

Disruptivne tehnologije u arhivima

prof. dr. sc. Hrvoje Stančić
Odsjek za informacijske i komunikacijske znanosti
Filozofski fakultet Sveučilišta u Zagrebu
Ivana Lučića 3, Zagreb
hstancic@ffzg.hr

Sažetak

Rad se uvodno osvrće na trenutno stanje u arhivima i detektira potrebu za korištenjem suvremenih tehnologija. Neke od njih donose korjenite promjene ili, pak, mijenjaju ustaljene načine poslovanja pa se zbog toga i nazivaju disruptivnim tehnologijama. Potom se detaljnije pojašnjava sam pojam i identificiraju osnovne tehnologije koje će imati značajan utjecaj na poslovanje arhiva. Analiziraju se umjetna inteligencija, ulančani blokovi, velika količina podataka i tehnologije digitalne participacije – najprije se daju osnovna objašnjenja kako pojedina tehnologija funkcionira, potom prikazuju primjeri te diskutira o potencijalnoj primjeni u arhivima. Zaključno rad donosi usporedbu očuvanja analognog u odnosu na digitalno gradivo, diskutira o pristupima financiranju te argumentira potrebu da se arhivi znatnije okrenu korištenju disruptivnih tehnologija.

Ključne riječi: arhiv, digitalno gradivo, umjetna inteligencija, ulančani blokovi, velika količina podataka, informacijski sustavi potpomognuti mnoštvom, igrifikacija

Uvod

Arhivi se već neko vrijeme nalaze u procesu digitalne transformacije. To znači da istovremeno funkcioniraju u “dva svijeta” – analognom i digitalnom. S jedne strane se, dakle, brinu za očuvanje tradicionalnoga gradiva koje će uvijek biti prisutno u arhivima i koje ima svoju vrijednost. S druge strane se nalaze dva procesa. Jedan se odnosi na digitalizaciju, tj. pretvaranje analognih arhivskih zapisa u digitalne radi njihove dostupnosti i/ili očuvanja. Drugi se odnosi na preuzimanje gradiva koje je netko drugi digitalizirao ili onoga koje je izvorno nastalo u digitalnome obliku. Oba procesa uključuju upravljanje velikim količinama gradiva. Velika količina podataka, ili engleski *big data*, jest termin kojim se označavaju situacije u kojima postoji mnoštvo digitalnih zapisa kojima treba upravljati, očuvati ih i osigurati im pristup, a da su pritom oni pohranjeni u mnogim, različitim formatima te da tog sadržaja ima više nego što se stigne pročitati i obraditi. S obzirom da je sve navedeno svakodnevica u kontekstu arhiva, može se odmah u uvodu ovog članka konstatirati da arhivi rade s velikim količinama podataka (informacija, dokumenata, zapisa). Stoga su arhivima potrebne suvremene tehnologije kako bi se uspješno nosili sa svojim temeljnim zadaćama. No, suvremene tehnologije mogu s jedne strane pomoći, ali s druge strane uzrokovati i neke korjenite promjene, odnosno dubokom promijeniti standardne načine poslovanja. Zbog toga se takve tehnologije nazivaju disruptivnim tehnologijama, odnosno tehnologijama koje iz temelja mijenjaju ustaljene procese. U nastavku ovoga rada identificirat će se takve tehnologije, analizirati njihova relevantnost za područje rada arhiva te procijeniti njihov potencijalni utjecaj.

Disruptivne tehnologije

Clayton M. Christensen, profesor Harvardske poslovne škole, skovao je termin “disruptivne tehnologije” u svojoj knjizi “The Innovator’s Dilemma” iz 1997. godine¹. To su tehnologije koje su nove, koje su zanimljive ograničenom broju korisnika, koje se još nisu dokazale u praksi, ali imaju veliki potencijal uzdrmati i promijeniti postojeću praksu. Disruptivne tehnologije nisu ništa novo, jer je tehnološki razvoj pun takvih primjera – razvoj fotokopirnog aparata doveo je do gašenja radnih mjesta daktilografa, a proizvodnja računala je ugasila proizvodnju pisanih mašina. Elektronička pošta je u vrlo značajnoj mjeri uzdrmla tradicionalne poštanske usluge. Današnji pametni telefoni objedinili su mnogo uređaja, ali i fizičkih pomagala u jednome uređaju. Oni tako danas objedinjuju telefon, fotoaparat, diktafon, video-kameru, reproduktor zvučnih zapisa (npr. mp3), radio, TV, kalkulator, kompas, GPS navigaciju, rokovnik, kalendar, planer, notes, adresar, foto-album, baterijsku svjetiljku, štopericu, budilicu, ručni sat, brojač učinjenih koraka, povećalo, ravnalo, libelu, zrcalo itd. Uz to danas oni zamjenjuju novac te kreditne ili debitne kartice.

PwC identificira osnovnih osam tehnologija koje će imati vrlo značajan utjecaj na današnje poslovanje te razvoj industrije i usluga. To su 1. umjetna inteligencija (engl. Artificial Intelligence, AI), 2. proširena stvarnost (engl. Augmented Reality, AR), 3. ulančani blokovi (engl. Blockchain, BC), 4. bespilotne letjelice (engl. Drone), 5. internet stvari (engl. Internet of Things, IoT), 6. roboti, 7. virtualna stvarnost (engl. Virtual Reality), te 8. 3D ispis (engl. 3D printing).² Potrebno je spomenuti da neke od ovih tehnologija u sebi objedinjuju niz drugih tehnologija. Tako, na primjer, umjetna inteligencija objedinjuje tehnologije strojnog učenja (engl. Machine Learning, ML), prepoznavanja uzoraka (engl. Pattern Recognition), strojne obrade prirodnog jezika (engl. Natural Language Processing, NLP) itd. Također je potrebno spomenuti da postoje i druge tehnologije koje imaju disruptivni karakter. To su velika količina podataka i podatkovna analitika, tehnologije digitalne participacije poput onih namijenjenih korištenju mnoštva (engl. Crowdsourcing) te igrifikaciji (engl. Gamification) i druge.

