

ZÁRÓBESZÁMOLÓ - FK: 128981

Kutatás Vezető: Dr Kmetty Zoltán

Projekt háttere

A digitális adatok társadalomtudományi használata a Számítógépes Társadalomtudomány (továbbiakban CSS) egyik legfontosabb kutatási iránya. Az elmúlt húsz évben jelentősen megnőtt a különböző digitális adatforrások elérhetősége. Ezek az adatforrások nagyon sokszínűek. Többek között ide tartoznak a különböző szenzorok és digitális eszközök által gyűjtött adatok, a hatalmas digitális tranzakciós adatbázisok (pl. bankok, egészségügy, államigazgatás), a korábbi korok digitalizált tartalmai (pl. régi újságcikkek, születési rekordok) és az online térben megjelentetett online tartalmak. Utóbbiak jelentős részét teszik ki a közösségi médiában megjelenő felhasználó által generált posztok, képek, kommentek és különféle aktivitások. Ezek a közösségi média tartalmak különösen izgalmasak társadalomtudományi szempontból, mert közvetlen megfigyelését teszik lehetővé az emberek viselkedésének, ami egy új nézőpontot ad a leginkább attitűdöket vizsgáló survey kutatásokkal szemben. A közösségi média adatokhoz való hozzáférés azonban nem triviális feladat. A 2010-es évek elején a hozzáférést alapvetően úgynevezett API-kon ((Application Programming Interface) történt, ami lehetővé tette azt, hogy a kutatók gyorsan és viszonylag olcsón férjenek hozzá nagy mennyiségű közösségi média tartalomhoz. Az API-k alapvetően a számítógépek egymás közötti kommunikációját/adatcseréjét segítik azzal, hogy autentikált csatornákon kaput nyitnak egy adatbázis bizonyos részéhez. A nyilvános API-k egyszerű hozzáférést biztosítanak nagy mennyiségű adathoz, de az adatok minősége változó és az is változó, hogy melyik platform mennyi adatot ad az API-kon keresztül. Bár bizonyos platformok esetekben – mint például a Twitter – ez a hozzáférési mód továbbra is az egyik leghatékonyabb adatelérési út, de más közösségi oldalak, például a Facebook vagy az Instagram esetében ezt a hozzáférési módot leállították vagy drasztikusan megnehezítették a platformok tulajdonosai (Breuer et al. 2021). Az API-k lezárása elsősorban a Cambridge Analytica botrány következménye, de a szigorodó adatvédelmi környezet mindentől függetlenül is abba az irányba terelte a platformokat, hogy szűkítsék az adathozzáférést. Ennek kapcsán Freelon (2018) azt írta, hogy a „Computational Social Science” belépett a „Post-API” korbá, Bruns (2019) pedig ezt az egész helyzetet „APIcalypse”-nek nevezte. Mások, mint Tromble (2021) vagy Puschmann (2019) viszont kiemelik ennek a folyamatnak a pozitív hatását, miszerint végre véget ért ezzel a „közösségi média kutatásának vadnyugata”.

Az itt bemutatott nehezedő adathozzáférési környezetben új modelleket kell kidolgozni a digitális adatokhoz való hozzáférés érdekében.

A NetGain Partnership (Shapiro et al. 2021) által publikált tanulmány két nagy ágat különbözteti meg a digitális adathozzáférésnek: Azok a megközelítések, amelyek együttműködnek a platformokkal és azok, amelyek nem.

A platform együttműködés a következő modellekre:

- API
- Differenciált adatvédelem (Differential Privacy – DP)
- Platformok direkt adatmegosztása publikálási megkötésekkel
- Kontrollált környezetben zajló hozzáférés

A platformtól független adatgyűjtési módok a következők

- Web-scrapelés
- App-scrapelés
- Adatdonáció

A 2018-ban induló kutatásunkban egy adatdonációs pilot kutatást kezdtünk el. Az adatdonációs modellben a kutatók megkérik a résztvevőket, hogy osszák meg velük a platformon tárolt adatokat. A nagy platformoknak a GDPR törvény megfelelés miatt lehetőséget kell adni a felhasználóknak, hogy elérjék és letöltsék a róluk tárolt adatokat, adatsomagok keretében (data download packages – DDP). A nagy nyugati platformok, mint a Google, a Facebook, az Instagram a WhatsApp vagy a Netflix felhasználóbarát módon adnak lehetőséget az adataink elérésére és letöltésére. Ezeket az adatokat amellet, hogy letöltheti a felhasználó akár további is oszthatja. Ez lehetőséget teremt a kutatóknak, hogy teljesen tiszta jogi környezetben férjenek hozzá közösségi média adatokhoz.

