsok,
- *sshpas* – SSH-jelszóalapú hitelesítés használata.

A szolgáltató a sikeres bérletet követően az alábbi adatokat biztosítja az ügyfél részére (minta):

„Your server #server_id (192.168.108.1) has been activated and is now ready for use. Currently it is booted in the Rescue System. You can access the server via SSH2 using the following details:

IPv4 Address: 192.168.108.1

IPv6 Address: fe80::250:56ff:fec0:8

Username: root

Password: xDm8MieTXsU9ue

Host key: MYsjNcUoTgGEVHSUYfeJhRW9BOmh1m
HDFTLQaZ7bh/c (ECDSA 256)
k0pj5G4hiE7xBgf6325xfoicLUERmN8Kl5kLmVCNpno (ED25519 256)
Qvty3unE+4mAgTS2RFA7XcicBm/OPx88cXv/Fb3tHfQ (RSA 3072)

Az előző táblázatban alkalmazott IPv4 és IPv6 címek módosításra kerültek, és helyükre értelemszerűen a szolgáltatótól kapott publikus címeket kell elképzelnünk.

Az elkészített megoldás

A megoldás esetében egy input nevű fájlban rögzítjük a telepítendő szerverszolgáltatótól kapott releváns paramétereket és az új szerver nevét:

```
server="192.168.108.1"
```

```
hostname="debian12server"
```

```
sshuser="root"
```

```
export SSHPASS="xDm8MieTXsU9ue"
```

```
export TERM="xterm-256color"
sshtimeout=20
sshhostkeyalg=ECDSA
sshhostkeysize=256
sshhostkeytype="SHA${sshhostkeysize}"
sshhostkeyfingerprint="MYsjNcUoTgGEVHSUYfeJhRW9BOmh1mHDFTLQaZ7bh/c"
```

1. ábra. Forráskód első része

```
1 #!/usr/bin/env bash
2
3 . input
4
5 providedfingerprint=${sshhostkeysize} ${sshhostkeytype} ${sshhostkeyfingerprint} ${server} ${sshhostkeyalg})"
6 actualfingerprint=$(ssh-keyscan -4 -t ecdsa "${server}" 2>/dev/null | ssh-keygen -lf -)
7
8 if [ "$providedfingerprint" = "$actualfingerprint" ]; then
9
10 # Add new server to known hosts
11 ssh-keygen -F "${server}" &> /dev/null
12 ret=$?
13
14 if [ $ret -eq 1 ]; then
15     ssh-keyscan -4 -t ecdsa "${server}" >> ~/.ssh/known_hosts 2> /dev/null
16 fi
17
18 # Generate installimage configuration
19 installconfig="installimage.conf"
20 cat << EOT > "${installconfig}"
21 DRIVE1 /dev/nvme0n1
22 DRIVE2 /dev/nvme1n1
23 SWRAID 1
24 SWRAIDLEVEL 1
25 HOSTNAME ${hostname}
26 USE_KERNEL_MODE_SETTING yes
27 PART /boot/efi esp 250M
28 PART swap swap 64G
29 PART /boot ext3 1024M
30 PART / ext4 all
31 IMAGE /root/.oldroot/nfs/install/.../images/Debian-1207-bookworm-amd64-base.tar.gz
32 EOT
33
34 # Upload installer config
35 command="cat > /${installconfig}"
36 cat "${installconfig}" | sshpass -e ssh -o 'StrictHostKeyChecking=yes' -o 'ConnectTimeout=${sshtimeout}' ${sshuser}@${server} "${command}"
37
38 # Execute installer
39 installimage="/root/.oldroot/nfs/install/installimage"
40 command="bash ${installimage} -o < /${installconfig}"
41 result=$(sshpass -e ssh -t -o 'StrictHostKeyChecking=yes' -o 'ConnectTimeout=${sshtimeout}' ${sshuser}@${server} "${command}")
42 echo $result
43
44 # Reboot the server
45 command="reboot"
46 result=$(sshpass -e ssh -t -o 'StrictHostKeyChecking=yes' -o 'ConnectTimeout=${sshtimeout}' ${sshuser}@${server} "${command}")
47 echo $result
48
49 # Remove old server from known hosts
50 ssh-keygen -f ~/.ssh/known_hosts -R "${server}"
```

Forrás: Saját ábra.

A szkript az alábbiakat végzi el:

- Beolvassa az input fájlbeli paramétereket.

- Ellenőrzi, hogy a szolgáltató által küldött SSH ECDSA 256 kulcs lenyomata egyezik-e azzal, amit az aktuális kapcsolódáskor kapott a szkript. Amennyiben nem, akkor a szkript az ujjenyomatok értékeivel fog visszatérni.
- Amennyiben egyezik az ujjenyomat, akkor az alábbiak kerülnek végrehajtásra:

2. ábra. Forráskód második része

```
debian12_install_publ_ok.sh - hetzner_installer - Visual Studio Code

$ debian12_install_publ_ok.sh

51
52 # Check server rebooted
53 command="true"
54 until sshpass -e ssh -o 'StrictHostKeyChecking=no' -o 'ConnectTimeout=${sshtimeout}' ${sshuser}@${server} "${command}" 2>/dev/null;
55 do
56     sleep 30
57 done
58
59 # Ansible installer
60 ansible_installer="ansible_install.sh"
61 cat << 'EOT' > "${ansible_installer}"
62 #!/bin/bash
63 # Task:
64 # - setup SSH and ansible user on a new Debian 12.x server
65 # - install ansible
66 #
67 # Author: Kálmán Hadarics <kalman@hadarics.hu>
68 # -----
69
70 # Constants
71 admin_group="admins"
72 admin_gid=200
73 ansible_user="ansibleadmin"
74 ansible_user_home="/home/${ansible_user}"
75
76 # Setup SSH port and IPv4 only
77 sed -i \
78 s/"#Port 22/Port 32025/ \
79 s/"#AddressFamily any/AddressFamily inet/ \
80 ' /etc/ssh/sshd_config
81
82 systemctl restart ssh.service
83
84 # Add admin group and ansibleadmin user
85 groupadd --gid $admin_gid $admin_group
86 useradd --create-home --home-dir $ansible_user_home --gid $admin_gid $ansible_user --shell /bin/bash
87
88 # Upgrade system
89 apt-get update
90 apt-get -y install firmware-realtek
91 apt-get -y install firmware-amd-graphics
92 apt-get -y dselect-upgrade
93
94 # Install sudo
95 apt-get install sudo
96
97 # Add ansible user sudo
98 echo "${ansible_user} ALL=(ALL) NOPASSWD: ALL" > /etc/sudoers.d/${ansible_user}
99
100 # Generate and caten SSH-key
```

Forrás: Saját ábra.

- Az új szerver hozzáadása a kliens `.ssh/known_hosts` fájljához.
- Az `installimage` [2] konfigurációjának (`installimage.conf`) dinamikus létrehozása.
- Az `installimage.conf` fájl feltöltése SSH-kliens segítségével.
- Az `installimage` parancs távoli végrehajtása SSH segítségével a szerveren. Ez elvégzi a konfiguráció alapján a telepítést és a hálózat alapbeállításait.

- A telepített szerver újraindítása.
- A korábbi tárolt ujjlenyomatok törlése a `.ssh/known_hosts` fájlból. A szerver újrategelítése új SSH-kulcsokat hoz létre, így értelemszerűen az ujjlenyomatok is változnak.
- Várákozás az szerver újraindulására, 30 másodpercenként csatlakozási kísérlet.
- Az `ansible_installer.sh` szkript dinamikus létrehozása. Ez a szkript az alábbiakat végzi:
 - admins csoport és rendszer felhasználó (`ansible_admin`) létrehozása (`$ansible_user` változó tartalma)
 - az alapértelmezett OpenSSH-port módosítása (biztonsági megfontolásból);
 - a feltelepített rendszer frissítése,
 - sudo tegelítése és `ansible_admin` sudo privilégiumának konfigurációja,
 - SSH-kulcspár létrehozása az `ansible_admin` felhasználó esetében.
 - az ansible csomag tegelítése, amely lehetővé teszi a távoli adminisztrációt SSH felhasználásával.
- Az `ansible_installer.sh` szkript távoli végrehajtása SSH segítségével a szerveren.
- Az `ansible_admin` számára generált kulcspár privát kulcsának letöltése a helyi rendszerbe, majd a távoli szerverről a privát kulcs törlése.
- Ansible inventory fájl létrehozása, amely felhasználásával a távoli szerver Ansible segítségével távolról vezérelhető lesz.