U nastavku će se detaljnije analizirati one od identificiranih tehnologija koje će imati utjecaj u području upravljanja dokumentima i zapisima te arhiviranja i čije će uvođenje i korištenje donijeti pozitivne promjene, olakšati svakodnevno poslovanje te rezultirati novom kvalitetom. Tako će biti riječ o umjetnoj inteligenciji, ulančanim blokovima, velikoj količini podataka i tehnologijama digitalne participacije.

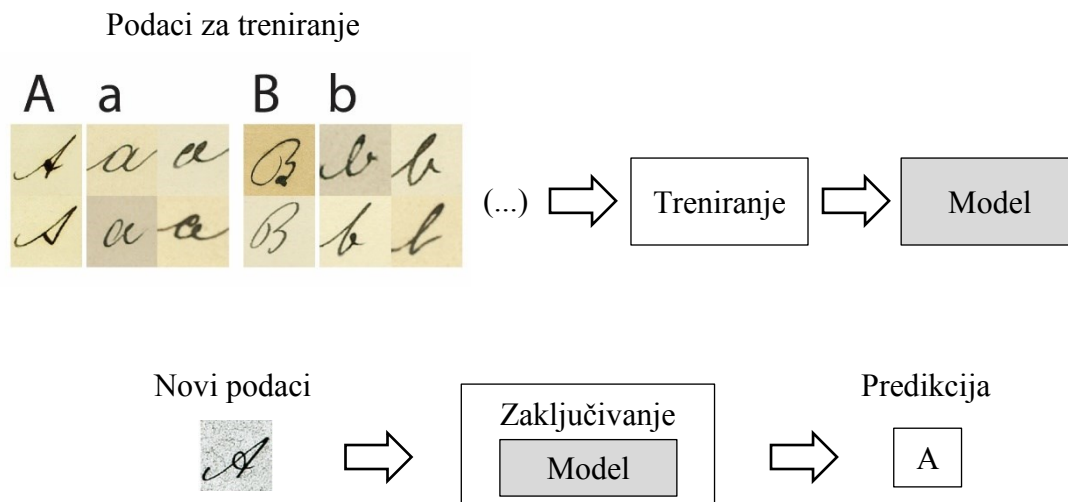
Umjetna inteligencija

Kao što je već rečeno, pojam umjetne inteligencije je vrlo širok i u sebi sadrži druge tehnologije. Trenutno stanje razvijenosti rješenja koja pripadaju ovoj kategoriji je usko specijalizirano. Tako su, primjerice, rješenja za tehničku podršku korisnicima koja rade po principu razmjene poruka između korisnika i institucije (engl. chatbot) uvijek ograničena na jedno područje znanja. Drugi sustavi umjetne inteligencije realizirani su zapravo kao ekspertni sustavi pa pomažu npr. liječnicima u dijagnostici. No, još je daleki put do razvoja umjetne inteligencije koja bi bila namijenjena za više različitih stvari istovremeno. Sustave za umjetnu inteligenciju je potrebno trenirati, tj. učiti razlikovati pa u konačnici i prepoznavati.

¹ Prema: Rouse, Margaret, *Disruptive Technology*, TechTarget, prosinac 2016., URL: <https://whatis.techtarget.com/definition/disruptive-technology> (30.4.2019.)

² The Essential Eight, PwC, 2017., URL: <https://www.pwc.com/gx/en/issues/technology/essential-eight-technologies.html> (30.4.2019.)

Treniranje sustava, odnosno proces strojnog učenja, provodi se tako da se sustavu “pokaže” npr. element na slici, rukom pisano slovo i sl. i onda se poveže s konceptom ili znakom (slika 1). Tako stvoreni model potom se primjenjuje u novim situacijama. Sustav s vremenom postaje sve sigurniji i pouzdaniji, jer svaki novi primjer koji je ispravno prepoznat nadograđuje model koji tako postaje točniji.



Slika 1. Proces strojnog učenja i zaključivanja svojstven sustavima umjetne inteligencije (primjer rukopisa Stjepana Gjurašina (1867.-1936.), hrvatskog botaničara; izvor: Dević Hameršmit, Marta, *Analiza i optičko prepoznavanje rukopisa s herbarijskih etiketa u zbirci Herbarium Croaticum*, diplomski rad, Filozofski fakultet Sveučilišta u Zagrebu, 2018.)

Ovakav proces, primjerice, koristi program Transkribus³ koji pripada kategoriji programa za optičko prepoznavanje rukopisa (engl. Handwritten Text Recognition, HTR), ali i mnogo sofisticiraniji Google Cloud Vision⁴ koji omogućava formiranje vlastitih modela strojnog učenja.

Umjetna inteligencija u arhivima

Objašnjeni koncepti i pristupi koji se koriste u izgradnji umjetne inteligencije svakako mogu imati široku primjenu u arhivima. Za mrežnu komunikaciju s korisnicima, koji vrlo često imaju ista ili slična pitanja, mogu se izgraditi *chatbot* sustavi razmjene poruka između arhiva i korisnika. Nadalje, arhivi prilikom digitalizacije snimaka s mikrofilmova vrlo jednostavno stvore na desetke tisuća slika na mrežnom poslužitelju koje je potrebno obraditi. Sustavi umjetne inteligencije mogu prepoznavati početak i kraj dokumenta, iščitavati relevantne informacije s unaprijed određenih pozicija, provoditi (polu)automatsku klasifikaciju i sl. Ukoliko je, pak, riječ o velikoj količini slika takvi sustavi mogu prepoznavati ljude i objekte na slikama, ili ih barem međusobno razlikovati. Već i to je velika pomoć kod opisivanja, tj. pridruživanja potrebnih metapodataka.