Projekt célkitűzések

A kutatásunk abból az itt bemutatott problémából indult ki, hogy a korábbi API alapú adathozzáférés már nem működik a Facebook esetében, és ezért új adatgyűjtési módokat kell keresni. A projekt kezdetekor azt a célt tűztük ki, hogy egy 150 fős kísérleti jellegű Facebook kutatást végezzünk, ahol az adatdonációs megközelítés technikai kivitelezhetőségét teszteljük, kialakítsuk egy olyan technikai és tartalmi keretet, ami alkalmas a nagyon diverz jellegű közösségi média adatok elemzésére, valamint megvizsgáljuk hogyan lehet összekötni a digitális térben keletkező adatokat survey adatokkal. A kutatásban két (kevésbé hangsúlyos) kutatási célt is kijelöltünk. Az egyik arra vonatkozott, hogy a közösségi média adatok jelentős részét kitevő szöveges adatokat milyen módszerekkel lehet hatékonyan feldolgozni és beépíteni a vizsgálati fókuszba. Az utolsó projekt célkitűzés pedig arra a kérdésre kereste a választ, hogy hogyan lehet ezekben az elemzésekben külső szereplőktől megvásárolt közösségi média adatokat is beilleszteni.

Adatgyűjtés

A 2018-ban induló adatgyűjtés egy teljesen új kutatási irány egyik első megvalósult kutatása volt. A pilot kutatásunk előtt tudomásunk szerint (a mai napig elérhető publikációk alapján) nem készült hasonló jellegű adatfelvétel, a mai napig is csak 1-2 kutatással találkozhatunk, ami ezt a megközelítést alkalmazza. Az innovatív megközelítés miatt számos kérdés esetében nem volt korábbi kutatói tapasztalat, ami miatt egy nagyon körültekintő kutatás tervezésre volt szükség. Ezért a projekt első félében a kutató csoport a projekt előkészítését végezte, ahol leteszteltük, hogy milyen adatok érhetőek el az adatdonációs megközelítéssel, mik azok az adatok amelyek elkérhetőek, hogyan lehet megoldani az adatfelvétel technikai kivitelezését és hogyan lehet biztosítani a különféle etikai és adatvédelmi szabályok maximális betartását. Az adatfelvétel részletes leírása megtalálható a Metszetekben megjelent tanulmányunkban (Kmetty 2021). A zárójelentésben csak a legfontosabb pontokat emeljük ki. A kutatás lebonyolításával egy professzionális piackutató céget bízunk meg. A tervezett (és végül meg is valósult) minta nagyság 150 fő volt. A kutatásban egy gyenge életkori kvótát használtunk, annyit kötöttünk ki, hogy a minta legalább 40 százaléka 30 év feletti kell, hogy legyen. A minta nem valószínűségi minta, tehát nem reprezentálja a Facebook használókat. A terepmunka 2019 áprilisa és szeptembere között zajlott. A kutatási alanyok 3000 forintot kaptak a részvételért. Azoknak a résztvevőknek, akik beleegyeztek a vizsgálatban való részvételbe, elküldtük a projekt leírását, és kaptak egy meghívót a piackutató cég helyi irodájába.

Az adatdonációs megközelítés egyik legnagyobb előnyének emeltük ki, a tiszta jogi környezetet és a jól érvényesíthető adatvédelmi szempontokat. Ennek fontos alapja a résztvevők beleegyező nyilatkozata a kutatásba, illetve az adatkezelési elvek elfogadása. Ehhez igazodva a vizsgálat megkezdése előtt a résztvevőknek alá kellett írniuk egy részvételi beleegyező nyilatkozatot. A munka során kifejlesztettünk egy R kódot, ami a letöltött JSON fájlokat előfeldolgozta, átalakította CSV fájlkká és eltávolította az adatokból a megtalált neveket. Fontos cél volt, hogy ne kelljen személyes adatot kezelni az adatfelvétel után, a személyes adatok kezelése teljes egészében a piackutató cég feladata legyen. Ehhez a kutatási alanyok nevét, és a Facebook barátaiknak a nevét elfedtük az algoritmussal, hashelés segítségével (lásd hasonló kapcsán: Mancosu – Vegetti, 2020).

