

Dimitrij Reja¹

POSSIBILITIES OF SYNERGISTIC EFFECTS OF ARTIFICIAL INTELLIGENCE AND METADATA

Abstract

Purpose: Review the state of art intelligence in the public sector.

Method/approach: A review of the latest published articles in the field of artificial intelligence and metadata.

Results: An overview of concrete examples of the use of artificial intelligence with metadata and their actual use with all the advantages and disadvantages.

Conclusions/findings: In the process of processing archival material using metadata, the artificial intelligence or tools used in various procedures help us significantly.

Key words: artificial intelligence, metadata, archival sciences, public sectorUvod

POSSIBILITÀ DI EFFETTI SINERGICI DELL'INTELLIGENZA ARTIFICIALE E DEI METADATI

Sintesi

Scopo: rivedere lo stato dell'arte dell'intelligence nel settore pubblico. **Metodo/approccio:** una rassegna degli ultimi articoli pubblicati nel campo dell'intelligenza artificiale e dei metadati.

Risultati: una panoramica di esempi concreti di utilizzo dell'intelligenza artificiale con i metadati e il loro effettivo utilizzo con tutti i vantaggi e gli svantaggi.

Conclusioni/risultati: Nel processo di elaborazione del materiale d'archivio tramite metadati, l'intelligenza artificiale o gli strumenti utilizzati nelle varie procedure ci aiutano in modo significativo.

Parole chiave: intelligenza artificiale, metadati, scienze archivistiche, settore pubblico

1 Dimitrij Reja, študent doktorskega študija 3. stopnje arhivskih znanosti na Alma Mater Europaea, ECM, e-mail: dimitrij.reja@gmail.com

MOŽNOSTI SINERGIJSKIH UČINKOV UMETNE INTELIGENCE IN METAPODATKOV

Izvleček

Namen: Pregledati stanje umetne inteligence na področju arhiviranja/metapodatkov v javnem sektorju.

Metoda/pristop: Pregled zadnjih objavljenih člankov na področju umetne inteligence in metapodatkov.

Rezultati: Pregled konkretnih primerov uporabe umetne inteligence z metapodatki ter njihova dejanska uporaba z vsemi prednostmi in slabostmi.

Sklepi/ugotovitve: V procesu obdelave arhivskega gradiva z uporabo metapodatkov nam umetna inteligenca ali orodja, ki jih uporabljajo, pri različnih postopkih znatno pripomorejo pri vsakodnevem delu.

Ključne besede: umetna inteligenca, metapodatki, arhivske znanosti, javni sektor

Veliko pred definiranjem pojma Big Data² so arhivisti imeli uveljavljeno splošno mersko enoto, npr. km - kilometer za količino arhivskih podatkov. Z uporabo svetovnega spleta in novih digitalnih tehnologij je število digitalnih podatkov pričelo skokovito naraščati. V osemdesetih in devetdesetih letih dvajsetega stoletja in v začetku enaindvajsetega stoletja, se prakse arhivistov s pričo pojava novih tehnologij niso kaj dosti spremenile. Ena izmed bistvenih sprememb je bil prehod iz kartičnega sistema iskanja v digitalni način iskanja v arhivih. V nadaljevanju so nastajali najrazličnejši spletni katalogi, s pomočjo katerih so postali podatki bolj dostopni (Moss et al., 2018). Vendar je kljub temu bilo vrednotenje, razvrščanje podvrženo ročnemu delu, torej arhivist je bil tisti, ki je vse te kilometre gradiva pregledal in razvrščal.

Z množičnim generiranjem digitalnih podatkov se količina arhivskega gradiva znatno povečuje in človek (arhivist) ne uspe obdelati vsega v enakem časovnem okviru in z enako natančnostjo, kot pred prihodom digitalne revolucije. Imeli bomo vedno večje količine izvirne digitalne vsebine, ki jo bo treba obdelati, indeksirati, analizirati in s pomočjo le teh izračunati sekundarne podatke. Ob procesu obdelave izvirne digitalne vsebine skozi čas bo nastajalo tudi vedno več podatkov o podatkih ali metapodatkov.