3. ábra. Forráskód harmadik része

```

100 # Generate and setup SSH-key
101 ssh_dir="${ansible_user_home}/ssh"
102 ssh_key_private="${ssh_dir}/${(hostname) id_rsa}"
103 ssh_key_public="${ssh_dir}/${(hostname) id_rsa.pub}"
104 ssh_key_comment="${(hostname) id_rsa}"
105 ssh_auth_keyfile="${ssh_dir}/authorized_keys"
106
107 sudo -u ${ansible_user} mkdir -m 700 $ssh_dir
108 sudo -u ${ansible_user} echo -e y\N | sudo -u ${ansible_user} ssh-keygen -t rsa -b 4096 -C "${ssh_key_comment}" -f "${ssh_key_private}" -N ""
109 sudo -u ${ansible_user} cp "${ssh_key_public}" "${ssh_auth_keyfile}"
110 sudo -u ${ansible_user} chmod 600 "${ssh_auth_keyfile}"
111
112 # Install ansible
113 apt-get -y install ansible
114 EOF
115
116 # Upload Ansible installer
117 command="cat > /root/${ansible_installer}"
118 cat "${ansible_installer}" | sshpass -e ssh -o 'StrictHostKeyChecking=no' -o 'ConnectTimeout=${ssh_timeout}' ${sshuser}@${server} "${command}"
119
120 # Execute Ansible installer
121 command="bash /root/${ansible_installer}"
122 result=$(sshpass -e ssh -t -o 'StrictHostKeyChecking=no' -o 'ConnectTimeout=${ssh_timeout}' ${sshuser}@${server} "${command}")
123 echo $result
124
125 # New ssh port
126 ssh_port=32025
127
128 # Get generated Ansible admin private key
129 ansible_user="ansibleadmin"
130 ansible_user_home="/home/${ansible_user}"
131 ssh_dir="${ansible_user_home}/ssh"
132 ssh_key_private_file="${(hostname) id_rsa}"
133 ssh_key_private_path="${ssh_dir}/${ssh_key_private_file}"
134
135 command="cat $ssh_key_private_path"
136 sshpass -e ssh -o "Port=${ssh_port}" -o 'StrictHostKeyChecking=no' -o 'ConnectTimeout=${ssh_timeout}' ${sshuser}@${server} "${command}" > "${ssh_key_private_file}"
137
138 # Delete remote Ansible admin private key
139 command="rm $ssh_key_private_path"
140 sshpass -e ssh -o "Port=${ssh_port}" -o 'StrictHostKeyChecking=no' -o 'ConnectTimeout=${ssh_timeout}' ${sshuser}@${server} "${command}"
141
142 # Generate Ansible inventory related to host
143 inventory="inventory ${hostname}"
144 cat << EOF > "${inventory}"
145
146     ${hostname}:
147         ansible_host: ${server}
148         ansible_ssh_private_key_file: /home/kami/ansible_project/keys/${ssh_key_private_file}
149 EOF
150

```

Forrás: Saját ábra.

Összegzés

Aholgyakori üzemeltetési feladat Linux-szerverek telepítése, mindenképpen gondolkodni kell azon, hogy milyen módon valósítható ez meg automatizált módon. On-premise környezetben az automatizált telepítés megoldása nagyban függ a szolgáltató által biztosított feltételektől és lehetőségektől. Némi szkript-írási és integrálási tapasztalatokkal elkészíthető olyan megoldás, amely rugalmasan felhasználható automatikus telepítés kivitelezésére.

A bemutatott megoldás képes arra, hogy az újonnan telepített rendszer Ansible által vezérelhető legyen, a használt Ansible-inventory kibővítésével.



Adatbázistrendek és technológiák: Hagyományos és modern megközelítések

Összefoglalás: Az adatbáziskezelési technológiák fejlődése az elmúlt évtizedekben forradalmasította az adattárolás és -kezelés módszereit. A hagyományos SQL-alapú relációs adatbázisok évtizedek óta dominálnak, de a megváltozott adatigények és a Big Data korában megjelenő új kihívások miatt a NoSQL-adatbázisok egyre nagyobb népszerűségnek örvendenek. A cikk áttekinti az SQL- és NoSQL-adatbázisok közötti főbb különbségeket, különös figyelmet fordítva az ACID- és a CAP-elvekre, amelyek meghatározzák a különböző adatbázisok teljesítményét és megbízhatóságát. Bemutatásra kerülnek a hibrid adatbázisok, amelyek a két technológia előnyeit ötvözik, új megoldásokat kínálnak, alkalmazva a relációs modellek struktúráját és a NoSQL rugalmasságát. A hibrid rendszerek nemcsak költséghatékony megoldásokat jelentenek, hanem lehetőséget adnak a különböző adattípusok integrált kezelésére is. A cikk végül bemutatja az egyes adatbázistípusok leggyakoribb alkalmazási területeit, hangsúlyozva, hogy a megfelelő megoldás kiválasztása az adott üzleti igények és kihívások alapján történik.

Kulcsszavak: SQL, NoSQL, relációs adatbázis, hibrid adatbázis, dokumentum-alapú adatbázis, gráfolapú adatbázis, kulcsérték, ACID, CAP.

Abstract: The development of database management technologies in recent decades has revolutionized the methods of data storage and management. Traditional SQL-based relational databases have dominated for decades, but due to changed data requirements and new challenges in the era of Big Data, NoSQL databases are gaining more and more popularity. The article reviews the main differences between SQL and NoSQL databases, paying particular attention to the ACID and CAP principles that determine the performance and reliability of the different databases. Hybrid databases are presented, which combine the advantages of the two technologies and

* Dunaújvárosi Egyetem, Informatikai Intézet, Szoftverfejlesztési és Alkalmazási Tanszék
E-mail: dudasn@uniduna.hu

offer new solutions, applying the structure of relational models and the flexibility of NoSQL. Hybrid systems are not only cost-effective solutions, but also offer the possibility of integrated management of different data types. Finally, the article presents the most common application areas of each database type, emphasizing that the right solution is selected based on the specific business needs and challenges. **Keywords:** SQL, NoSQL, relational database, hybrid database, document-based database, graph-based database, key-value, ACID, CAP.

Bevezető

Az elmúlt évtizedekben az adatbáziskezelési technológiák jelentős fejlődésen mentek keresztül. Míg az SQL-alapú (Structured Query Language) relációs adatbázisok hosszú ideig uralták az adattárolás világát, az utóbbi években a NoSQL-adatbázisok (Not Only SQL) is egyre népszerűbbé váltak a dinamikus, nagy adatmennyiséggel dolgozó rendszerek igényeinek kielégítésére.

Az SQL- és NoSQL-adatbázisok közötti választás sokáig kompromisszumot jelentett a struktúra, rugalmasság, skálázhatóság és teljesítmény terén.

Azonban a technológiai fejlődés lehetővé tette az olyan hibrid adatbázisok megjelenését, amelyek egyesítik az SQL- és NoSQL-rendszerek előnyeit. Az olyan adatbázisok, mint a CockroachDB és az ArangoDB, mind a relációs, mind a nem-relációs adatok kezelésére alkalmasak, így lehetővé teszik, hogy a felhasználók a legjobbat kapják mindkét technológiából.