A Facebook-adataik megosztása mellett a résztvevőknek egy online kérdőívet is ki kellett tölteniük amíg az irodában tartózkodtak. A kérdőív nagyon változatos témákat fedett le, tartalmazott kérdéseket a politikáról, a médiahasználatról, az önreprezentációról, a mentális egészségről, a szabadidős tevékenységekről és a zenei preferenciákról is. A két adatforrás összekapcsolása érdekében ugyanazt az azonosító kódot használtuk az online kérdőívben és a Facebook-adatok tárolásakor. A nyers Facebook-adatokat az anonimizálási folyamat után törölte a kutató cég. A piackutató cég az anonimizált adatokat csak a vizsgálat vezető kutatójával osztotta meg.

Az adatok teljes feldolgozása közel két évig tartott a nagyon komplex adatszerkezet miatt. A végső adatbázisokat bizonyos megkötésekkel elérhetővé tettük a KDK kutatási archívumban. A kutatás a következő linken érhető el: <https://openarchive.tk.mta.hu/508/>

Kutatási irányok

A komplex adatbázis és a széles témaszett miatt több fókusz is volt a projektnek. A következőkben bemutatjuk az egyes kutatási irányokat és az ezekhez kapcsolódó szakmai megjelenéseket.

Módszertani kutatások

A kutatásunkban használt módszertan egyedi választ ad arra a problémára, hogyan lehet közösségi média adatokhoz hozzáférni a korábbi alapvetően API alapú csatornák bezárása után. A kutatás nemzetközi szinten is az első között használta az adatdonációs megközelítést. A kialakított módszertani keretet 2019-ben két vezető nemzetközi konferencián is bemutattuk.

Zoltán Kmetty, Ádám Stefkovics, Anna Vancsó, Dániel Váry: Acquiring personal Facebook data - Is it still possible? 8'th Conference of the European Survey Research Association, 15'th-19'th July 2019, Zagreb

Zoltán Kmetty, Ádám Stefkovics, Anna Vancsó, Dániel Váry: Acquiring personal Facebook data - Is it still possible? 4'th European Conference on Social Networks, 9-12 September, 2019, Zürich

A módszer nemzetközi disszeminációjának legfontosabb eleme a GESIS kutatócsoporttal közös publikációnk. A közös tanulmányban azt mutatjuk be, hogyan lehet különböző adatdonációs megoldásokkal adatokat gyűjteni. A közös tanulmány arra fókuszál milyen előnyök és lehetőségek rejlenek a közösségi média használók direkt beleegyezésén alapuló kutatásokban. A közös munka eredményeként elkészült tanulmányunk az Information Communication and Society (IF:5.4) lapban revízió alatt van (a második körös lektori vélemények alapján 2022 áprilisában küldjük vissza a tanulmányt).

Johannes Breuer, Zoltán Kmetty, Mario Haim, and Sebastian Stier: User-focused approaches for collecting Facebook data in the "post-API age": Experiences from two studies and recommendations for future research (revise/resubmit alatt)

A kutatás hazai disszeminációját is fontosnak tartottuk. Itt elsősorban az motivált minket, hogy a hazai tanítási gyakorlatban legyen olyan elérhető magyar tanulmány, amit BA és MA szinten is ki lehet adni hallgatóknak. A magyar nyelvű tanulmányunk a *Metszetek* folyóiratban jelent meg.

Kmetty, Z. (2021). Új utak a társadalom megismerésében. Metszetek, 10(3).