Digitalna preobrazba spreminja stare in nove arhive v podatke. Lahko bi bolj natančno zapisali, da jih spreminja v bolj dostopne in med seboj povezljive podatke. Veliko se v zadnjem času govori o digitalni preobrazbi, vendar kaj to pravzaprav predstavlja? Meni osebno najbolj plastično prikazuje digitalno preobrazbo naslednja definicija: Digitalna preobrazba je sprememba načina uporabe digitalnih tehnologij za razvoj novih digitalnih poslovnih modelov, ki pomagajo ustvarjati nove vrednosti in hkrati povečuje obstoječe vrednosti podjetij (Verhoef et al., 2021).

Posledično lahko povzamemo, da v današnjem času in ob velikih količinah podatkov, arhivisti pri svojem delu potrebujejo pomoč digitalnih tehnologij. Arhivi so postali svojevrsne Big Data organizacije in kot Big Data organizacije je potrebno v del svojega poslovanja vključiti tudi umetno inteligenco, ki se v večini izraža v strojnem učenju (Colavizza et al., 2021, str. 22).

2 Big data je pojem, ki se nanašajo na podatke, ki so tako veliki, hitri ali zapleteni, da jih je težko ali nemogoče obdelati s tradicionalnimi metodami

1. VPLIV TEHNOLOGIJE NA DELO ARHIVISTOV

Visoko kakovostni metapodatki igrajo izjemno vlogo pri izboljšanju rezultatov iskanja. Toda proces prepričevanja ljudi, da dosledno vnašajo kakovostne metapodatke, za večino predstavlja Sizifovo delo. Ena izmed možnosti rešitve nastale zagate je tudi samodejno ustvarjanje metapodatkov s pomočjo umetne inteligence. Avtomatizacija dela poslovanja je neizogibna in razumeti moramo, kako uporabljati stroje, njihove prednosti in slabosti ter kdaj mora človek posredovati (Seles, 2019).

Nedvoumno lahko zatrdimo, da se s prihodom novih digitalnih tehnologij delo arhivista in način dela v samem arhivu korenito spreminja. Arhivisti se bodo morali pri svojem delu vse bolj opirati na novodobne tehnologije in ob tem bodo primorani spremeniti sam pogled na primarno vsebino. Arhivi postajajo vse bolj podobni Big Data centrom, v katerih je rudarjenje ali podatkovno rudarjenje³ primarnega pomena. Ob tem ne smemo pozabiti na sodelovanje z ostalimi specialisti informacijskih tehnologij npr. administratorji za spletne strani, administratorji podatkov baz, ipd. (Moss et al., 2018).

Umetna inteligenca v službi arhiva je bila do sedaj v grobem razdeljena na sisteme, ki temeljijo na pravilih, statističnih modelih in modelih z globokim učenjem⁴ (Bunn, 2020). Sistemi, ki temeljijo na pravilih, so dokaj enostavni in vpogled v njihovo delovanje je preprost. Statistični modeli v večini primerov delujejo in predvidevajo s pomočjo določene stopnje zaupanja. V primeru modelov z globokim učenjem so rezultati pridobljeni z uporabo črne skrinjice⁵, kar rezultira, da je zelo težko natančno interpretirati delovanje modela (Rollan, 2019, v Bunn, 2020, str. 144). Ob tem trčimo na problem, če smo do sedaj natančno vedeli, kako procesi delujejo, imamo sedaj z uporabo umetne inteligence in globokim učenjem težavo. Procesov znotraj modela ne poznamo in ga ne znamo opisati. Postavljeni smo pred dejstvo, da smo prisiljeni, slepo zaupati modelom, ki uporabljajo črno skrinjico.

Za rešitev nastalega problema je v zadnjih letih vzniknil nov model umetne inteligence, eXplainable Artificial Intelligence ali XAI⁶. Razložljiva umetna inteligenca deluje prav tako na principu črne skrinjice, vendar kot že povzema samo ime, rezultati, procesi in razlogi so razumljivi ljudem, kar je bistvenega pomena. Strinjamo se, da sama interpretacija rezultatov ni dovolj, če želimo zaupati metodam črne skrinjice, potrebujemo t. i. pojasnjevalne modele, ki so sposobni povzeti razloge, zakaj se je nevronska mreža odločila, kot se je in s tem zagotovo pridobi določeno stopnjo našega zaupanja (Gilpin et al., 2018).