Propuštanje sustava umjetne inteligencije da analizira digitalizirane video-snimke u arhivu BBC-a rezultiralo je pravim malim otkrićem. Naime, unutar snimke jedne nogometne utakmice otkrivena je do tada nepoznata snimka vezana uz Marshallov plan. Tijekom

³ Transkribus, <https://transkribus.eu/Transkribus/> (3.5.2019.)

⁴ Google Cloud Vision, <https://cloud.google.com/vision/> (3.5.2019.)

utakmice je prijenos bio prekinut radi izvanrednih vijesti, no to nije bilo nigdje navedeno u opisu snimke. Vjerojatno je netko pogledao početak i kraj te zaključio da je riječ o cjelovitoj snimci nogometne utakmice. Program umjetne inteligencije analizirao je samo zvučni dio video-snimke te ga semantički analizirao. Izričaj i vokabular se iznenada promijenio kad je program prekinut te se ponovo vratio u okvire očekivanoga po nastavku prijenosa. Tako je detektirano “nešto neočekivano”, tj. sadržaj koji nije bio dio utakmice.

Sustavi umjetne inteligencije mogli bi pomoći prilikom automatiziranja redakteure preuzetih digitalnih zapisa, tj. automatskog prepoznavanja i zacrnjivanja osobnih podataka i sl. To je vrlo složen problem koji ne daje konzistentne rezultate čak niti onda kad različiti arhivisti obrađuju isto gradivo! Da bi sustavi umjetne inteligencije mogli automatizirati redakturu potrebno je mnogo toga, kao što je to prepoznavanje osobnih imena, naziva institucija, geografskih mjesta, kodova (npr. poštanskih i sl.) itd. (engl. Named Entity Recognition, NER) što pripada području strojne obrade prirodnog jezika. Ovaj zadatak je vrlo složen, jer ne ovisi samo o spomenutom prepoznavanju već i o kontekstu. Tako će, na primjer, u dokumentu koji se odnosi na učiteljicu, majku četvero djece koja živi u velikome gradu biti dovoljno ukloniti ime i prezime, dok to u malome mjestu s jednom školom neće biti dovoljno jer bi nedvojbeno bilo jasno o kome je riječ. Izazov provođenja automatske redakteure još uvijek nije kvalitetno riješen pa prostora za unaprjeđenje algoritama umjetne inteligencije svakako ima.

Ovakvi i slični primjeri pokazuju da se sustavi umjetne inteligencije mogu vrlo inovativno primijeniti u arhivima – bilo kao pomoć u efikasnijem obavljanju svakodnevnih zadataka, bilo kao napredni analitički alat.

Ulančani blokovi (blockchain)

Ulančani blokovi su tehnologija koja je najpoznatija u kontekstu kriptovaluta, ali ona ima bitno širu primjenu. Neki joj pogrešno zamjeraju da je za njezino korištenje potrebno mnogo električne energije, ali to je točno samo kod rudarenja kriptovaluta, jer se time dokaz rada (opreme, tj. struja, engl. *proof-of-work*) mijenja za vrijednost (kriptovalute). Korištenje ulančanih blokova bez dokaza rada ne zahtijeva povećanu potrošnju električne energije. Tehnologiju ulančanih blokova omogućuju četiri osnovna koncepta: 1) izračun kriptografskog sažetka (engl. *hash*), 2) Merkleovo stablo (engl. *Merkle tree*), 3) međusobno ulančavanje korijenskih blokova i 4) distribuirani konsenzus (engl. *peer-to-peer consensus*). Potrebno je razumjeti kako se ovi koncepti međusobno nadograđuju i ostvaraju ulančanost blokova. U nastavku će ova četiri koncepta biti objašnjena na vrlo jednostavan način, bez ulaženja u složene detalje, s ciljem razumijevanja ove tehnologije i prepoznavanja njezine potencijalne dobrobiti u kontekstu arhiva.⁵

Kriptografski sažetak je jednosmjerna matematička funkcija koja datoteku bilo koje veličine ili niz znakova bilo koje duljine pretvara u niz znakova točno određene duljine pa će tako, na primjer, primjena algoritma SHA256 za izračun kriptografskog sažetka⁶ za PDF datoteku *Upute suradnicima* za predaju radova za 51. Savjetovanje hrvatskih arhivista koja je objavljena na stranicama Hrvatskog arhivističkog društva rezultirati sljedećim nizom znakova:

⁵ Za više detalja vidjeti: Stančić, Hrvoje. New Technologies applicable to Document and Records Management: Blockchain. *Lligall. Revista Catalana d'Arxivística. Noves perspectives en matèria de gestió documental*. 41 (2018); str. 56-72. URL:

https://www.researchgate.net/publication/332849198_New_Technologies_applicable_to_Document_and_Records_Management_Blockchain (3.5.2019.)

⁶ Za izračun je korišten besplatni internetski servis OnlineMD5, <http://onlinemd5.com/> (3.5.2019.)

7460E378E46985D39CFBF0AFD2E8F974206ED1B06EDF54B782845F1700400B77

S obzirom da je funkcija jednosmjerna nitko ne može na temelju tog niza znakova obrnutim procesom doznati izvorni sadržaj. Osim toga, ako netko, slučajno ili namjerno, napravi i najmanju izmjenu u toj datoteci rezultirajući niz će biti značajno drugačiji. Također, gotovo je potpuno nemoguće da dvije različite datoteke rezultiraju istim nizom. Iz ovoga se može zaključiti da je ovakav niz jedinstven za svaku datoteku baš kao što je to otisak prsta za ljude, pa se i naziva *digitalnim otiskom prsta*. On osigurava provjeru cjelovitosti (integriteta) bilo kojeg zapisa, tj. pruža sigurnost da izvornu datoteku nitko nije promijenio.