Zenei ízlés

Részben módszertani, részben tartalmi fókuszúak azok az elemzések, amelyben Németh Renátával közösen azt vizsgáljuk hogyan lehet operacionalizálni a zenei ízlést Facebook adatok alapján. A Facebookon rengeteg olyan adat keletkezik (pl: követett oldalak listája), ami közvetve információt nyújt arról, hogy milyen zenei ízlése van egy felhasználónak, sőt akár ennek a zenei ízlésnek a változása is nyomon követhető. Elsősorban azt vizsgáljuk ebben a kutatásban, hogy a survey adatokból kinyert zenei ízlés mennyiben köthető össze a FB alapján kiszámolt zenei ízléssel. A kutatásunkat több nemzetközi konferencián is előadtuk, publikációkat, pedig elfogadta a BMS (Q2), a Nemzetközi Szociológia Társaság (ISA) módszertani csoportjának (RC33) lapja.

Zoltán Kmetty, Renáta Németh: Same or different? Comparing Facebook data with survey answers on music taste, Workshop on the Concept of. Quality for Big Data, 25-26 February, 2020, Budapest

Zoltán Kmetty, Renáta Németh: Same or different? Comparing Facebook data with survey answers on music taste, Ic2s2 Conference, 17-20 July, 2020, online

Zoltán Kmetty, Renáta Németh: Which is your favorite music genre? A validity comparison of Facebook data and survey data. Bigsurv Conference, 06-27 November, 2020, online - <https://youtu.be/5JfPs-1hzos>

Kmetty, Z., & Németh, R. (2020). Which is your favorite music genre? A validity comparison of Facebook data and survey data. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 07591063211061754.

Depressziós és FB használat

Az elmúlt 5-10 évben sok kutatás foglalkozott azzal, hogy a közösségi média használatból hogyan lehet következtetni különböző depressziós tünetekre. A kérdőívbe elhelyeztünk egy külön blokkot a depresszió mérésére (PHQ-9 mérési módszer). A felhasználóktól származó nagyon részletes FB használati adatok pedig lehetővé teszik, hogy megvizsgáljuk, mik azok a FB használati jellemzők, amelyek összefüggést mutatnak magasabb depressziós szinttel. Bozsonyi Károllyal (KRE) közös munkánk eredményeit vezető nemzetközi konferenciákon mutattuk be – a kutatásra épülő publikáció pedig 2022-ben jelent meg a Social Scineces (Q2) folyóiratban. Ehhez a kutatási irányhoz kapcsolódott Lovák Zsanett (ELTE Survey MSC) szakdolgozata is.

Zoltán Kmetty, Károly Bozsonyi: *Identifying depression related behaviour in Facebook – an experimental study*, Sunbelt Conference, 13-17 July, 2020, online

Zoltán Kmetty, Károly Bozsonyi: *Identifying depression related behaviour in Facebook – an experimental study*, Ic2s2 Conference, 17-20 July, 2020, online

Zoltán Kmetty, Károly Bozsonyi: *Identifying depression related behaviour in Facebook – an experimental study*, Bigsurv Conference, 06-27 November, 2020, online - <https://youtu.be/ofE--0mBnds>

Kmetty, Z., & Bozsonyi, K. (2022). *Identifying Depression-Related Behavior on Facebook—An Experimental Study*. *Social Sciences*, 11(3), 135.

Lovák Zsanett 2020: *Depressziós jelek azonosítása Facebook adatokkal. Szakdolgozat, ELTE, Survey szak.*
Témavezető: Kmetty Zoltán.
https://fbpilot.tatk.elte.hu/media/05/34/f8ae791b4418685123686cbc923001a7e61454efdbd1e659b2ea769677ed/Lovák%20Zsanett_GST2DU_szakdolgozat_2020-21-1.pdf

A depresszióval kapcsolatos kutatásaink másik irányát azok a munkák jelentették, amelyekben azt vizsgáltuk, hogy milyen nyelvi jegyekkel lehet azonosítani a depressziós tüneteket mutató résztvevőket. Ez az irány szorosan kapcsolódik, ahhoz a kutatási célkitűzésünkhöz, amiben a különböző szöveganalitikai megoldások lehetséges használatát céloztuk meg. Németh Renáta vezetésével Farkas Dalma 2020-ban írt szakdolgozatot a témából, Kmetty Zoltán vezetésével pedig Kelemen Sára 2021-ben az ELTE Survey Msc szakán. A legújabb szöveganalitikai megoldásokat használó tanulmányunk eredményeit a XVIII Magyar Számítógépes Nyelvészeti konferencián (MSZNY) mutattuk be 2022-ben.