A cikk célja, hogy bemutassa a hagyományos SQL- és a NoSQL-alapú adatbázisok közötti főbb különbségeket, illetve ismertesse az új hibrid adatbázisok adta előnyöket és kihívásokat. Valamint rávilágítson arra, hogy milyen helyeken érdemes alkalmazni ezeket a megoldásokat.

Relációs adatbázisok

A relációs adatbázisok (RDBMS) az 1970-es években váltak népszerűvé, miután Edgar Frank „Ted” Codd kidolgozta a relációs modell alapjait. Ted Codd új elgondolása matematikailag megalapozott eszközöket használ, ezáltal pontosabb leírást tesz lehetővé. Megalkotta a relációs algebrát, mely a relációkon végezhető műveleteket definiálja. Fő gondolata az volt, hogy az adatokat kétdimenziós táblázatok (relációk) formájában kell tárolni, így az egymással logikai kapcsolatban lévő táblázatok rendszere alkotja majd magát az adatbázist. Tehát amíg a korai modellek – a hálós és a hierarchikus modell – megvalósításában az egyes rekordok között fizikai kapcsolat van (mutatókon keresztül), addig a táblázatok közötti kapcsolat logikai.

Természetesen akkoriban – ahogy ma is – a szakmai világ kétkedve fogadta az újdonságokat. Az IBM eleinte nem is alkalmazta Codd kutatási eredményeit, nem fogadták el az ötleteit, azaz az adatbáziskezelés relációs elvű megközelítését. Minimum 10 évébe került, hogy meggyőzze a szakmai közvéleményt. Pedig az általa lefektetett alapelvek még a mai napig érvényesek, és szinte egyeduralkodóvá vált a relációs elvű adatbáziskezelés. Nem igazán terjedt el a más szemléletű adatbáziskezelés, kivéve a legutóbbi időben.

NoSQL-adatbázisok

A NoSQL-adatbázisok (Not Only SQL) olyan adatbázis-technológiák, amelyek a relációs adatmodellek alternatívájaként jöttek létre a 2000-es évek elején. A NoSQL-adatbázisok elsősorban nagy méretű, gyorsan változó, és nem strukturált adatok kezelésére specializálódtak.

Ezek az adatbázisok nem kötődnek szigorúan strukturált séma követelményekhez, és nagyobb rugalmasságot kínálnak az adattárolás és kezelés terén is.

A NoSQL-adatbázisok négy fő kategóriába sorolhatók:

- Dokumentum-alapú adatbázisok, ahol az adatok dokumentumok formájában tárolódnak (pl. JSON, BSON formátumban).
- Kulcsérték tárolók, amelyek egyszerű kulcsérték párok alapján tárolják az adatokat.
- Oszlopalapú adatbázisok, amelyek táblázatos formában, de oszlop-orientáltan tárolják az adatokat.
- Gráfalapú adatbázisok, amelyek az adatok közötti kapcsolatokat gráfként kezelik. Az adatokat csomópontok és élek formájában tárolják, ahol a csomópontok entitásokat képviselnek, az élek pedig az ezek közötti kapcsolatokat jelölik.

[1] Buza A. (2015):
Az adatbáziskezelés
alapjai. Dunaújváros:
Rolling.

1. táblázat. NoSQL-alapú adatbázistípusok

Adatbázis típus	Adatbáziskezelő	Alkalmazási területek
Dokumen- tum-alapú	– MongoDB – CouchDB – RavenDB	Webalkalmazások, E-kereskedelem (pl: termékín- formációk), Mobilalkalmazások, Keresőmotorok, Tartalomkezelő rendszerek (CMS)
Kulcsértéktároló	– Redis – Amazon Dyna- moDB – Riak – Firebase Realtime Database	Gyorsítótárak (pl. webes alkalmazások gyorsítása), Munkamenet-kezelés (pl. felhasználói munkamene- tek tárolása), Valós idejű adatok (pl. hőmérsékleti szenzoradatok) feldolgozása, Konfigurációs beállítások kezelése
Oszlopalapú	– Apache Cassand- ra, – HBase – Google Bigtable	Big Data elemzések (pl.: Webes elemzések), Napló- fájlok feldolgozása (pl.: szerverlogok) Valós idejű analitika (pl.: pénzügyi tranzakciók elemzése)
Gráfalapú	– Neo4j	Szociális hálózatok (pl. kapcsolatok és interakciók modellezése), Hálózatelemzés (pl. kiberbiztonsági elemzés), Ajánló rendszerek (pl. termék- vagy tartalomajánlás)

Az 1. táblázatban látható néhány adatbáziskezelő, és a főbb alkalmazási terü-
leteik. Elmondható, hogy NoSQL-adatbázisok rugalmas adatmodellt használnak,
amely lehetővé teszi az adatok strukturálatlan (pl.: különféle bináris fájlok: képek,
hangfájlok, videók, ...) vagy félig strukturált (pl.: JSON formátumú dokumentu-
mok) formában történő tárolását. De akár a strukturált adatok (oszlopalapú ada-
tok) tárolására is alkalmasak. Az egyik legfontosabb jellemzőjük a sémamentes-
ség, amely azt jelenti, hogy az adatok szerkezetének előzetes meghatározása nem
szükséges, és különböző típusú adatok tárolhatók ugyanabban az adatbázisban. [1]

A NoSQL-rendszerek általában elosztott architektúrát használnak, amely le-
hetővé teszi az adatok több szerveren vagy adatközpontban történő tárolását és
elosztását.

Ez növeli a skálázhatóságot, és lehetővé teszi a nagy adatmennyiségekkel (Big Data) történő hatékony munkavégzést. Például a webes alkalmazások, a közösségi média és az online szolgáltatások mind-mind hatalmas adatmennyiséget generálnak, melyet a relációs adatbázisok már nem tudnak hatékonyan kezelni.

[2] Ullman Widom–Jennifer Widom (Szerk.) (1998): *Adatbázis-rendszerek alapvetés*, Budapest: Panem.

Hagyományos SQL- és NoSQL-adatbázisok főbb különbségei

Az SQL-adatbázisok, mint például a MySQL, PostgreSQL, és az Oracle Database, a relációs modellre épülnek, és olyan környezetekben használjuk őket, ahol az adatok szerkezete jól meghatározott, és ahol a tranzakciós integritás kritikus.

A relációs adatbázisokat az SQL (Structured Query Language) nyelven keresztül kezeljük, amely lehetővé teszi az adatok hatékony lekérdezését, manipulálását és karbantartását. Az SQL támogatja a tranzakciók ACID (Atomicity, Consistency, Isolation, Durability) tulajdonságait, amelyek biztosítják a megbízható és konzisztens adatkezelést. Azaz egy vagy több adatbázisműveletet egy tranzakcióba csoportosíthatunk, ez lesz egy olyan munkaegység, amit atomosan (oszthatatlan módon) és más tranzakciótól látszólag elkülönítve (elzártan, rejtve) kell végrehajtani. Miközben minden művelet végrehajtása során rendszerünk egy konzisztens állapotból egy konzisztens állapotba kerül. Valamint az adatbáziskezelő a tartósságot is garantálja, mely során befejezett, sikeres tranzakciók eredményei nem veszhetnek el. Ha egy tranzakció sikertelen, akkor a kezdeti állapotnak kell megmaradnia. [2]

A NoSQL-adatbázisok nem kötődnek a relációs modellhez, és nem igényelnek előre meghatározott sémát. Ez lehetővé teszi számukra, hogy sokkal rugalmasabban kezeljék a strukturálatlan vagy félig strukturált adatokat. Emellett a NoSQL-rendszerek nagy előnye a horizontális skálázhatóság, amely lehetővé teszi, hogy a rendszerek nagy mennyiségű adatot kezeljenek elosztott architektúrákon keresztül.