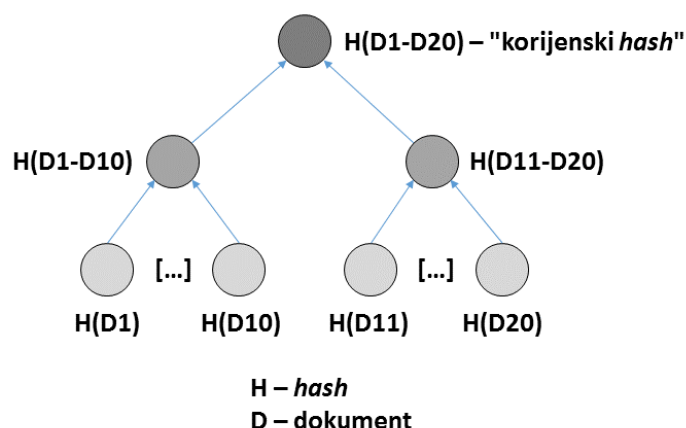
Merkleovo stablo je nazvano po Ralph C. Merkleu koji je 1980. predložio izračun kriptografskog sažetka na temelju više kriptografskih sažetaka, tj. njihovo objedinjavanje na hijerarhijski višoj razini. Ako se uzme primjer cjelovitih zbornika sa 47., 48. i 49. Savjetovanja hrvatskih arhivista koji su dostupni na internetskoj stranici Hrvatskog arhivističkog društva (<https://had-info.hr/publikacije>) i za njih izračunaju kriptografski sažeci koristeći SHA256 to će redom rezultirati sljedećim nizovima znakova:

EA24C7A40926638FB006FBDF7A7B6EB3837BD644F659BCA8A47BB616CCAF7408
B2AF3B7B1EAA55050A11907223EEE8708F1E98A6A2D90F61C2676113873D21EA
BEDF7A51CA345383C5BA7666ACFA5694E93300D87B5EAE122043BB5C2E9495D6

Kad se potom izračuna kriptografski sažetak za koji se kao ulazne vrijednosti uzmu ta tri niza kriptografskih sažetaka tada se umjesto tri dobiva samo jedan niz znakova koji se naziva korijenskim kriptografskim sažetkom (engl. *root hash* ili *top hash*):

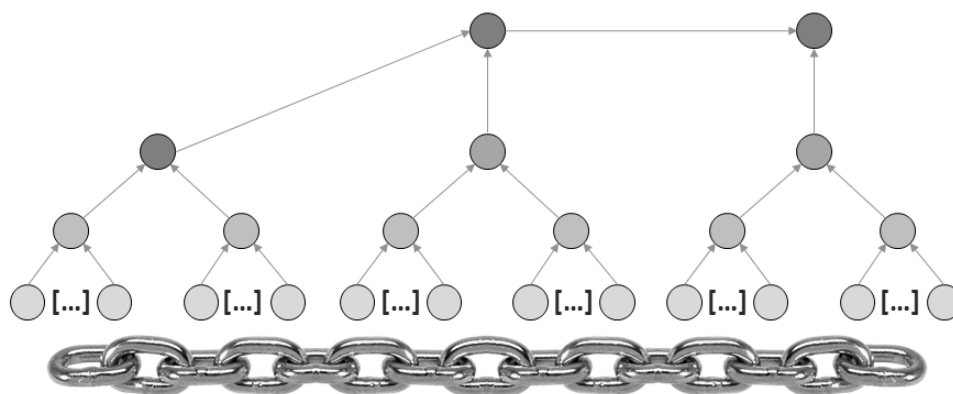
A74DE5FF208CD755391FC5D889361670EDB0BF056F06B0DE1396E2384092FAE3

Drugim riječima, ako dođe do promjene u bilo kojem od tri zbornika, promijenit će se ovaj konačni niz znakova. Sada se ovaj posljednji niz znakova koji objedinjuje prethodna tri može uzeti kao jedan od nizova u sljedećem koraku itd. Tako se svaki put stvara hijerarhijski viša razina koja objedinjuje *hash* vrijednosti niže razine (slika 2).



Slika 2. Koncept Merkleovog stabla

Međusobno ulančavanje korijenskih blokova koncept je koji povezuje korijenske kriptografske sažetke i na temelju njih stvara novi korijenski sažetak za koji, da bi bio ispravan, svi prethodni kriptografski sažeci objedinjeni u prethodne korijenske sažetke moraju biti ispravni (slika 3). Promjena u bilo kojoj od prethodnih datoteka rezultirat će promjenom njezinog kriptografskog sažetka i posljedično svih koji se na njega oslanjaju u svojim izračunima te će tako lanac blokova od tog mjesta nadalje postati nevjerodostojan.



Slika 3. Povezivanje korijenskih sažetaka i stvaranje ulančanih blokova

Distribuirani konsenzus je koncept koji osigurava da svaki blok bude pohranjen na svim računalima koja su uključena u mrežu ulančanih blokova stvarajući takozvanu distribuiranu glavnu knjigu (engl. *distributed ledger*). Tek nakon što kvalificirana većina (50% + 1 računalno) potvrdi da je izračun sljedećeg korijenskog kriptografskog sažetka ispravan taj postaje formalno potvrđen i svi ga zapisuju. Time se osigurava da ne postoji centralno mjesto na kojem se čuvaju dokazi integriteta pa samim time ne postoji niti centralno mjesto na kojem se nešto može neovlašteno promijeniti.