Farkas Dalma (2020): *A depresszió tüneteinek felismerése szövegelemzés módszerével Facebook posztok alapján.* Szakdolgozat, ELTE, Survey szak. Témavezető: Németh Renáta
https://fbpilot.tatk.elte.hu/media/71/18/1dce0055c58ba2141bfabaeae7cb4438c1e2544308df02305fec97b67639/Farkas%20Alma%20Dalma_SVHKR4_2019-20-2.pdf

Kelemen Sára (2021): *Facebook posztokra alapozott nyelvhasználati kutatás.* Szakdolgozat, ELTE, Survey szak. Témavezető: Kmetty Zoltán

Kelemen, Sára Bernadett ; Yang, Zijian Győző ; Kmetty, Zoltán (2022): *Nemek közötti nyelvhasználati különbségek a Facebook bejegyzésekben*
In: Berend, Gábor; Gosztolya, Gábor; Vincze, Veronika (szerk.) XVIII. Magyar Számítógépes Nyelvészeti Konferencia Szeged, Magyarország : Szegedi Tudományegyetem, Informatikai Intézet (2022) 656 p. pp. 573-585. , 13 p.

Network és dinamika

A kutatási adatoknak elsősorban a társadalomtudományi alkalmazhatóságára fókuszáltunk, de ezzel párhuzamosan két olyan projektünk is elindult, amelyben fizikusokkal és számítógépes mérnökökkel közösen gondolkodtunk az adatok további felhasználhatóságán.

Az első projekt egy szimulációs megközelítéshez csatlakozott. Nagy gráfokat viszonylag egyszerűen lehet készíteni, de olyan gráfokat, amelyeknek a tulajdonságai hasonlóak a közösségi média hálózatokhoz már sokkal nehezebb. Az ilyen gráfok technikailag nagyon fontosak, mert segítségükkel jól lehet tesztelni különböző felhasználói viselkedéseknek komplexitását és ezáltal jobban lehet hálózati architektúrákat építeni. Egy ilyen témára fókuszáló szimulációs projekthez adtunk adatbázisunkból származó meta adatokat. Az adatainkat felhasználó publikáció meg is jelent a 2020-as évben.

Waudby, J., Steer, B. A., Prat-Pérez, A., & Szárnyas, G. (2020, June). Supporting Dynamic Graphs and Temporal Entity Deletions in the LDBC Social Network Benchmark's Data Generator. In *Proceedings of the 3rd Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA)* (pp. 1-8).

A második projektben a CEU Department of Network and Data Science (DNDS) kutatóival közösen a Facebook aktivitások temporális dinamikáját vizsgáltuk. Az elmúlt években heves vita bontakozott ki abban a kérdésben, hogy a különböző online aktivitás mögötti temporális dinamikát elsősorban a cirkadián ritmus szabályozza, vagy a közösségi média platformok belső logikája. A kutatásunk lehetőséget ad arra, hogy nagyon részletesen megvizsgáljuk a Facebook aktivitás időbeli mintázatát. Karsai Mártonnal (CEU), Kertész Jánossal (CEU) és Koltai Júliával (ELTE, TK CSS RECENS) közös kutatásunkat több vezető nemzetközi konferencián bemutattuk
Zoltán Kmetty, Márton Karsai, Julia Koltai, János Kertész: *Context of burstiness in Facebook activities*, *Ic2s2 Conference, 17-20 July, 2020, online*

Zoltán Kmetty, Márton Karsai, Julia Koltai, János Kertész: *Context of burstiness in Facebook activities*, *Sunbelt and NetSci Conference, 6-11 July, 2021, online*