A NoSQL-adatbázisok leggyakrabban a CAP-tétel (Consistency, Availability, Partition tolerance) alapján működnek, és inkább a rendelkezésre állást és a partíciótoleranciát preferálják, szemben a teljes adatkonzisztenciával, azaz kompromisszumot kötnek az adatok konzisztenciájában. Ez lehetővé teszi az adatok gyors hozzáférését, még akkor is, ha a rendszer egyes részei nem elérhetők.

[3] <https://db.bme.hu/~gajdos/2012adatb2/3.%20eloadas%20NoSQL%20adat-b%20E1zisok%20doc.pdf>,” [Online]. Available: <https://db.bme.hu/~gajdos/2012adatb2/3.%20eloadas%20NoSQL%20adatb%20E1zisok%20doc.pdf>. [Hozzáférés dátuma: 2024. október 17.]

Maga a CAP-tétel elosztott adatbázisokra vonatkozik, és arra mutat rá, hogy egy elosztott rendszerben a konzisztencia (Consistency), rendelkezésre állás (Availability) és partíciótürés (Partition tolerance) három tulajdonságából egyszerre legfeljebb kettőt lehet biztosítani. [3]

- 1. Konzisztencia:** Az adatok minden lekérdezéskor ugyanazok, tehát a frissítések azonnal láthatók a teljes rendszerben.
- 2. Rendelkezésre állás:** A rendszer minden lekérdezésre választ ad, még akkor is, ha nem tudja garantálni a legfrissebb adatokat.
- 3. Partíciótürés:** A rendszer képes tovább működni akkor is, ha hálózati szakadás történik, és a csomópontok közötti kommunikáció megszakad.

Skálázhatóság

A relációs adatbázisok hagyományosan vertikálisan (függőlegesen) skálázhatók, ami azt jelenti, hogy a teljesítmény növelése érdekében a szerver hardveres kapacitását kell növelni (pl. több memória, gyorsabb CPU). Ez a megközelítés azonban költséges és gyakran nehéz, különösen, ha az adatbázis mérete és a felhasználói terhelése gyorsan növekszik.

A relációs adatbáziskezelő rendszerek jól teljesítenek kisebb adatmennyiségeknél és egyszerű tranzakcióknál, de problémák merülhetnek fel, amikor a rendszer terhelése növekszik. Az ACID-tulajdonságok fenntartása nagy adathalmazok esetén lassíthatja a rendszert, mivel az adatok konzisztenciáját és integritását minden egyes tranzakció során biztosítani kell.

A NoSQL-adatbázisok egyik legnagyobb előnye a horizontális (vízszintes) skálázhatóság, ami azt jelenti, hogy a teljesítmény és a tárhelykapacitás növelése érdekében további szerverek adhatók a rendszerhez. Ez lehetővé teszi, hogy a NoSQL-adatbázisok könnyen kezeljenek hatalmas adatmennyiségeket és nagy terhelést, anélkül, hogy drága hardverfrissítésekre, nagyobb szerverek beszerzésére lenne szükség. Ezáltal lehetővé vált az olcsóbb hardvereken való elosztott adatkezelés.

Hibrid adatbázisok: Az SQL és NoSQL integrációja

A hagyományos SQL-alapú és NoSQL-rendszerek különböző célokat szolgálnak, és gyakran kiegészítik egymást a különböző alkalmazási területeken. A hibrid adatbázisok olyan megoldások, amelyek az SQL- és NoSQL-adatbázis-technológiák egyesítésével működnek, kihasználva és kombinálva a két adatkezelési módszer előnyeit. A hibrid adatbázisok kialakulása a 2010-es évek környékén kezdődött, amikor a vállalatok egyre nagyobb adatkezelési kihívásokkal szembesültek. A relációs adatbázisok már nem voltak elegendők a Big Data korszakában jelentkező, skálázhatósági és teljesítményigények kielégítésére. Azonban az SQL-alapú adatbázisok robusztus tranzakciókezelése és adatkonzisztenciája továbbra is nélkülözhetetlen volt. Ezért alakultak ki olyan adatbázisok, amelyek egyesítik a hagyományos relációs rendszerek és a NoSQL rugalmasságát.

ELŐNYÖK ÉS HÁTRÁNYOK

Ha röviden kellene összefoglalni a hibrid rendszerek előnyeit, akkor ezek a következők lennének:

1. *Rugalmasság:* A hibrid adatbázisok képesek strukturált – relációs adatbázisokban megszokott – és nem strukturált vagy félig strukturált – NoSQL-alapú – adatokat is kezelni. Így egy rendszer képes különböző típusú adatok feldolgozására és tárolására.
2. *Skálázhatóság és teljesítmény:* A NoSQL-technológia biztosítja a hibrid adatbázisok számára a horizontális skálázást (több csomópont hozzáadásával), ami lehetővé teszi a nagy méretű adatforgalom kezelését.
3. *Tranzakciók és konzisztencia:* A hagyományos SQL-alapú adatbázisok által kínált ACID-tulajdonságok a hibrid rendszerekben is elérhetőek, így biztosított a biztonságos és megbízható tranzakciókezelés.
4. *Egyszerű integráció:* A hibrid rendszerek lehetővé teszik, hogy a különböző típusú adatok (strukturált és nem strukturált) egyetlen platformon keresztül legyenek elérhetőek és elemezhetőek, ami egyszerűíti a fejlesztést és az adatkezelést.
5. *Költséghatékonyság:* A hibrid megoldások csökkenthetik az adatkezelési költségeket, mivel egy platformon lehet integrálni a különféle adatokat, ami megszünteti az igényt több rendszer fenntartására.

Viszont akadnak a hibrid rendszereknek hátrányai is:

1. *Komplexitás:* A hibrid adatbázisok kezelése és karbantartása bonyolultabb lehet, mint a tisztán SQL- vagy NoSQL-rendszereké, mivel mindkét technológiát integrálni kell. Ez növeli a fejlesztés és karbantartás költségeit.

2. *Optimalizálás nehézségei*: Az eltérő adattípusok és használati esetek miatt a hibrid rendszerek optimalizálása kihívást jelenthet, különösen akkor, ha az adatbázisnak egyszerre kell kezelnie nagyméretű strukturált és nem strukturált adatokat.
3. *Magasabb belépési küszöb*: A fejlesztőknek és rendszergazdáknak mind SQL-, mind NoSQL-technológiák ismeretére is szükségük van, ami növeli az oktatási és szakértelmi követelményeket.
4. *Teljesítményproblémák*: A hibrid rendszerek esetenként kompromisszumokat kötnek a teljesítmény terén, ha mindkét típusú adatkezelési technológia előnyeit akarják kihasználni.

Alkalmazási területek adatbázis típusonként

A relációs adatbázisok tipikus alkalmazási területei:

- *Pénzügyi rendszerek*: ahol az adatkonzisztencia és a tranzakciók integritása kritikus.
- *Üzleti alkalmazások*: pl. ERP-rendszerek (Enterprise Resource Planning), ahol a tárolt adatok struktúrája szigorúan szabályozott.
- *Kis- és középállalkozások webes alkalmazásai*: kisebb adatbázisok, ahol a relációs modell jól kezelhető.

A NoSQL-adatbázisok leggyakoribb alkalmazási területei:

- *Big Data rendszerek*: ahol hatalmas adatokat kell tárolni és gyorsan elérni (pl. social media platformok, streaming szolgáltatások).
- *IoT (Internet of Things)*: ahol az eszközök folyamatosan hatalmas mennyiségű, különböző típusú adatokat generálnak.
- *E-kereskedelem és felhasználóiinterakció-alapú rendszerek*: ahol a felhasználói adatok nagy mennyiségben és dinamikusan változnak (pl. Amazon, eBay).