Do sedaj smo govorili le o obdelavi podatkov, kakšno je stanje s teorijo? Vsekakor moramo upoštevati tudi spremembe v arhivski znanosti kot takšni. Izhajati moramo iz stališča, da nam nove informacijske tehnologije omogočajo razkorak med uveljavljenimi arhivskimi načeli in praksami ter zmožnostmi tehnologije.

Avtorica Kenneth Thibodeau v svojem delu z naslovom »Breaking Down the Invisible Wall to Enrich Archival Science and Practice«, plastično opisuje razkorak med ustoličenimi načeli in prakso z novodobno tehnologijo. Argumenta za razkorak sta po avtoričinem

3 Podatkovno rudarjenje je sistematično iskanje informacij v veliki količini podatkov.

4 Globoko učenje je področje umetne inteligence, ki obravnava globoke večslojne nevronske mreže

5 Črna skrinjica je naprava, sistem ali program, ki vam omogoča, da vidite vhod in izhod, vendar ne daje vpogleda v procese in delovanje med njimi.

6 XAI - eXplainable Artificial Intelligence, je umetna inteligenca, katere rezultati so razumljivi ljudem

mnenju dva. Prvič, uveljavljeni koncepti v arhivistiki niso niti ustrezni, niti neustrezni za številne vrste digitalnih informacij niti za vse načini uporabe informacij. Drugič, tehnologija ponuja možnosti za izboljšanje arhivske prakse, vendar lahko neprimerni koncepti ovirajo zmožnost razvoja (Thibodeau, 2016).

Odveč je bojazen, da bi arhivistom umetna inteligenca prevzela delo ali še bolj rigorozno razmišljanje, da bi arhivistom zmanjkalo dela. Theimer (2018) razpravlja, da bodo v bližnji prihodnosti, kjer je uporaba umetne inteligence prisotna vsakodnevno, če že niso to tudi sedaj, arhivisti postali »mojstri podatkov« ali kot je še dodala avtorica »podatkovni znanstveniki«. Arhivisti bodo morali razumeti, kako uporabiti orodja in tehnike Big data do podatkov, ki jih generiramo s svojim delovanjem.

NEKAJ O UMETNI INTELIGENCI

Umetna inteligenca ne predstavlja Arnolda Schwarzeneggerja v vlogi Terminatorja T1000, vse prej kot takšna pošast, je popolna inteligentna naprava, prilagodljivo, racionalno sredstvo, ki zaznava svoje okolje in sprejema odločitve, ki povečajo njegove možnosti za doseg zastavljenih ciljev (iPROM, 2022).

Splošno prepričanje, da je umetna inteligenca nekaj oddaljenega v prihodnost, vendar temu ni tako. Umetna inteligenca je prisotna od leta 1940 naprej in se vztrajno razvija. Do danes je vtkana v vse pore našega življenja, npr. finančna, storitvena industrija (Han, 2021).

Arhivisti imamo zelo pomembno vlogo pri svetovanju ob ustvarjanju in hrambi digitalnih zapisov in podatkov. Upoštevati moramo, da bodo danes ustvarjeni digitalni podatki zelo kmalu obdelani s strani umetne inteligence. Postavlja se vse več vprašanj, kakšen nasvet bi danes podali glede ustvarjanja in hrambe generiranih podatkov s pomočjo umetne inteligence?

Orodja umetne inteligence v grobem lahko razdelimo na:

1. **Nevronske mreže**, ki posnemajo delovanje človeških možganov. Nevroni so celice, ki so med seboj povezani z nitkami ali sinapsami. Nevroni si med seboj, preko sinaps, pošiljajo električne dražljaje. Za sinapse je značilno, da se med seboj razlikujejo po električni prevodnosti, ki se med učenjem spreminja. Torej je znanje, pridobljeno med učenjem, nakopičeno v sinapsah oziroma v njihovi prevodnosti. Če je vsota signalov, ki prispejo v posamezni nevron preko sinaps, dovolj velika, pride do vžiga posameznega nevrona. To pomeni, da ta nevron pošlje na svoj izhod signal, ki se preko sinaps prenese naprej v ostale nevrone (Ploj, 2017).
2. **Obdelava naravnega jezika** - zmožnost razumevanja besedila in izgovorjenih besed na približno enak način kot pri ljudeh.
3. **Statistično strojno učenje** - analiza podatkovnih nizov s pomočjo uporabe statističnih modelov.
4. **Globoko učenje** - gre za učenje enakih večslojnih nevronske mreže, vendar s sodobnejšimi učnimi postopki.