Nyilvános Facebook adatok

A Facebook adatokhoz való hozzáférés nem egyetlen módja az adatdonációs megközelítés, a nyilvános Facebook oldalak tartalmak elérhetők különböző Social Listening szolgáltatókon keresztül. Fontos kutatási kérdésünk volt, hogy ezek a tartalmak milyen elemzési célokra alkalmasak, milyen társadalomtudományi kérdéseket lehet megválaszolni ezeken az adatokon keresztül. A kutatásunk fókuszában a nyelvhasználat, ezen belül is a káromkodások álltak. Szabó Gabriellával (TK PTI) és Molnár Emesével (ELTE BTK) indított kutatásunkban azt vizsgáltuk milyen káromkodási formák azonosíthatók a közösségi média oldalakon és az online megjelenő újságcikkek hozzászólásaiban. Ehhez elkészítettünk egy részletes káromkodás szótárt és a szótár segítségével több mint 17 millió kommentet elemeztünk. Kutatásunk megjelent az *International Journal of Communication* lapban (Q1).

A témát Knap Árpáddal tovább vittük és a 2020-21-es koronavírus kapcsán megjelent kommentekben is megvizsgáltuk a káromkodások arányát és alakulását. Tanulmányunk egy Szabó Gabriella által szerkesztett hazai kötetben jelent meg 2022-ben.

Szabó, G., Kmetty, Z., & Molnár, E. K. (2021). *Politics and Incivility in the Online Comments: What is Beyond the Norm-Violation Approach?*. *International Journal of Communication*, 15, 26.
<https://ijoc.org/index.php/ijoc/article/viewFile/16411/3403>

Kmetty, Zoltán ; Knap, Árpád: *Trágárság mint érzelmi válasz a COVID-19-járvány idején* In: Szabó, Gabriella *Érzelmelek és járványpolitizálás. Politikai érzelmenedzserek és érzelmszabályozási ajánlataik Magyarországon a COVID-19 pandémia idején.* Budapest, Magyarország : ELTE Eötvös Kiadó (2022) 199 p. p. 173

A projekt célkitűzésének teljesülései

A projekt elsődleges célkitűzése az volt, hogy technikai és tartalmi oldalról megvizsgálja az adatdonációs megközelítés társadalomtudományi használhatóságát. Az adatfelvételt a terveknek megfelelően az első évben megvalósítottuk, a módszertani keretrendszer több nemzetközi konferencián bemutattuk, a tervezett publikációk részben megjelentek. A q1 lapba beadott módszertani tanulmányunk várhatóan 2022 tavaszán/nyarán fog megjelenni. Fontos eredménynek tartjuk, hogy a nagyon komplex adatrendszernek egy olyan struktúrát tudunk kialakítani, ami lehetővé tette a repozitóriumi elhelyezését. Az adatok a KDK repozitóriumában elérhetők a vezető kutató engedélyével. Az adatgyűjtésre épülő tartalmi elemzések két témát érintettek, a zenei stílust, valamint a depresszió azonosíthatóságát. Mindkét esetben nemzetközi konferenciákon mutattuk be az eredményeket, a tanulmányok Q2 lapokban jelentek meg. Itt fontos megjegyeznünk, hogy a magasabb presztízsű lapokban való publikálást nagyban megnehezítette, a minta kis mérete, valamint az, hogy nem reprezentatív a minta a magyar Facebook használókra.

Célként tűztük ki a projekt kezdetekor, hogy megvizsgáljuk hogyan viszonyulnak a külső adatforrásból beszerzett Facebook adatok az adatdonáció során elérhetővé vált adatokhoz. A kutatás korai

szakaszában világossá vált, hogy a különböző adatforrások nem elemezhetők ugyanabban a kutatási keretben, más típusú kérdések megválaszolására alkalmasak az adatdonációból származó adatok és más típusú kérdések megválaszolására használhatók a Social Listening szolgáltatókon beszerezhető adatok. Utóbbi adatokra építve elsősorban a káromkodások kommentekben való megjelenését vizsgáltuk. Ebből a projektből egy hazai kötet tanulmány és egy nemzetközi Q1 publikáció jelent meg. Ezek a tanulmányok szervesen összekapcsolódnak a szöveganalitikai fókuszú célkitűzésünkkel. A szöveganalitikai megközelítés azonban elsősorban a depresszió nyelvi jegyeit vizsgáló munkánkban jelent meg. Ezeknek az eredményeknek a disszeminációja nem zárul le a projekttel, az eredményeinket 2022 nyarán a CSS vezető konferenciáján, az IC2S2 konferencián fogjuk bemutatni.