Važno je istaknuti da se u distribuiranu glavnu knjigu pohranjuju samo kriptografski sažeci i eventualno minimalna količina metapodataka za njihovo jednostavnije pronalaženje, a arhivsko gradivo se i dalje čuva u digitalnome arhivu same institucije. Potrebno je obratiti posebnu pažnju da se svi podaci koji podliježu Općoj uredbi o zaštiti podataka (GDPR) ne bilježe u ulančane blokove jer se odande neće moći kasnije izbrisati⁷. Zbog toga je tehnologija ulančanih blokova izuzetno pogodna za provjeru cjelovitosti sigurnosno klasificiranoga gradiva ili onoga koje još nije dostupno javnosti kao i onoga koje je javno dostupno. Važno je i spomenuti da za izračun kriptografskih sažetaka i njihovo bilježenje u *blockchain* niti jedna datoteka ne napušta matični sustav već se dalje šalju samo kriptografski sažeci iz kojih se nikako ne može rekonstruirati sadržaj izvornoga gradiva.

⁷ Više o ovoj problematici vidjeti u: Hofman, Darra; Lemieux, Victoria L.; Joo, Alysha; Alves Batista, Danielle. "The margin between the edge of the world and infinite possibility": Blockchain, GDPR and information governance, *Records Management Journal*, 29(2019), 1/2, str. 240-257. URL: <https://doi.org/10.1108/RMJ-12-2018-0045> (3.5.2019.)

Ulančani blokovi u arhivima

Tehnologiju ulančanih blokova iskoristio je, primjerice, Sirijski arhiv⁸ koji čuva digitalnu AV dokumentaciju o povredama ljudskih prava u Siriji. Za svaki video zapis koji je pohranjen u tom arhivu izračunava se kriptografski sažetak koji se pohranjuje u *blockchain*.⁹ Ako jednom dođe do sudskog procesa moći će se dokazati da izvorno pohranjeni AV zapisi nisu naknadno mijenjani, izrezivani, montirani i sl.

U arhive će sve više ulaziti digitalno potpisani zapisi, a njihovi potpisni certifikati vrijede 2-5 godina što je suviše kratak rok. Ulančani blokovi se mogu iskoristiti za bilježenje valjanosti digitalnog potpisa pa se kasnije, kad njegova valjanost istekne, potvrda valjanosti ubilježena u *blockchain* može koristiti kao valjani digitalni potpis¹⁰.

Tehnologija ulančanih blokova u arhivima može doprinijeti potvrđivanju cjelovitosti nakon što je arhivski zapis stvoren, identiteta potpisnika, autentičnosti zapisa¹¹, potvrđivanju da zapis nije nastao nakon trenutka bilježenja u *blockchain*, redosljeda nastanka zapisa te neporecivosti. Ova tehnologija će uskoro biti standardizirana kao ISO norma¹², ali je već sada moguće vrlo jednostavno postojeći digitalni arhiv ili repozitorij nadograditi *blockchain* funkcionalnošću kojom se izračunate vrijednosti kriptografskih sažetaka registriraju u *blockchain*.¹³

Velika količina podataka

Arhivi su uvijek radili s velikim količinama gradiva i iznalazili su način kako ga obraditi. No, intenzivnijim zahvaćanjem digitalnih zapisa kojih ima eksponencijalno više nego analognih, koji se stvaraju znatno brže i pritom su u različitim formatima (takozvani model 3V – *Volume, Velocity, Variety*) potrebno je početi primjenjivati suvremene pristupe automatizaciji procesa preuzimanja, obrade, pohrane, čuvanja, izlučivanja, arhiviranja te davanja na korištenje. Osim toga termin *big data* vrlo često uključuje aspekte napredne analize, primjerice uočavanja uzoraka i otkrivanja povezanosti, vizualizacije i sl. Ovi aspekti su napose važni, primjerice, kod analize digitalnoga gradiva poput e-pošte, objava na internetskim stranicama ili društvenim mrežama i drugih oblika suvremenoga gradiva. Tako se kod analize poruka e-pošte može analizirati tko je kome slao poruke, je li druga strana odgovarala, može se

⁸ Syrian Archive, <https://syrianarchive.org/en> (3.5.2019.)

⁹ Syrian Archive – Technology, https://syrianarchive.org/en/tools_methods/technology/ (3.5.2019.)

¹⁰ Više o mogućnostima korištenja tehnologije ulančanih blokova u kontekstu problematike očuvanja digitalno potpisanoga gradiva vidjeti u:

Bralić, Vladimir; Kuleš, Magdalena; Stančić, Hrvoje. A model for long-term preservation of digital signature validity: TrustChain. U: Atanassova, Iana i dr. (ur.). *INFUTURE2017: Integrating ICT in Society*. Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia, Zagreb, 2017., str. 89-103. URL:

https://www.researchgate.net/publication/321171227_A_Model_for_Long-term_Preservation_of_Digital_Signature_VValidity_TrustChain (3.5.2019.)

Stanić, Hrvoje i dr. *TRUSTER Preservation Model – Final Report*. Projekt InterPARES Trust, 2018., 43 str.

URL: [https://interparestrust.org/assets/public/dissemination/TRUSTERPreservationModel\(EU31\)-Finalreportv_1_3.pdf](https://interparestrust.org/assets/public/dissemination/TRUSTERPreservationModel(EU31)-Finalreportv_1_3.pdf) (3.5.2019.)