NEKAJ O KONKRETNIH PRIMERIH

Dejstvo je, da se arhivisti srečujemo tudi z nestrukturiranimi in ne kategoriziranimi zapisi. Slednjih zapisov je vedno več. Za plastičen primer si poskusimo predstavljati, koliko digitalnih dokumentov se producira ob javnem naročilu; vsa nešteta komunikacija, popravki, verzije dokumentov, elektronska sporočila ipd. Vsekakor morajo biti dokumenti urejeni v zadevi z repom in glavo, vendar je velikokrat vse prej kot bi moralo biti. Umetna inteligenca ima moč in zmožnost arhivistom pomagati pri obvladovanju takšnih količin podatkov. Kako izluščiti in razlikovati ustrezne podatke v vsej množici raznovrstnih

in zapletenih oblik zapisov in vsebine podatkov (Makhlouf Shabou et al., 2020), to je eno izmed večnih vprašanj arhivistov.

V primeru elektronskih sporočil, ki so po mojem prepričanju še vedno najboljši kanal v poslovnem komuniciranju, je v letu 2021 samo eni minuti odposlanih 197,6 mio elektronskih sporočil (Jenik, 2021). V našem primeru moramo razlikovati, katera sporočila imajo potencialno poslovno vrednost in katera ne. S pomočjo avtomatičnega kategoriziranja elektronskih sporočil sta (Vellino in Alberts, 2016) izvedla tudi študij primera. V zaključku sta avtorja zapisala, da je ocenjevanje elektorskih sporočil zahtevna naloga, čeprav je nujno potreben, saj je potrebno v poslovni komunikaciji ločiti zrno od plevla.

Problem pri veliki količini elektronskih podatkov se izkaže tudi na naslednjem primeru. Arhiv zvezne države Viktorije v Avstraliji ima v svojih arhivih 67.000 trakov ali 28 PB elektronskih sporočil. Od leta 1990 je v uporabi sistem IBM Lotus Notes za vse državne institucije in v 20 letih uporabe se je nabrala zgoraj omenjena količina podatkov (Rolan et al., 2019). Logično lahko sklepamo, da obseg podatkov ne more biti učinkovito uporabljen ali še bolj plastično zelo težko si predstavljamo, kako s preprostim iskalnim nizom iščejo podatke. Vpeljava umetne inteligence za takšne namene je povsem upravičena in zaželena. Upravičenost vpeljave umetne inteligence ali orodij, s katerimi bi zmanjšali podvojene, potrojene vnose, izločili neustrezne in avtomatično dodali bolj uporabne metapodatke je tudi prikazala raziskava, ki so jo izvedli avtorji v članku z naslovom »More human than human? Artificial intelligence in the archive«.

