

Ide nekem a nagyit!

Hozzászólás „A szintetikus dokumentumok és a jövő archívumai” vitához

A 20. Század Hangja Archívum nem gyűjt, és egyelőre nem is szándékozik szintetikus dokumentumokat gyűjteni. Egy olyan rekordtípusra specializálódott, amely azonban a legfrissebb technológiák és az MI kitüntetett figyelmét élvezi. Miért is? Mi interjúkat gyűjtünk: korábbi párbeszédék írásos, hangzó- vagy videós rekordjait. Ezek a számítástudomány (computational science) szemében olyan komplex adatgyűjtemények, amelyek a „valósághoz” nagyon közel esnek, és azt szokatlanul árnyaltan – szemben, mondjuk, egy TikTok-bejegyzéssel vagy egy YouTube-reelsszel – képesek bemutatni. Mivel ezek általában hosszú, gyakran több órás elbeszélések/beszélgetések, az érzelmektől a gesztusokon át a nyelvi és diskurzív sajátosságokig kincseshányái a mai modern technológiáknak.

A (szintetikus) beavatkozásnak számos megjelenési formájával lehet találkozni, először vegyük át ezeket. Kezdjük azokkal a beavatkozásokkal, melyeket mi magunk, azaz az Archívum végezhet el ezeken a fizikai forrásokon. Először is ezeket digitalizáljuk, s ezzel egy korábban csak helyben kutatható forrást teszünk globálisan megoszthatóvá – másfelől, ezzel máris megfosztottuk a forrást a saját aurájától (már akinek ez számít, lásd Walter Benjamin szép esszéjét a fotográfiáról). Hogy a kutatást elősegítsük, az elhangzottakat AI segítségével leírjuk, ezzel máris egy új, hibrid rekordot hozunk létre, amely nem az interjúval keletkezett. Ezzel általában senki sem törődik, és ettől a pillanattól fogva „írott forrásként” tekint az interjúra. Az OCR-es PDF-ekben már mindenre rá lehet keresni: kezdetben ez a klasszikus *cherry-picking*-logika szerint működött, azaz tetszőleges interjúrészek kerültek bele a tudományos publikációkba, elsősorban azon elv szerint, hogy az adott kifejezés szerepelt-e a szövegben. A kutatók többsége egyáltalán nem vette a fáradságot, hogy reflektáljon a saját eljárására, azaz arra, hogy mit jelent egy egykori dialógus (amelynek hossza, dinamikája van) szavakra bontása. Egyáltalán nem volt szélsőséges az a felhasználási mód sem, amelyben az elbeszélő maximum a lábjegyzetben jelent meg, és ott is csak annyi, mikor vették fel vele az interjút – magáról az interjúról mint egészről nem esett szó.

Az elmúlt évtizedben a természetesnyelv-feldolgozás, az NLP szakadatlan fejlődése egyre nagyobb szövegegységek komplex elemzését teszi lehetővé, itt látok egy kicsi esélyt arra, hogy az MI előbb-utóbb olyan összefüggésekre világít rá, amelyekkel a történészek, szociológusok, nyelvészek stb. a hagyományos módszerekkel nem boldogultak. Mondok egy példát nem az interjú források köréből: ma már szinte az összes nemzeti bank történeti archívuma digitalizálva van. Feltételezem, hogy ha az 1929–1933 közötti jegyzőkönyveket a számítástudomány szakértőinek a kezébe adnánk, olyan összefüggésekre mutatnának rá a nagy gazdasági válságról, amelyek a történészek számára mindeddig láthatatlanok voltak.

De térjünk vissza az archívumunkhoz: egy felhasználóbarát archívum sokat ad arra, hogy a kollektívái jól kereshetők legyenek. Ehhez olyan metaadatokra van szüksége, melyeket eddig egyéni kurálással, kézzel hozott létre. A mai technológiák kedveznek a hibrid megoldásoknak: az MI egyre „okosabban” szűri le egy-egy interjúból a lényeges tárgyszavakat, ezek megfelelő ellenőrzés után is hatékonyabb és vélhetően semlegesebb metaadatokra nyújtanak lehetőséget, mint eddig. Kevésbé vagyok optimista a rövid összefoglalások esetében: ott éppenséggel az egyes interjú sajátos zsánere veszik el a vegytiszta összefoglalásban. Ugyancsak a felhasználónak tudunk segíteni azzal, ha a saját gyűjteményünket más gyűjteményekkel automatizált formában tudjuk összekötni (itt nem csak a geoadatok, személyek, intézmények stb. összekapcsolására gondolok, hanem beszédmódok, kulturális kifejezési formák és normák összekapcsolására is). Egy bizonyos időszakban az egész keleti blokkban készültek például nagy volumenű, szociológiai interjúk kutatások, melyek rekordjai többnyire fennmaradtak: ezek összekapcsolása, „egy korpuszként” kezelése ma még talán kihívás a számítástudománynak, de valószínűleg lehetséges. Az így létrejövő hibrid vagy szintetikus szövegek (a szegénység, a kulturális azonosságok, a munka, a család stb. elbeszélései) egészen biztosan gazdagítanak szociológiai tudásunkat.

Végül az archívumunknak – és tágabban minden, *testimony*t gyűjtő archívumnak – szembe kell néznie azzal, hogy adományozóink is vonzódnak az MI-hez és a legfrissebb technológiákhoz.

A USC Shoah Foundation volt az első, amely virtuális konverzációs platformot hozott létre, amelyben a hologramszerűen kivetített Pinchas Guttert faggathatjuk (immár virtuálisan is) a holokausztban megélt tapasztalatairól (Iwitness). S noha születése óta óriási szakmai vita kíséri ezeket a technológiai fejlesztéseket, vitathatatlan, hogy a túlélők örömmel ültek be napokra a zöld szobába, hogy legyártsák a több tízezer kérdésre a választ, és a múzeumok is gombamód veszik át ezt az edukációs technológiát. Ez ugye hibrid abban az értelemben, hogy lekövethető az emberi szándék a keletkezéstörténetben, abban azonban teljesen szintetikus, hogy ilyen dialógus a való életben soha nem zajlott le. A technológia iszonyatosan sok pénzbe kerül, de annyira vonzó, hogy elképzelhető, hogy hamarosan egy kedves családtagunkkal készítünk majd ilyen virtuális beszélgetést, hogy aztán, amikor már nem lehet közöttünk, meginvitáljuk a vasárnapi ebédhez, hogy meséljen az unokáknak. Így válik aztán egy teljesen szintetikus rekord a valós családi emlékezet részévé. Feltételezem, ha ez divatba jön, a 20. Század Hangja Archívum lesz az első, amely ezeket a „hologramokat” gyűjteni fogja.

Hivatkozások

Iwitness. University of Southern California Shoah Foundation. Utolsó hozzáférés: 2026. március 26.

<https://iwitness.usc.edu/dit/pinchas>

Kovács Éva kutatóprofesszor az ELTE TK Szociológiai Intézetében és megbízott tudományos igazgató a bécsi Wiesenthal Intézetben. A 20. Század Hangja Archívum és Kutatóműhely alapítója, az EHRI (European Holocaust Research Infrastructure) szakértője. Főbb kutatási területei: 20. századi történelempolitika és társadalmi emlékezet. Emellett oktatja és kutatja a kvalitatív társadalomkutatási módszereket, különös tekintettel az interjú technikákra.

Így hivatkozzon erre a cikkre:

<https://dx.doi.org/10.22503/inftars.XXVI.2026.1.5>