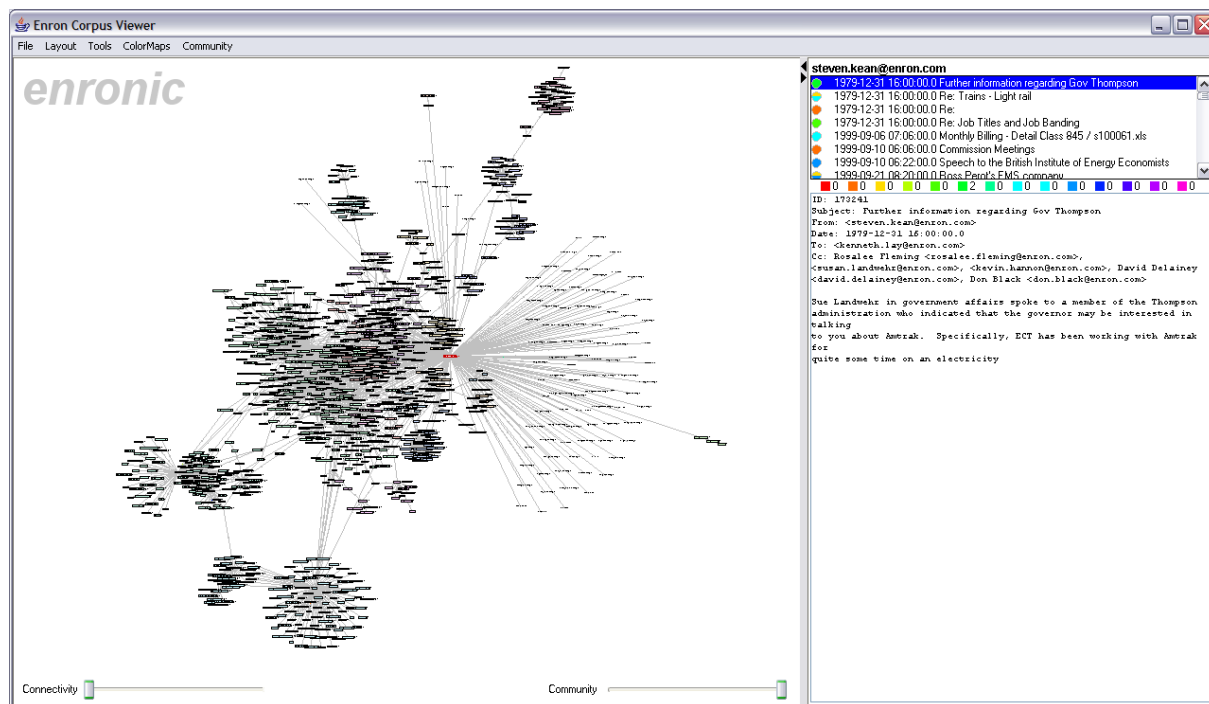
¹¹ Lemieux, Victoria L. Blockchain and Distributed Ledgers as Trusted Recordkeeping Systems. An Archival Theoretic Evaluation Framework. *Future Technologies Conference (FTC) 2017*, Vancouver, Kanada, str. 1-8. URL:

https://www.researchgate.net/publication/317433591_Blockchain_and_Distributed_Ledgers_as_Trusted_Recordkeeping_Systems_An_Archival_Theoretic_Evaluation_Framework (3.5.2019.)

¹² ISO/TC 307 Blockchain and distributed ledger technologies, <https://www.iso.org/committee/6266604.html> (3.5.2019.)

¹³ Npr. EnigioTime, <https://www.enigio.com/> (3.5.2019.)

analizirati tekst poruka, uočavati ključne riječi itd. i primijeniti teoriju grafova, analizu mreža i vizualizaciju kako bi se prikazale grupacije npr. zaposlenika koje su međusobno komunicirale, tko su bili zaposlenici kojima se dostavljalo informacije (tj. šefovi), te tko je što znao u kojem trenutku. Baš tako je analizirano pola milijuna poruka e-pošte u aferi Enron¹⁴ te su detektirane zajednice zaposlenika (slika 4) pri čemu se uspjelo dokazati tko je u kojem trenutku sasvim sigurno morao imati određenu informaciju.



Slika 4. Prikaz društvenih zajednica dobiven analizom razmijenjenih poruka e-pošte među zaposlenicima (Izvor: Heer, Jeffrey. *Exploring Enron*. 2004. URL: <https://homes.cs.washington.edu/~jheer/projects/enron/v1/> (4.5.2019.))

Velike količine podataka u arhivima

Primjer analize poruka e-pošte samo je jedan od mnogih primjera analize i vizualizacije velikih količina podataka koje se danas pojavljuju u arhivima. Pružanje pristupa velikim količinama digitalnoga gradiva, ali i stvaranje programskih rješenja za njihovu analizu, doprinijet će kvalitetnijim istraživačkim rezultatima zbog mogućnosti zahvata i obrade takve količine gradiva koja, kad bi bilo riječ o gradivu na papiru, naprosto ne bi bila moguća. Ovakvi trendovi također jako dobro pokazuju koja znanja moraju imati kako oni koji rade u arhivima tako i povjesničari te svi ostali koji žele analizirati digitalno arhivsko gradivo. To ujedno predstavlja i priliku arhivima da se nametnu kao mjesta koja pružaju ovakve usluge te nude suvremeni pristup analizi gradiva koji može postati osnova za uočavanje trendova i donošenje kvalitetnih odluka.

¹⁴ Više o aferi Enron u: Segal, Troy. Enron Scandal: The Fall of a Wall Street Darling. *Investopedia*, 1.5.2019., URL: <https://www.investopedia.com/updates/enron-scandal-summary/> (4.5.2019.)