Zanimiv primer uporabe umetne inteligence je tudi prepoznavanje obrazov v digitalnih arhivih knjižnic, muzejev in ostalih kulturnih ustanovah. Raziskovalni projekt avtorjev (Bakker et al., 2020) je bil izveden v prvi vrsti, da s pomočjo testiranja različnih rešitev prikaže dejanske prednosti in pomanjkljivosti le teh rešitev. Vzrok za avtomatično klasifikacijo fotografij v različnih formatih je povsem trivialen, namreč metapodatki o fotografijah so v večini zelo osiromašene. In razlog? Povsem preprost., Če fotografije ob samem vnosu oz. prejemu niso bile opremljene z vsemi potrebnimi podatki, se na njih povsem pozabi. S pomočjo umetne inteligence in ustrezne zbirke podatkov so v veliki meri zelo uspešno popravili ali dodali relevantne meta podatke. (Chabin, 2020) obravnava dogajanje med veliko nacionalno debato. Analiza pripomb in predlogov francoskih državljanov med veliko nacionalno debato se je odvijala med protesti rumenih jopičev in je vsekakor zelo zanimiv projekt. Projekt francoske vlade je imel za enega od ciljev pridobiti odziv državljanov. V ta namen so organizirali različne načine participiranja (preko spleta, pisnih predlogov ter audio posnetkov pogovorov). Časovna komponenta je bila zelo pomembna zaradi razgrete situacije v državi. Izbira orodij, ki so s pomočjo umetne inteligence ustrezno klasificirala in ovrednotila prispevke, je bila popolnoma logična. Ob reviziji projekta je bilo poudarjeno, da bi v mnogih primerih prepoznavanja rokopisov, hitreje izvedli postopek, če bi rokopise enostavno prebrali. Na tej točki se lahko ustavimo in potegnemo črto ter sklenemo, da umetna inteligenca potrebuje posredovanje človeka. Kot smo na začetku članka zapisali, s tem, ko vemo kaj se dogaja se nam poveča določena stopnja zaupanja v umetno inteligenco.

METAPODATKI, PODATKI O PODATKIH V DIGITALNI DOBI

V digitalni dobi, kjer kraljujejo nove digitalne tehnologije, so arhivi pred težkimi izzivi. Zagotavljati morajo nove načine dostopa do digitalnih arhivskih zapisov in nove možnosti nadaljnje računalniške analize. Ob tem ne smemo pozabiti, da moramo arhivisti razmišljati daljnosežno. Ne smemo se zadovoljiti le z rešitvijo trenutnih problemov, nastavitve je treba vzvode, s katerimi bodo zanamci enostavno in hitro nadgradili obstoječe stanje.

Arhiviste že tradicionalno zelo zanima izvor njihovih zbirk tj. od kod prihajajo zgodovinski zapisi in kdo jih je ustvaril. V bližnji prihodnosti ali pa že sedaj nas enako zanima izvor metapodatkov.

Nacionalni arhiv Velike Britanije je že v letu 2018 podal predlog novega modela metapodatkov, ki se nanaša na delovanje državnih arhivov. Poudarek je na delovanju državnega arhiva, saj se v nadaljevanju določeni metapodatki nanašajo izključno na državni arhiv.

Kakovostni metapodatki, ki so shranjeni v digitalnih repozitorijih, so bistvenega pomena za njihovo nadaljnjo obdelavo in interoperabilnost⁷ (Ochoa in Duval, 2009). V primeru, da so metapodatki slabi, je otežen dostop do digitalnih vsebin. Zelo preprosto lahko to prikažemo na primeru. Velika razlika je, če opišemo metapodatek »naslov« z naslovom predavanja npr. ATLANTI 2021- Dimitrij Reja, Sinergijski učinki metapodatkov na umetno inteligenco ali samo ATLANTI, str. 25, Predavanje 1.

Nov predlog modela so poimenovali »The seven pillars of metadata«, kar bi lahko prevedli kot sedem stebrov metapodatkov. Pri oblikovanju novega modela so se naslonili na značilnosti in izvor metapodatkov. Takšen pristop je malce drugačen od dosedanjega, ki temelji izključno na funkciji metapodatkov. Razlog za takšen odklon od uveljavljenega načina definiranja izhaja iz zavedanja, da posamičen metapodatek lahko vsebuje več funkcij. V prihodnosti bodo zagotovo nastale nove funkcije metapodatkov, ki pa jih danes še ne poznamo (The National Archives, 2018).

Model sedmih stebrov metapodatkov je razdeljen na podedovane, primarne, sekundarne, dodatne, izpeljane, kontrolne in metapodatke.

Podedovani metapodatki

Podedovani metapodatki so vsi tisti podatki, ki se nanašajo na kontekstualne metapodatke. V sklop kontekstualnih metapodatkov lahko prištejemo revizijske sledi o digitalnem dokumentu.

Primarni metapodatki

Primarni metapodatki so neločljivi del digitalnih dokumentov; v ta sklop sodijo npr. ime datoteke, format datoteke, dimenzije, datum in čas ipd.