A hibrid adatbázisok leggyakoribb alkalmazási területei:

- *E-kereskedelmi platformok*: A termékinformációk relációs adatok, míg az ügyfélértékelések és felhasználói tartalmak félig vagy nem strukturáltak.
- *IoT-rendszerek*: Nagy mennyiségű, változó struktúrájú érzékelőadatok és rögzített, strukturált metainformációk kezelése.
- *Adatvezérelt marketing rendszerek*: Strukturált ügyféladatok és nem strukturált interakciós adatok kombinálása valós idejű analitika és személyre szabott ajánlatok létrehozása céljából.

2. táblázat. Néhány hibrid adatbáziskezelő

Adatbáziskezelő	Típus	Alkalmazási Területek
ArangoDB [4]	Multimodel-adatbázis (dokumentum, kulcsérték, gráf)	Szociális hálózatok, összetett elemzések, webalkalmazások.
Couchbase [5]	Dokumentumalapú és kulcsérték tároló. Képes SQL-alapú lekérdezések futtatására is.	Mobilalkalmazások, e-kereskedelem.
Microsoft Azure Cosmos DB [6]	Globális elosztású multimodel adatbázis (dokumentum, kulcsérték, gráf)	Nagy léptékű webalkalmazások, IoT-alkalmazások, Big Data-elemzések.
OrientDB [7]	Multimodel-adatbázis (dokumentum, gráf, reláció)	Hálózatelemzés, szociális hálózatok, összetett adatkapcsolatok modellezése.

Összegzés

A relációs, NoSQL- és a hibrid adatbáziskezelők mind jelentős szerepet játszanak az adatok kezelésében, de különböző igényekhez és alkalmazási területekhez illeszkednek. Relációs-adatbázisok klaszikus, jól bevált megoldások, amelyek szigorú adatmodellezést és tranzakcióbiztonságot kínálnak az ACID-tulajdonságok révén.

Ideálisak olyan alkalmazásokhoz, ahol a strukturált adatok és a magas fokú adatkonzisztencia elengedhetetlenek, mint például pénzügyi rendszerek vagy vállalati erőforrás-tervezési (ERP) megoldások.

A NoSQL-adatbázisok viszont rugalmasabb megközelítést kínálnak, lehetővé téve a strukturált, félig strukturált és nem strukturált adatok hatékony kezelését. Az horizontálisan skálázható NoSQL-rendszerek kiválóan alkalmasak a Big Data-kihívások kezelésére, mint például a közösségi média, valós idejű analitika és IoT-alkalmazások, ahol a sebesség és a rugalmasság kulcsszerepet játszik.

A hibrid adatbázisok az SQL és NoSQL előnyeit ötvözik, lehetővé téve a felhasználók számára, hogy egyszerre kezeljenek különböző típusú adatokat. Ezek a rendszerek ideálisak a modern, összetett alkalmazások számára, amelyek változatos adatkezelési igényekkel rendelkeznek. Rugalmas adatmodellezésük és a tranzakciók biztonságának megőrzése segít a különböző adatok hatékony integrálásában.

Mindezek a megoldások különböző kihívásokat és előnyöket kínálnak, és a választásuk mindig az adott alkalmazás igényeitől, a tervezett adatmennyiségtől, a szükséges skálázhatóságtól és a teljesítmény elvárásoktól függ. A jövőben valószínű, hogy a különböző adatbáziskezelő rendszerek további fejlődésen mennek keresztül, reagálva a folyamatosan változó adatmennyiségi és kezelési igényekre. Az optimális megoldás kiválasztása érdekében a fejlesztőknek alaposan meg kell érteniük a rendelkezésre álló lehetőségeket, és figyelembe kell venniük a specifikus követelményeket és környezeti tényezőket.



A tanulóelemzés fejlődése és nemzetközi alkalmazása az oktatásban

Összefoglalás: A tanulóelemzés avagy tanulási analitika (Learning Analytics: LA) fogalma és gyakorlata technológiaalapúsága miatt viszonylag rövid múltra visszatekintő, interdiszciplináris kutatási terület, ami alapvetően az elektronikus tanulási környezet és annak relevanciájában a tanulás megértésével, optimalizálásával foglalkozik. Az Európai Bizottság magyar „Oktatás és képzés 2020” munkacsoportjának véleménye szerint „komoly potenciállal rendelkezik az oktatás és a tanulás minőségének javítása szempontjából”. [1] A tanulóelemzés egyszerre akadémiai terület és kereskedelmi piac, amely az elmúlt évtizedben gyorsan formálódott. Kutatási és oktatási területként a tanulás (pl. oktatáskutatás, tanulási és értékelési tudományok, oktatástechnológia), az analitika (pl. statisztika, vizualizáció, számítástechnika/adattudomány, mesterséges intelligencia) és az emberközpontú tervezés (pl. használhatóság, részvételen alapuló tervezés, szociotechnikai rendszerekben való gondolkodás) konvergenciáján helyezkedik el. Történetileg a tanulóelemzés legáltalánosabb alkalmazási területe a hallgatók tanulmányi sikerének előrejelzése, pontosabban azon hallgatók azonosítása, akiknél fennáll a veszélye, hogy megbuknak egy kurzuson vagy kiesnek a tanulmányaikból. Habár ezek az indukciós pontok is már elég reális igényt adnak a tanulóelemzés alkalmazásához, a kutatások és a gyakorlatból származó bizonyítékok azt mutatják, hogy a tanítás és tanulás támogatására ezen felül további potens és produktív lehetőségek rejlenek a használatában. [2]

Kulcsszavak: Learning Analytics, tanulóelemzés, adatvezérelt oktatás.

Abstract: Learning Analytics (LA) is a relatively young, interdisciplinary research field due to its technological foundations. It primarily focuses on understanding and optimizing learning within electronic learning environments and their relevance. According to the European Commission's

* Dunaújvárosi Egyetem, Tanárképző Központ; Pécsi Tudományegyetem, Oktatás és Társadalom Neveléstudományi Doktori Iskola
E-mail: kocsoe@uniduna.hu

[1] Európai Bizottság (2019): Az „Oktatás és képzés 2020” munkacsoportok főbb tevékenységei (2016–2017). https://education.ec.europa.eu/sites/default/files/document-library-docs/education-training-et2020-working-groups-2016-2017_hu.pdf

[2] SoLAR (2021. december 30.): What is Learning Analytics? Society for Learning Analytics Research (SoLAR). <https://www.solaresearch.org/about/what-is-learning-analytics/>

[1] Európai Bizottság (2019): *Az „Oktatás és képzés 2020” munkacsoportok főbb tevékenységei (2016–2017)*. https://education.ec.europa.eu/sites/default/files/document-library-docs/education-training-et2020-working-groups-2016-2017_hu.pdf

[2] SoLAR (2021. december 30.): What is Learning Analytics? *Society for Learning Analytics Research (SoLAR)*. <https://www.solaresearch.org/about/what-is-learning-analytics/>

[3] Németh A. (2013): A neveléstudomány főbb fejlődésmo-delljei és tudományos irányzatai. *Neveléstudomány*, 1., pp. 18–63. http://nevel-estudomany.elte.hu/downloads/2013/nevelestudo-many_2013_1_18-63.pdf

[4] Németh A. (2015): A neveléstudomány nemzetközi modelljei és tudományos irányzatai. *Magyar Pedagógia*, 115., (3.), 255–294. <https://doi.org/10.17670/MPed.2015.3.255>

[5] Csapó B. (2015): A PISA hatása a neveléstudomány fejlődésére. *Educatio*, 2015., (2), pp. 29–38 pp. <http://www.hier.iif.hu/hu/letoltes.php?fid=tartalom-sor/2426>.

Hungarian "Education and Training 2020" working group, LA "holds significant potential for improving the quality of education and learning" [1]. Learning analytics functions both as an academic field and a commercial market, which has rapidly evolved over the past decade. As a research and educational domain, it lies at the intersection of learning sciences (e.g., educational research, learning and assessment sciences, educational technology), analytics (e.g., statistics, visualization, computer science/data science, artificial intelligence), and human-centered design (e.g., usability, participatory design, socio-technical systems thinking).