Ahogy a rövid összefoglaló is mutatja a projekt elején vállalt célkitűzéseket meg tudtuk valósítani, eredményeinket vezető nemzetközi konferenciákon és Q1 /Q2 lapokban tudtuk publikálni, az adatbázist pedig a vállaltak megfelelően elérhetővé tettük a KDK repozitóriumában.

Erőforrások

A kutatásban a kutatás vezetőn kívül (Kmetty Zoltán) egy vezető kutató vett részt, Németh Renáta. A projektbe bevontunk számos fiatal PhD előtt álló kollégát, Boros Krisztiánt, Tarnói Csengét, Váry Dánielt, Mogyorósi Pálmát, Stefkovics Ádámot és Vancsó Annát.

A COVID járvány miatt az eredetileg tervezett konferencia utak egy része nem valósult meg, ezért NKFI engedéllyel a költségkeret egy részét átcsoportosítottuk bérekre. Ezeket az erőforrásokat arra használtuk, hogy javítsunk az adatminőségen és komplex képzett változókat hozunk létre a későbbi adatelemzésekhez (pl. Facebook aktivitások csoportosítása, érdeklődési kategóriák aggregálása, stb..). Az átcsoportosítás ellenére a projekt költségek egy része nem került elköltésre. A projekt keretében nagyobb összeget tettünk félre open-access publikációra, de a publikációinkat közlő lapok közül kettő (IJOE, BMS) ingyenes open-access publikációt tett lehetővé, a Social Science-ben megjelenő publikációnk a pénzügyi zárás (2022.02.28) után került elfogadásra, a módszertani tanulmányunk (ICOS) pedig még lektorálás alatt van.

Hasznosulás

A kutatás legfontosabb eredménye, az adatdonációs adatgyűjtés módszertani keretének a kidolgozása és tesztelése volt. Ez megágyaz a további hasonló jellegű kutatásoknak (lásd utolsó fejezet). A nemzetközi konferencia megjelenések, valamint a megjelenő nemzetközi publikációk jelentős láthatóságot biztosítottak a projektnek a tágabb tudományos közegben, a hazai megjelenések pedig hozzájárulnak a számítógépes társadalomtudomány (Computational Social Science) mélyebb integrálódásához a szakmában. Projektünk fontos küldetése, hogy a keletkezett újszerű adatokat minél szélesebb körben használják. Ennek a célnak a része volt az, hogy az ELTE TÁTK képzésében több kurzust is dedikáltunk a kutatás feldolgozásra, mind BA mind MA/MSZ szinten. A kurzusokon a hallgatók megismerkedtek az adatgyűjtéssel, az adatfeldolgozás lehetséges módjaival és 2-3 fős csoportokban egy saját témát is feldolgoztak. A kutatás alapján több nagyon érdekes elemzés készült, ezek egy részét elérhetővé tettük a honlapon is. A kurzusok mellett az anonimizált adatokat elérhetővé tettük szakdolgozóknak is. A projekt zárásáig 3 Msc és egy BA szakdolgozat készült el az adatdonációs adatbázisokra épülve.

A projekttel kapcsolatos anyagok elérhetők a projekt honlapján: <https://fbpilot.tatk.elte.hu>

Zárás helyett

A pályázat hivatalosan 2022 február 28-án zárult le. A pályázati zárás azonban nem jelenti az adatdonációs projekt befejezését. A vezető kutató (Kmetty Zoltán), sikeresen pályázott az ELKH kiemelt kutatás pályázatára, aminek keretében 60 millió forintot nyert el egy nagyobb léptékű adatdonációs projektre. Az új projektben 1000 fős reprezentatív mintán fog megvalósulni az adatgyűjtés és a Facebook mellett más platformokat is meg fog célozni (Instagram, Google, Twitter). Az új kutatás 2021 decemberében indult és szervesen épül az itt bemutatott OTKA kutatás eredményeire. Ezek a kutatások fontos sarokkövei annak a célkitűzésünknek, hogy Budapest az egyik központja legyen az újszerű digitális adatok kutatásának.