Sekundarni metapodatki

Sekundarni metapodatki vključujejo dodatne informacije o digitalnem dokumentu. Dodatne informacije so lahko kreirane ročno ali samodejno⁸. Vse zbrane in ustvarjene informacije so nato redno vzdrževane in so ločene od izvirnega digitalnega dokumenta. Med sekundarne metapodatke sodijo:

1. opisne informacije (navedba, opis),
2. sistemske informacije (ključi za razvrščanje, strojno berljivi datumi),
3. lokacijske informacije (mape datotek, pogon, nosilec),
4. informacije o dostopu (različne omejitve, pregled dostopov, pravni status)
5. revizijske informacije (izvor, zgodovina, prenos, sprememba)
6. referenčne informacije (semantične povezave, notranje povezave).

⁷ označuje sposobnost sistema ali izdelka, da brez posebnega napora potrošnika deluje z drugimi sistemi ali izdelki.

⁸ O samodejnem zbiranju sekundarnih podatkov je avtorica Gesa Büttner opisovala v članku z naslovom Auto-classification in an international organization: report from a feasibility study.

Dodatni metapodatki

Dodatni metapodatki se v primeru državnega arhiva nanašajo na informacije o digitalnem dokumentu, ki jih prispeva tretja oseba ali organizacija, ki ni del vladnega ustroja.

Izpeljani metapodatki

Izpeljani metapodatki so priloženi atributi k izvorni digitalni vsebini in so rezultat neke vrste programske analize ali poljubnega algoritemskega računanja. Shranjeni so v strukturirani obliki⁹ in se občasno osvežujejo. Dober primer so spletni iskalniki, ki imajo za lažje delovanje v ozadju izdelane najrazličnejše indekse.

Kontrolni metapodatki

Kontrolni metapodatki so informacije, ki so namenjene npr. kontroli skladnosti z izbranimi pravili ali standardi. Kontrola se lahko nanaša na povezan nabor navodil, ki določajo predstavitev predmeta v različnih okoliščinah, kot je slogovna lista. Metapodatki o metapodatkih

Metapodatki opisujejo spremembe metapodatkov, kajti metapodatki niso nujna fiksna entiteta. Metapodatki so podvrženi spremembam in vse te spremembe skozi čas bi bilo zelo priporočljivo zabeležiti. Še posebej, če upoštevamo umetno inteligenco, ki se v procesu samodejnega obdelovanja ali ustvarjanja podatkov nenehno dopolnjuje ali ureja. Ob celotni poplavi najrazličnejših podatkov bomo morali najti način, da bomo vsem tem podatkom tudi zaupali. Zagotovo je eden od načinov tudi ustvarjanje metapodatkov o metapodatkih.

Skoraj vsako področje našega raziskovanja ima definiran svoj nabor metapodatkov, ali so ga izpeljali bodisi iz ISO standarda ali Dublin Core¹⁰. Neizpodbitno dejstvo je, da je potrebno za lažje delo prilagajati nanizane rešitve. Razvoj podatkov o podatkih se bo čez čas razvijal in dopolnjeval in ob novih tehnologijah bomo pričali vedno večjemu avtomatiziranju pri kreiranju podatkov o podatkih.

⁹ Strukturirani podatki so podatki, ki so organizirani v obliki, ki jo brez težav uporablja baza podatkov ali drugačna tehnologija.

¹⁰ Dublin core je standard 15 osnovnih metapodatkovnih elementov za opisovanje na spletu dostopnih virov

2. ZAKLJUČEK

Količina digitalnih podatkov se enormno povečuje (Jenik, 2021) in to je nesporno dejstvo. Vsi ti podatki so razdrobljeni, na različne vrste dokumentov. V primeru dela v javni upravi govorimo o zadevah, kjer prevladujejo formati iz družine Microsoft Office™. Iz lastnih izkušenj pri delu z arhivskim gradivom lahko z veliko gotovostjo trdim, da je v fizičnih zadevah veliko balasta. Vendar se ob tem poraja vprašanje, ali je resnično vse balast? Arhivska praksa nas uči, da je potrebno vse shranjevat kronološko, za lažje delo naših zanamcev. Po drugi plati pa je iskanje po vseh nekončanih podatkih skoraj iluzorno. Težko si predstavljam, da bi za pregledno raziskavo, npr. stanje parnih kotlov v Republiki Sloveniji, podal med vse te papirje in oblikoval pregled dogajanja s parnimi kotli od konca druge svetovne vojne dalje. Vendar delo je delo in obstajajo tudi takšni junaki, ki se podajajo na takšne ekspedicije.