Historically, the most common application of learning analytics has been predicting student academic success, specifically identifying students at risk of failing a course or dropping out of their studies. While these predictive capabilities provide a strong rationale for adopting learning analytics, research and empirical evidence from practice indicate that its applications extend beyond this scope, offering further potent and productive opportunities to support teaching and learning [2].

Keywords: Learning Analytics (LA), Learning Optimization, Student Success

A tanulóelemzés tudományterületi elhelyezése

A tanulóelemzés gyökerei ugyan a 20. század első feléig is visszavezetnek, erőteljesebb megjelenése az empirikus neveléstudomány, illetve az abból kibontakozó kvantitatív jellegű kognitív pedagógia korszakában észlelhetők először. A kognitív neveléstudományi szemléletnek fontos elemei az oktatási rendszerek makrofolyamatainak nemzetközi és országos szinten végzett mérései, mint például az OECD 2000. óta végzett PISA-vizsgálatai. [3, 4] Csapó Benő [5] szerint ezek – az időközben részben, majd teljesen technológiaalapúvá vált mérések – vezetnek át a *Big Data* korszakába, ami a neveléstudományi kutatásokat is megújítani látszik az interdiszciplináris kooperáció fokozódásával, valamint az adatgyűjtés és -elemzés kultúrájának újjáteremtésével.

A pedagógiai értékelésnek ezen formáihoz a hagyományos diszciplináris kereteket átlépő társadalomtudományi kutatásokon belül olyan összefogásra van szükség, mely kihasználja a gazdaságtudomány, a szociológia, a pszichometria és háttérben az alkalmazott matematika, a statisztika és a számítástechnika eszközeit is. [6]

Az információs forradalmak, illetve az elektronikus rendszerek használatából eredő rengeteg adatot számos iparágban – üzleti, egészségügyi, közigazgatási, közlekedési, szórakoztatóipari, telekommunikációs szférában – jóval korábban elkezdtek elemezni, használni, mint azt tesszük mostanra az oktatási szférában. Ez már csak azért is különös, mert az ezek alapját tekintő, az olyan korszerűnek számító oktatástechnológiai erőforrások, mint a tanulástámogató rendszerek meglehetősen elterjedtek, különösen a felsőoktatásban. [7, 8]

A mai felsőoktatásban a képzési tartalom rögzített formában történő átadási felületei elsősorban elektronikus tanulástámogató keretrendszerek (Learning Management System: LMS), ahol a tanulók mindenekelőtt a tanuláshoz szükséges, általában írásos vagy multimédia-alapú tananyagokat, útmutatókat találják. [9]

Mindazonáltal, ezeket a virtuális tereket már széleskörben használják nemcsak tananyagok, oktatási tartalmak, hanem tanulási tevékenységek, tanulói aktivitások gyűjtőhelyeként is, például tesztekhez és interaktív feladatokhoz [10], vagy akár a többirányú, általában aszinkron kommunikációhoz. A képzési tartalmakat bizonyos fókig az oktatók maguk választják csakúgy, mint az oktatási módszereiket.

A tanulástámogató rendszerek lehetővé teszik a teljesen online, de segítik a blended típusú, illetve a jelenléti oktatást is. Az egyik – nemzetközi szinten – elterjedt Moodle-rendszer 4.1-es verziójában 8-féle tananyag- és 20-féle tevékenységtípus illeszthető be a kurzusok tartalmába. Az ezekkel a tartalmakkal történő tanulói aktivitásokat – úgymint az oktatási anyagok megnyitása, letöltése, videók megtekintése, tesztek, feladatok elvégzése, pontszáma, kommunikációban való részvétel stb. –, monitorozhatjuk, naplózhatjuk, majd elemezhetjük. Ilyen gyakorlatokról olvashatunk Námesztovszki Zsolt, Ollé János és munkatársaik publikációiban is. [11, 12]

[6] Csapó B. (2011): Az oktatás tudományos háttérének fejlődése. *Magyar Tudomány*, 172., (9.), pp. 1065–1076. <http://www.matud.iif.hu/2011/09/06.htm>

[7] Joksimović, S.–Kovanović, V.–Dawson, S. (2019): *The Journey of Learning Analytics*, 6., pp. 37–63. <https://www.herdsa.org.au/system/files/HERD-SARHE2019v06p37.pdf>

[8] Siemens, G. (2014): The Journal of Learning Analytics: Supporting and Promoting Learning Analytics Research. *Journal of Learning Analytics*, 1., (1.), pp. 3–5. <https://doi.org/10.18608/jla.2014.11.2>

[9] Nagy, B.–Váraljai, M.–Mihalovicsné, K. A. (2020): E-learning spaces to empower students' collaborative work serving individual goals. *Acta Polytechnica Hungarica*, 17., (2.), pp. 97–114.

[10] Takács, A. M.–Takács, A.–Augusztinyi, D. G.–Nagy, B. (2024): On some methods of digital assessment for mathematical courses. In: Öbudai Egyetem (Ed.): *2024 IEEE 7th International Conference and Workshop Óbuda on Electrical and Power Engineering (CANDO-EPE)*. Proceedings. pp. 313–316. IEEE Hungary Section.

[11] Námesztovszki Z.–Balázs P. D.–Kovács C.–Major L.–Karuovic, D. (2017): Tanulói aktivitás mintázatai három MOOC-képzés alapján. *Információs Társadalom*, 16., (4.), pp. 40–60. <https://doi.org/10.22503/infars.XVI.2016.4.3>

[12] Ollé, J. (2019, május 23): Különböző tevékenységtervezési sémával készülő digitális tananyagok hatékonyságának összehasonlító elemzése. *HERA: Prevenció, Intervenció és Kompenzáció*, Eger. http://hera.org.hu/wp-content/uploads/2019/05/HuCER_absztrakt2019.pdf

- [9] Nagy, B.–Váraljai, M.–Mihalovicsné, K. A. (2020): E-learning spaces to empower students' collaborative work serving individual goals. *Acta Polytechnica Hungarica*, 17., (2.), pp. 97–114.
- [13] Baker, R. S. J.–Yacef, K. (2009): The State of Educational Data Mining in 2009. A Review and Future Visions. *Journal of Educational Data Mining*, 1., (1.), pp. 3–17. <https://doi.org/10.5281/zenodo.3554658>
- [14] JISC (2016): *Learning Analytics in Higher Education. A review of UK and international practice. CASE STUDY A: Traffic lights and interventions: Signals at Purdue University. Joint Information Systems Committee.* <https://analytics.jiscinvolve.org/wp/files/2016/04/CASE-STUDY-A-Purdue-University.pdf>
- [15] Picciano, A. G. (2012): *The Evolution of Big Data and Learning Analytics in American Higher Education.* Online Learning, 16(3): <https://doi.org/10.24059/olj.v16i3.267>
- [16] Sclater, N.–Mullan, J. (2017): *Jisc briefing: Learning analytics and student success – assessing the evidence.* https://repository.jisc.ac.uk/6560/1/learning-analytics_and_student_success.pdf
- [17] Sclater, N.–Peasgood, A. (2016): *Learning analytics in higher education. Jisc.* <https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education>
- [18] Wong, B. T. M. (2017): Learning analytics in higher education: An analysis of case studies. *Asian Association of Open Universities Journal*, 12., (1.), pp. 21–40. <https://doi.org/10.1108/AAOUJ-01-2017-0009>
- [19] Katonane Gyonyoru, K. I. (2024): The Role of AI-based Adaptive Learning Systems in Digital Education. *Journal of Applied Technical and Educational Sciences*, 14., (2.), pp. 1–12.
- [20] Microsoft. (2020): A magyar kormány az AI és a Big Data segítségével fejleszti az oktatási rendszert. *Microsoft Customers Stories.* <https://customers.microsoft.com/en-us/story/1350551297886456734-digitalis-jolet-non-profit-kft-education-azure-hu-hungary>
- [21] Vincent-Lancrin, S.–Vlies, R. van der. (2020): Trustworthy artificial intelligence (AI) in education: *Promises and challenges. OECD Education Working Papers*, No. 218. OECD Publishing, Paris. <https://doi.org/10.1787/a6c90fa9-en>
- [22] Bognár, L.–Fauszt, T.–Nagy, B. (2021): Machine learning model building techniques for small and medium-sized university courses. *International Journal of Artificial Intelligence*, 19., (2.), pp. 20–43.