Tehnologije digitalne participacije

Tehnologije digitalne participacije koriste informacijske sustave potpomognute mnoštvom. Riječ je o sustavima koji se naslanjaju na današnju kulturu sudjelovanja u stvaranju digitalnih sadržaja i koji omogućavaju da se poslovi za koje bi postojećim zaposlenicima u arhivima trebalo mnogo vremena obave u relativno kratkom vremenu. Tako se, na primjer, korištenjem suradničkih sustava može opisati veliki broj digitaliziranih fotografija za koje bi zaposlenicima trebalo nekoliko desetaka godina. Korisnici, odnosno zainteresirana javnost spaja se na mrežnu stranicu i otvara im se neka, do tada neopisana, fotografija. Kroz sučelje koje se može sastojati od prostora za slobodan upis teksta ili, bolje, strukturiranih polja za upis podataka, korisnici opisuju što se nalazi na fotografiji, je li ona crno-bijela ili u koloru, prepoznaju li koje osobe, predmete, lokaciju koju fotografija prikazuje i sl. Mnoštvo korisnika će trebati opisati mali broj fotografija, ali će se ovim pristupom ukupno obraditi veliki broj. Drugi primjer koji zorno prikazuje mogućnosti ovakvog pristupa je projekt arhiva Sveučilišta Stanford za prijepis rukopisnoga gradiva – pisama i dokumenata¹⁵. Kod transkripcije korisnicima je poželjno omogućiti odabir gradiva kojem se žele posvetiti, jer je mnogima određena tematika zanimljivija, a i tada se tijekom vremena naviknu na rukopis neke osobe. Takvi sustavi uobičajeno u jednom dijelu prikazuju digitalizirani izvornik dok je u drugome dijelu prostor za prijepis. Pritom je moguće pratiti ukupni status prepisanosti (slika 5). Jedan od najpoznatijih primjera prepisivanja sadržaja potpomognutih mnoštvom je *reCaptcha*¹⁶ – sustav koji s jedne strane provjerava prijavljuje li se na neki sustav čovjek (tzv. Turingov test) dok s druge strane pomaže u digitalizaciji knjiga, novina i sl. koje se ne mogu obraditi OCR-om na način da iz slike stranice izrezuje riječi pa tako korisnici prijavom u neki sustav istovremeno digitaliziraju, tj. prepisuju knjige riječ po riječ (slika 6). Tako su prepisana stara izdanja novina The New York Times pa je sada dostupno 13 milijuna članaka iz perioda od 1851 do danas¹⁷.

Tehnologije digitalne participacije uobičajeno se dijele na one koje koriste mnoštvo (tada se uobičajeno koristi engleski termin *crowdsourcing*) i one koje koriste iste principe, ali ih oblikuju kao računalne igre (tada se uobičajeno koristi engleski termin *gamification*). Igrifikacija, dakle, koristi način razmišljanja i mehaniku primjerene računalnim igrama kako bi se korisnici angažirali te za neke odrađene radnje dodjeljuje bodove, zvjezdice, napredniju razinu i sl. pa tako korisnici napreduju ili se međusobno natječu. Nacionalni arhiv Australije u okviru projekta *arcHIVE* tako ima objavljenu rang-listu osoba koje su prepisale najviše tekstualnih zapisa, a stečeni bodovi se mogu zamijeniti za virtualne bedževe (početnik, entuzijast, ekspert) odnosno za primjerke izdanja arhiva ili postere.¹⁸

¹⁵ *Archives launches transcription crowdsourcing project*. 12.10.2017., URL: <http://library.stanford.edu/blogs/special-collections-unbound/2017/10/archives-launches-transcription-crowdsourcing-project> (10.5.2019.)

¹⁶ *reCaptcha*, URL: <https://www.google.com/recaptcha/intro/v3.html> (10.5.2019.)

¹⁷ *reCAPTCHA*. *Wikipedia*, 1.5.2019., URL: <https://en.wikipedia.org/wiki/ReCAPTCHA> (10.5.2019.)

¹⁸ *Projekt arcHIVE*. National Archives of Australia, <http://transcribe.naa.gov.au/> (10.5.2019.)

Facsimile	Transcription
	to Mr and Mrs Stanford that I should do so. Dr Harkness has done me the honor to invite me to lecture at the Academy of Sciences, which I have accepted with the understanding that your university has the prior claim. Will you kindly let me know what if any evening will be agreeable for the lecture to be given, and I will immediately inform Dr Harkness who will make his announcement accor

21 pages: 0% indexed, 43% transcribed, 57% needs review

Slika 5. Sustav za transkripciju i praćenje statusa prepisanosti (izvor: Stanford University Archives, <https://www.fromthepage.com/stanforduniversityarchives/people>)

The Norwich line steamboat train, from New-London for Boston, this morning ran off the track seven miles north of New-London.

morning

morning overlooks

Type the two words:

Slika 6. reCaptcha – prijepis knjiga riječ po riječ (izvor: Cyclifier, 1.10.2006., URL: <https://www.cyclifier.org/project/recaptcha/>)

Tehnologije digitalne participacije u arhivima

Tehnologije digitalne participacije, bez obzira je li rješenje za aktivnosti potpomognute mnoštvom nadograđeno principima primjerenim računalnim igrama ili ne, mogu znatno pomoći kako u obavljanju svakodnevnih poslova u arhivima tako i u međusobnom zabavnom natjecanju koje svima donosi korist. Takva rješenja mogu iskoristiti samo zaposlenici ili ona mogu biti namijenjena korisnicima i drugoj zainteresiranoj javnosti ili, pak, objema skupinama. Najvažnije je da se ovim pristupom može obraditi mnogo materijala – onoliko koliko se u arhivu ne bi moglo obraditi desetljećima. Naravno, najvažnije je pritom osmisliti dobru marketinšku kampanju, ispričati motivirajuću priču, detektirati zainteresirane grupe korisnika, eventualno pronaći sponzore itd. No, sve to je ipak daleko jednostavnije nego obraditi veliku količinu gradiva. Ipak, pritom ne treba zaboraviti da će sve rezultate ovakvih aktivnosti netko trebati temeljito pregledati pa je angažman arhiva neophodan.