Danes, ko je podatkov mnogo več, so takšne ekspedicije časovno vedno bolj mukotrpne. Tukaj nam lahko vskoči na pomoč, umetna inteligenca. Žal trenutno še ni tako pametna, da bi bila povsem samostojna in bi jo kot v znanstveno fantastičnih filmih, samo spraševali za podatke, vse delo bi opravila sama, žal.

Orodja za lažje delo so prisotna že danes in se vedno bolj izpopolnjujejo. Poseg človeka ali pomoč človeka je zato nujno potrebna. Največkrat je potrebno pripraviti referenčni okvir znotraj katerega nato primerja svoje ugotovitve. Rezultati v projektih, ki sem jih predelal, prikazujejo velike uspehe ob takšnem delu torej s pomočjo referenčnih okvirjev.

Ob prebiranju vseh projektov in člankov je razvidno nekaj, podatki o podatkih so živa stvar in se spreminjajo skozi čas. Nekoč je veljal dokaj enostaven nabor metapodatkov, sedaj ko se arhivirajo tudi generirani podatki iz vrst umetne inteligence je potrebno le te drugače opremiti. Zaupanje v izračunane podatke umetne inteligence je vedno vprašljivo. Ljudje smo v večini skeptiki in nezaupljivi, evolucija je opravila svoje. Menim, da določena kritična masa nezaupanja mora biti prisotna, vendar bo potrebno sprejemati nove tehnologije.

Podatki o podatkih nam pripomorejo k temu zaupanja, več kot smo seznanjeni o določeni vsebini in o postopkih, ki jih do določene mere lahko kontroliramo, bolj jim zaupamo, saj na nek način obvladujemo situacijo. Sinergija, kot pojav, da dva ali več vplivov deluje skupaj in s tem se poveča učinkovitost se plastično prikazuje med metapodatki in umetno inteligenco. Več, kot je podatkov o podatkih, bolj je lahko umetna inteligenca učinkovita in vice versa, umetna inteligenca lahko producira, boljše podatke o podatkih. Vse to skupaj rezultira večjo uporabno vrednost vsebine in ne nazadnje lahko olajša delo arhivista.

3. VIRI IN LITERATURA

- The National Archives. (14.03.2018). *The National Archives - Digital archiving: The seven pillars of metadata*. The National Archives Blog. <https://blog.nationalarchives.gov.uk/digital-archiving-seven-pillars-metadata/>
- Bakker, R., Rowan, K., Hu, L., Guan, B., Liu, P., Li, Z., He, R., in Monge, C. (2020). AI for Archives: Using Facial Recognition to Enhance Metadata. *Works of the FIU Libraries*. <https://digitalcommons.fiu.edu/glworks/93>
- Bunn, J. (2020). Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI). *Records Management Journal*, 30(2), 143–153. <https://doi.org/10.1108/RMJ-08-2019-0038>
- Chabin, M. A. (2020). The potential for collaboration between AI and archival science in processing data from the French great national debate. *Records Management Journal*, 30(2), 241–252. <https://doi.org/10.1108/RMJ-08-2019-0042>
- Colavizza, G., Blanke, T., Jeurgens, C., in Noordegraaf, J. (2021). Archives and AI: An Overview of Current Debates and Future Perspectives. *arXiv:2105.01117 [cs]*. <http://arxiv.org/abs/2105.01117>
- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., in Kagal, L. (2018). Explaining Explanations: An Overview of Interpretability of Machine Learning. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)* (str. 80–89). <https://doi.org/10.1109/DSAA.2018.00018>
- Han, K. (2021). Research and Exploration of Metadata in Artificial Intelligence Digital Library. *Journal of Physics: Conference Series*, 1915(2). <https://doi.org/10.1088/1742-6596/1915/2/022061>
- Jenik, C. (30. 07. 2021). A Minute on the Internet in 2021. *Statista Infographics*. <https://www.statista.com/chart/25443/estimated-amount-of-data-created-on-the-internet-in-one-minute/>
- Makhlouf Shabou, B., Tièche, J., Knafo, J., in Gaudinat, A. (2020). Algorithmic methods to explore the automation of the appraisal of structured and unstructured digital data. *Records Management Journal*, 30(2), 175–200. <https://doi.org/10.1108/RMJ-09-2019-0049>
- Moss, M., Thomas, D., in Gollins, T. (2018). The Reconfiguration of the Archive as Data to Be Mined. *Archivaria*, 86, 118–151.
- Ochoa, X., & Duval, E. (2009). Automatic evaluation of metadata quality in digital libraries. *International Journal on Digital Libraries*, 10, 67–91. <https://doi.org/10.1007/s00799-009-0054-4>
- Ploj, B. (2017). Umetna inteligenca—Nevronske mreže (3.del). *Elektrotehniška revija*, 1, 45–48.
- Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsouponova, T., in Stuart, K. (2019). More human than human? Artificial intelligence in the archive. *Archives and Manuscripts*, 47(2), 179–203. <https://doi.org/10.1080/01576895.2018.1502088>
- Seles, A. (25.11.2019). *A Brave New World: Artificial Intelligence and Archives*. http://www.archives.go.jp/english/news/pdf/20191125_27e_01.pdf
- Theimer, K. (2018). It's the end of the archival profession as we know it, and I feel fine. V C. Brown (ur.), *Archival Futures* (str. 1–18). Facet. <https://doi.org/10.29085/9781783302192.002>