Nemzetközi viszonylatban leginkább az Amerikai Egyesült Államok, Ausztrália, az Egyesült Királyság és Kanada felsőoktatási, felnőttképzési gyakorlatában megtalálhatóak a tanulástámogató rendszerekbe ágyazott, a tanulási-tanítási folyamatok vizsgálatára alkalmazott tanuláselemzési példák [13, 14, 15, 16, 17], de elterjedése a világban egyértelmű, például Hong Kong Nyitott Egyetemén történt megvalósulásról is fellelhető publikáció. [18]

Ezekhez a tanuláselemzési módszerekhez kapcsolódóan már a mesterséges intelligencia adta lehetőségeket is kihasználják, ami pár éve például a Digitális Jólét Program keretein belül Magyarországon is megjelent, mely során egy kísérleti program eredményeként a korai iskolaelhagyók arányát 20%-kal csökkentették. A tanulói tevékenységek megfigyeléséből, statisztikai elemzéséből rengeteg információhoz juthatunk, ami oktatóknak és döntéshozóknak is igen hasznosak lehetnek. [19] Ezen tanuláselemzési metódusok által képesek lehetünk a proaktív adatvezérelt oktatásra is. [20] Bár ezek a mesterséges intelligenciát használó alkalmazások még sok esetben kidolgozási fázisban vannak, és rendszerszinten egyelőre inkább csak kísérleti és lokális kontextusban használják. [21]

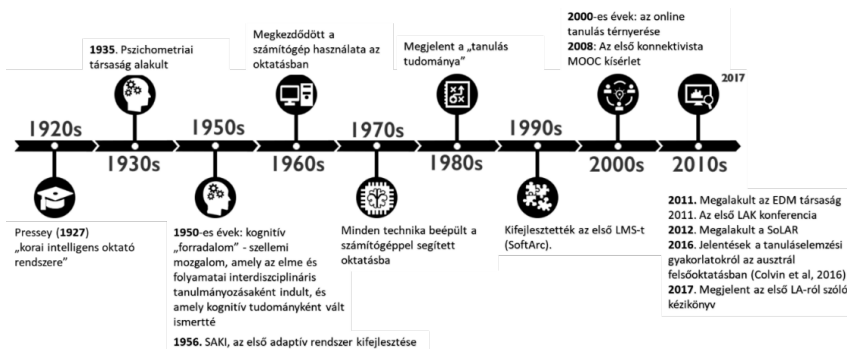
Az adatok statisztikai elemzéseiből, illetve a mesterséges intelligencia segítségével predikciókat készíthetünk, amivel előre jelezhetjük a hallgatói teljesítményeket, és feltérképezhetjük azokat, a lemorzsolódás szempontjából veszélyeztetett („at risk”) hallgatókat, akik a kurzus(ok) teljesítésének nehézségeivel küzdenek. [22, 9] Az elemzések során visszacsatolást kaphatunk a képzési tartalomra, az oktatási módszerekre vonatkozóan is. A mesterséges intelligenciával dolgozó gépi tanulási modellek számára az LMS-ben lévő tananyagok és tevékenységek tanulói felhasználása szolgál forrásadatként, így azok fejlesztésével nagyobb valószínűségű prognosztikákhoz juthatunk.

Az előrejelzések alapján az oktatók proaktív tanulástámogatási metódusokat iktathatnak be még a képzés korai szakaszában a tárgy teljesítése szempontjából különböző fokon veszélyeztetett hallgatók számára.

A tanuláselemzés kialakulásának csomópontjai

A tanuláselemzés történetének összefoglalója a Dél-Ausztráliai Egyetem oktatóinak munkája, a „*The Journey of Learning Analytics*” című tanulmánya, ami a Higher Education Research and Development Society of Australasia (Ausztrál-ázsiai Felsőoktatási Kutatási és Fejlesztési Társaság) „*Review in Higher Education*” (Felsőoktatási Szemle) folyóiratában jelent meg 2019-ben. Az alábbi történeti áttekintés elsősorban erre a publikációra hagyatkozik. (1. ábra)

1. ábra. A tanuláselemzés genealógiája



Forrás: [7] alapján saját ford. és szerk.

Srecko Joksimović, Vitomir Kovanović és Shane Dawson véleménye szerint a tanuláselemzés koncepciója Sidney L. Pressey, az Ohio-i Állami Egyetem pszichológia professzorának munkájára vezethető vissza, aki az 1920-as években kifejlesztett egy automatizált oktatógépet tesztek adminisztrálására. Ez egyfajta vizsgáztatógépként működött feleletválasztásos kérdéscsoportokkal aktivizálva a tanulót, a közvetlen megerősítés és egyéni haladás módszerét alkalmazva.

Az volt a célja, hogy a pedagógusok munkáját megkönnyítse a rutinszerű tanulás, az elméleti tananyag elsajátíttatásának gépiesítésével.

Az vitatott, hogy Pressey gépe tekinthető-e az intelligens oktatórendszerek kiindulópontjának, mindamellett, hogy az intelligens oktatórendszerek a tanulóelemzések eredetének egyik kulcsfontosságú területe.

Az *American Psychological Association* (Amerikai Pszichológiai Társaság: APA) 1935-ös ülésén alakult meg – eredetileg egy folyóirat indításának szándékával – a mai napig is tevékeny *Psychometric Society* (Pszichometriai Társaság), aminek alapító tagjai az oktatásban és a pszichológiában alkalmazott kvantitatív módszerek témája iránt érdeklődtek. Mára a társaság egy nonprofit szakmai szervezetként működik, amely a kvantitatív mérési gyakorlatok fejlesztésével foglalkozik a pszichológiában, az oktatásban és a társadalomtudományokban. (*History of The Psychometric Society* é. n.)

Egy másik mérvadó hatás volt a tanulóelemzések kialakulásában a már az előző pontban is említett kognitív tudomány megjelenése, az 1950-es években lezajló kognitív „forradalom”, ami – *Joksimović* és munkatársa szerint – *George A. Miller*, a kognitív pszichológia egyik megalapítójának munkájából, valamint a számítástechnika és a mesterséges intelligencia új fejlesztéseiből ered. *Gordon Pask*, felesége, *Elisabeth* és *Robin McKinnon-Wood* publikálta 1961-ben az első adaptív oktatásigép-rendszer, a SAKI (Self-Adaptive Keyboard Instructor: Önadaptív billentyűzetoktató) elkészítését.

A SAKI úgy optimalizálta a tanulási sebességet, hogy a feladatok nehézségét a tanuló teljesítményéhez igazította. Ugyancsak fontos felismerés volt *Benjamin Bloom* „2-sigma” elmélete a személyre szabott oktatás előnyeiről, mely további találmányokkal együtt az intelligens oktatórendszerek és a számítógéppel támogatott oktatás (computer-assisted instruction: CAI) területének fejlődését eredményezte.

A táv- és az online oktatás növekedése ugyancsak hozzájárult a tanulóelemzés fejlődéséhez. A postaszolgáltatások fejlődése, majd a különböző technológiai fejlesztések (rádió, TV, videó, CD/DVD, internet) tovább csökkentették a Michael G. Moore által „tranzakciós távolságnak” nevezett, a tartalom elérése, illetve a tanárunkkal és a többi diákkal való interakció közötti időt.