Zaključak

Ulaskom digitalnih zapisa u arhive, količina gradiva koju je potrebno obraditi eksponencijalno se povećava. I ne samo to – digitalno gradivo je tehnički drugačije pa je potrebno uvoditi novu tehnologiju, nove principe djelovanja, zapošljavati nove vrste stručnjaka, ali i mijenjati ustaljene načine financiranja. Dok se očuvanje analognoga gradiva moglo financirati kroz projekte, npr. mikrofilmiranja, jer su oni, kad su jednom dovršeni, osiguravali dugotrajno očuvanje, očuvanje digitalnoga gradiva potrebno je financirati kroz programe, dakle kroz trajno financiranje, a ne kroz kratkotrajno ili jednokratno. S obzirom da je mnogo jednostavnije dobiti novac za projekte, jer su oni ograničenoga trajanja i s vrlo jasnim ciljem, novac za programe i trajno financiranje arhivi moraju pronaći u svojem redovitom budžetu. S obzirom da rijetko dolazi do povećanja budžeta i znatnijeg povećanja broja zaposlenih, potpuno je jasno da se arhivi, pred koje se postavlja zadatak preuzimanja i dugoročnog očuvanja digitalnih zapisa, nalaze u svojevrsnoj krizi. Zbog toga se trebaju okrenuti novim, disruptivnim tehnologijama koje im mogu pomoći da svoje zadaće obavljaju jednostavnije, efikasnije i stvore nove vrijednosti koje bi im trebale pomoći da dođu do potrebnih financijskih sredstava.

Literatura

Archives launches transcription crowdsourcing project. 12.10.2017., URL:

<http://library.stanford.edu/blogs/special-collections-unbound/2017/10/archives-launches-transcription-crowdsourcing-project> (10.5.2019.)

Bralić, Vladimir; Kuleš, Magdalena; Stančić, Hrvoje. A model for long-term preservation of digital signature validity: TrustChain. U: Atanassova, Iana i dr. (ur.). *INFuture2017: Integrating ICT in Society*. Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia, Zagreb, 2017., str. 89-103. URL:

https://www.researchgate.net/publication/321171227_A_Model_for_Long-term_Preservation_of_Digital_Signature_VValidity_TrustChain (3.5.2019.)

Dević Hameršmit, Marta. *Analiza i optičko prepoznavanje rukopisa s herbarijskih etiketa u zbirci Herbarium Croaticum*. diplomski rad, Filozofski fakultet Sveučilišta u Zagrebu, 2018., URL:

<http://darhiv.ffzg.unizg.hr/id/eprint/10261/1/Devic%20Hamersmit%2C%20Marta%2C%20diplomski%20rad%2C%20zavrsna%20verzija.pdf> (3.5.2019.)

- Duff, Wendy M.; Haskell, Jessica. New Uses for Old Records: A Rhizomatic Approach to Archival Access. *The American Archivist*, proljeće/ljeto 2015., vol. 78, br. 1, str. 38-58, URL: <https://americanarchivist.org/doi/pdf/10.17723/0360-9081.78.1.38> (10.5.2019.)
- Heer, Jeffrey. *Exploring Enron*. 2004. URL: <https://homes.cs.washington.edu/~jheer/projects/enron/v1/> (4.5.2019.)
- Hofman, Darra; Lemieux, Victoria L.; Joo, Alysha; Alves Batista, Danielle. "The margin between the edge of the world and infinite possibility": Blockchain, GDPR and information governance, *Records Management Journal*, 29(2019), 1/2, str. 240-257. URL: <https://doi.org/10.1108/RMJ-12-2018-0045> (3.5.2019.)
- Lemieux, Victoria L. Blockchain and Distributed Ledgers as Trusted Recordkeeping Systems. An Archival Theoretic Evaluation Framework. *Future Technologies Conference (FTC) 2017*, Vancouver, Kanada, str. 1-8. URL: https://www.researchgate.net/publication/317433591_Blockchain_and_Distributed_Ledgers_as_Trusted_Recordkeeping_Systems_An_Archival_Theoretic_Evaluation_Framework (3.5.2019.)
- Merkle, Ralph C. *Protocols for public key cryptosystems*. *IEEE Symposium on Security and Privacy*, 1980., str. 122-134
- Projekt archIVE. National Archives of Australia, <http://transcribe.naa.gov.au/> (10.5.2019.)
- Rouse, Margaret. *Disruptive Technology*. TechTarget, prosinac 2016., URL: <https://whatis.techtarget.com/definition/disruptive-technology> (30.4.2019.)
- Stančić, Hrvoje. New Technologies applicable to Document and Records Management: Blockchain. *Lligall. Revista Catalana d'Arxivística. Noves perspectives en matèria de gestió documental*. 41 (2018); str. 56-72, URL: https://www.researchgate.net/publication/332849198_New_Technologies_applicable_to_Document_and_Records_Management_Blockchain (3.5.2019.)
- Stančić, Hrvoje i dr. *TRUSTER Preservation Model – Final Report*. Projekt InterPARES Trust, 2018., 43 str. URL: [https://interparestrust.org/assets/public/dissemination/TRUSTERPreservationModel\(EU31\)-Finalreportv_1_3.pdf](https://interparestrust.org/assets/public/dissemination/TRUSTERPreservationModel(EU31)-Finalreportv_1_3.pdf) (3.5.2019.)
- The Essential Eight*. PwC. 2017., URL: <https://www.pwc.com/gx/en/issues/technology/essential-eight-technologies.html> (30.4.2019.)

Disruptive Technologies in the Archives

Summary

The paper gives introduction to the current state of affairs in the archives and detects the need for usage of the cutting-edge technologies. Some of them will bring significant changes or will completely change standard operative procedures. Therefore, those technologies are called the disruptive technologies. The paper proceeds to explain the term in detail and identifies the essential eight technologies that will have the disruptive impact on the standard archival operations. Further, the technologies are grouped and explained in detail. The artificial intelligence (AI), blockchain and distributed ledger technologies, big data and the technologies for digital participation (crowdsourcing and gamification) are analysed. First, the explanation of how each technology works is given. Then, the example projects or applications are detailed and explained. Finally, the possibilities for usage and implementation in the archives are discussed. The paper concludes with the comparison of digital preservation in the analogue and digital form, discussion about the approach to financing, and provides argumentation why the archives need to engage both themselves and their communities by wider usage of disruptive technologies.

Key words: *archive, digital records, artificial intelligence, blockchain, big data, crowdsourcing, gamification*