Thibodeau, K. (2016). Breaking down the invisible wall to enrich archival science and practice. *2016 IEEE International Conference on Big Data (Big Data)* (str. 3277–3282). <https://doi.org/10.1109/BigData.2016.7840986>

iPROM. (2022). *UI; Umetna inteligenca*. <https://iprom.si/slovar/umetna-inteligenca/>

Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Qi Dong, J., Fabian, N., in Haenlein, M. (2021). Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research*, 122, 889–901. <https://doi.org/10.1016/j.jbusres.2019.09.022>

Summary

The amount of digital data is increasing enormously (Jenik, 2021) and this is an indisputable fact. All this information is fragmented into different types of documents. In the case of work in public administration, we are talking about matters where the formats from the Microsoft Office family predominate™. From my own experience in working with archival material, I can say with great certainty that there is a lot of ballast in physical matters. However, this raises the question of whether everything is really ballast? Archival practice teaches us that it is necessary to store everything chronologically, to facilitate the work of our descendants. On the other hand, the search for all the unfinished data is almost illusory. I have a hard time imagining that for transparent research, e.g. the state of steam boilers in the Republic of Slovenia, passed among all these papers and formed an overview of events with steam boilers from the end of the Second World War onwards. However, work is work and there are also such heroes who embark on such expeditions.

Today, when there is much more data, such expeditions are increasingly time-consuming. Artificial intelligence can come to our aid here. Unfortunately, at the moment it is not so smart to be completely independent and, as in science fiction movies, it would just be asked for information, it would do all the work herself, unfortunately.

Tools for easier work are already present today and are becoming more sophisticated. Human intervention or human assistance is therefore urgently needed. In most cases, it is necessary to prepare a reference framework within which he then compares his findings. The results in the projects I have reworked show great successes in such work, therefore with the help of reference frameworks.

Reading all the projects and articles show that data about data is a living thing and changes over time. Once upon a time, a fairly simple set of metadata was considered, but now that the generated data from the types of artificial intelligence are also archived, it is necessary to equip them differently. Confidence in the calculated AI data is always questionable. Humans are mostly skeptical and distrustful. I believe that a certain critical mass of mistrust must be present, but new technologies will have to be adopted.

Data about data helps us to build trust, the more we are aware of a certain content and of procedures that we can control to a certain extent, the more we trust them, because in some way we manage the situation. Synergy, as a phenomenon that two or more influences work together and thus increase efficiency, is plastically displayed between metadata and artificial intelligence. The more data there is, the more artificial intelligence can be effective and vice versa, artificial intelligence can produce, better data. All this together results in greater useful value of the content and, last but not least, can facilitate the work of the archivist.