A távoktatás történetének kulcsfontosságú mérföldköve volt a kétirányú kommunikációs technológiák fejlődése az 1980-as években, ami elősegítette az elmozdulást az egyszerű információtovábbítástól a szociálkonstruktivista tanulás és a tanulók–oktatók közötti minőségi interakciók felé. Az 1990-es években a World Wide Web megjelenése webalapú távoktatási rendszereket hozott létre, amik mára a modern kor online tanulásává, majd a tömeges nyílt online kurzusokká (Massive Open Online Courses: MOOCs) fejlődtek. Az 1990-es és 2000-es években az internet elterjedése vezetett a már említett, webalapú Learning Management Systems távoktatási technológiákhoz, amelyeket egyre gyakrabban kezdtek használni a hagyományos tantermi oktatás támogatására, majd az online tevékenységek és értékelések egyre szélesebb körben elterjedt alkalmazásával kevert (blended) tanítási-tanulási formák születtek.

Konkréten a tanuláselemzés fogalmának megjelenése az elmúlt évtizedben bontakozott ki, de egy lépéssel megelőzte a *Big Data* jelenségek és módszerek mentén kialakuló, feltörekvő tudományág, az oktatási adatbányászat (Educational Data Mining; EDM), amely az oktatási környezetből származó egyedi és egyre nagyobb léptékű adatok feltárására szolgáló módszerek kidolgozásával, használatával, valamint a tanulók és a tanulási környezet jobb megértésével foglalkozik. A tanuláselemzés egy közeli ágának, az oktatási adatbányászatnak a folyóirata, a „*Journal of Educational Data Mining*” 2009-ben indult, közösségből pedig 2011-ben alakult meg az *International Educational Data Mining Society* (Nemzetközi Oktatási Adatbányászati Társaság). [23] Az első nemzetközi konferenciát, ami már kifejezetten a tanuláselemzés témájával foglalkozott, 2011-ben hívta életre egy kisebb kutatócsoport. „*Conference on Learning Analytics and Knowledge*” (A tanuláselemzés és a tudás nemzetközi konferenciája, LAK '11) címmel Kanadában került megrendezésre.

A konferencia közleményében megfogalmazottak szerint ennek a fórumnak a létrehozását az alábbi okok generálták:

1. A tudással, a tanítással és a tanulással kapcsolatban olyan mennyiségű adat születik, amely már meghaladja a szervezetek azon képességét, hogy megértsek azokat.
2. Az oktatási intézmények és vállalatok kevésbé használják ki azokat az adatokat, amelyeket a tanulók „kidobnak” a tananyagokhoz való hozzáférés, az oktatókkal és társaikkal való interakció, valamint az új tartalom létrehozása során.
3. Egy olyan korban, amikor az oktatási intézményekre egyre nagyobb nyomás nehezedik a költségek csökkentésére és a hatékonyság növelésére, az analitika egy olyan hasznos szemüvegnek ígérkezik, amelyen keresztül láthatóak és tervezhetőek a változások a kurzusok és az intézmények szintjén is.

Fentiek mentén a konferencia elsődleges célja az volt, hogy meghatározzák és megvizsgálják azokat a kutatási területeket, melyek a tanulók tanulásának a gépi tanulás, az adatbányászat és az adatvizualizációs módszerek használatával történő megértésével foglalkoznak. Ennek az első konferenciának eredményeképp 2012-ben megalakult a *Society for Learning Analytics Research* (SoLAR, Tanuláselemzési Kutatócsoport), és megszületett a tanuláselemzés definíciója: „adatok feltárása, gyűjtése, elemzése és jelentések készítése a tanulókról, valamint kontextusukban a tanulás, illetve a tanulás környezetének megértése, optimalizálása céljából.” [24]

[23] International Educational Data Mining Society. (é. n.): *Educational Data Mining*. Elérés: 2024. november 1., Forrás: <https://educationaldatamining.org/>

[24] Lang, C., Siemens, G.–Wise, A.–Gasevic, D. (Szerk.) (2017): *Handbook of Learning Analytics* (First). *Society for Learning Analytics Research* (SoLAR). <https://doi.org/10.18608/hla17>

2011. óta minden évben megtartották a LAK konferenciákat, bővítvén a téma szakirodalmának listáját. Ezenkívül a SoLAR „nyári egyetemeket” szervez (Learning Analytics Summer Institute: LASI), számos kezdeményezést indított a tanulóelemzés körüli együttműködésen alapuló és nyílt kutatások támogatására, a tanulóelemzési kutatások közzétételének és terjesztésének előmozdítására, valamint részt vesz az állami, tartományi és nemzeti kormányokkal való tanácsadásokban és konzultációkban.

A „*Handbook of Learning Analytics*” első kiadása 2017-ben jelent meg, célja a tanulóelemzés új és növekvő területének bemutatása volt. A 355 oldalas dokumentum a kutatás akkori állását tükrözte, és számos kiemelkedő szerző munkáját tartalmazta a tanulóelemzés és oktatási adatbányászat közösségéből.

A fejezetek négy fő témakörbe rendeződtek: 1. Alapvető fogalmak, 2. Technikák és megközelítések, 3. Alkalmazások, 4. Intézményi stratégiák és rendszerperspektívák. A fejezeteket a terület elkötelezett szakértői lektorálták, és a Society for Learning Analytics Research (SoLAR) valamint a International Society for Educational Data Mining jóváhagyásával jelentek meg.

2022-ben jelent meg a kézikönyv második kiadása, melynek célja, hogy kiegyensúlyozza a tudományos szigorúságot, a minőséget, a nyílt hozzáférést és a széles körű vonzerőt, valamint bevezetést nyújtson a kutatások jelenlegi állásába. A második kiadás a 2022-es év pillanatfelvételét nyújtja, és a tanulóelemzés közösségének számos prominens szerzőjének munkáját tartalmazza.

A Society for Learning Analytics Research (SoLAR) 2019-ben indította el a SCImago rendszerében mind az oktatás, mind az alkalmazott számítástudományok területén Q1-es szintű „Journal of Learning Analytics” folyóiratát. A folyóirat azóta is a tanulóelemzés vezető platformja, amely magasan jegyzett tudományos cikkeket és nemzetközi publikációkat közöl. A „*Journal of Learning Analytics*” célja, hogy elősegítse a tanulóelemzési kutatások fejlődését, és interdiszciplináris megközelítést kínáljon, összekötve az oktatási technológiák, az adatelemzés és a pedagógiai kutatások területét. 2023-ra a folyóirat növekvő impaktfaktorral rendelkezik, tovább erősítve pozícióját a nemzetközi tudományos közösségben.

Összefoglalás

Az elmúlt évtizedben a tanulóelemzés technológiaalapú, interdiszciplináris kutatási területté fejlődött, amely az elektronikus tanulási környezetekből származó adatok elemzésével optimalizálja a tanulási folyamatokat. A Big Data és az oktatástechnológiai innovációk fejlődése nyomán a tanulóelemzés mára olyan eszközöket kínál, amelyekkel előre jelezhető a tanulási siker vagy a lemorzsolódás kockázata.

Az LMS-rendszerekből gyűjtött adatok és a mesterséges intelligencián alapuló predikciók lehetőséget adnak az oktatóknak proaktív, adatvezérelt támogatási stratégiák alkalmazására, így javítva a tanulás minőségét és hatékonyságát.

A Society for Learning Analytics Research nemzetközi közösség tagjai különböző országokból és régiókból származnak. Ez a globális jelenlét lehetővé teszi a tanulóelemzés területén szerzett tapasztalatok és tudás széles körű megosztását és az együttműködést. A SoLAR által szervezett konferenciák és kiadványok meghatározó szerepet töltenek be a terület fejlődésében. Az LA interdiszciplináris megközelítése – ötvözve a kognitív tudományokat, az adatbányászatot és az oktatástechnológiát – a modern oktatás egyik kulcsterületévé teszi, amely a hallgatói sikeresség és a pedagógiai innováció előmozdítását szolgálja.

Galéria

Duma Bálint fotói (Bécs)































