Etičnost u korištenju velikih podataka

Šušnjara, Kristina

Master's thesis / Diplomski rad

2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Department of Croatian Studies / Sveučilište u Zagrebu, Hrvatski studiji**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:111:263649

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**

SVEUČILIŠTE U ZAGREBU
HRVATSKI STUDIJI
UNIVERSITY OF ZAGREB
CENTER FOR CROATIAN STUDIES

Repository / Repozitorij:

Repository of University of Zagreb, Centre for Croatian Studies

zir.nsk.hr

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

SVEUČILIŠTE U ZAGREBU
HRVATSKI STUDIJI

Kristina Šušnjara

ETIČNOST U KORIŠTENJU VELIKIH PODATAKA

DIPLOMSKI RAD

Zagreb, 2019.

SVEUČILIŠTE U ZAGREBU
HRVATSKI STUDIJI
ODSJEK ZA KOMUNIKOLOGIJU

Kristina Šušnjara

ETIČNOST U KORIŠTENJU VELIKIH PODATAKA

DIPLOMSKI RAD

Mentor: prof. dr. sc. Danijel Labaš

Zagreb, 2019

Sadržaj

1. Uvod 1
2. Definiranje pojma veliki podaci (Big Data) 3
2.1. Što su veliki podaci? 5
2.2. Što sve nisu veliki podaci? 6
3. Etika i etičnost korištenja Velikih Podataka 7
3.1. Etička pitanja u industriji velikih podataka 10
3.2. Poslovna etika 11
3.3. Zakon o zaštiti osobnih podataka 14
4. Veliki podaci za veće dobro (Data for good) ili podaci kao kršitelji privatnosti? 18
4.1. Edward Snowden i novinarstvo 21
5. Medijsko pravo na pristup informacijama 23
5.1. Data Novinarstvo (podatkovno novinarstvo) 24
5.2. Etičnost u podatkovnom novinarstvu 26
6. Metodologija istraživanja 28
6.1. Ciljevi i hipoteze istraživanja 28
6.2. Analiza i interpretacija rezultata 31
6.2. Diskusija 50
7. Zaključak 53
Popis literature 55
Prilozi 58

SAŽETAK

Rad *Etičnost u korištenju Velikih podataka* sadrži teorijski i istraživački dio. U teorijskom dijelu definiran je pojam „veliki podaci“, što su veliki podaci te što nisu, etika i etičnost u korištenju podataka, etička pitanja u industriji velikih podataka, etika u poslovanju, zakon koji se odnosi na zaštitu podataka i zakon o pristupu na informacije, prikaz podataka za veće dobro ili podaci kao kršitelji privatnosti, a na kraju se govori o slučaju Edwarda Snowdena te podatkovnom novinarstvu i etičnosti primjene takve vrste novinarstva.

Istraživački dio sadrži rezultate analize sadržaja hrvatskih portala 24sata i Tportal.hr te inozemnih portala BBC i The Guardian.

Postoji mnogo primjera velikih podataka koji se koriste u praksi. Tvrtke proširuju svoje tradicionalne skupove podataka s podacima društvenih medija, preglednika i sličnim podacima kako bi bolje razumjele svoje klijente, njihova ponašanja i preferencije. Trgovci mogu optimizirati dionice na temelju predviđanja generiranih iz velikih podataka, a sve više gradova koristi podatke s cestovnih senzora i kamera kako bi optimizirali protok prometa. Ovo su samo neki od primjera u kojima pronalazimo velike podatke te na koji način se mogu koristiti. Budući da su se veliki podaci već ukorijenili u naše živote i same sustave poslovanja, pitanje je koliko je njihovo korištenje etično i je li etično zabraniti uporabu velikih podataka ako oni mogu pomoći javnom dobru?

Uz pomoć analize sadržaja istražit ćemo što o velikim podacima pišu hrvatski i inozemni portali, a potom usporediti njihovo izvještavanje. Jednako tako, istražit ćemo upozoravaju li novinski članci korisnike na njihovu štetnost, odnosno nude li javnosti savjete zaštite i prikazuju li članci i njihovu pozitivnu stranu korištenja.

Ključne riječi: veliki podaci, podaci, etičnost, etika, privatnost, zaštita podataka, krađa podataka, osobni podaci, gdpr, internetski portali

ABSTRACT

Research paper *Ethics in usage of Big Data* consists of a theoretical and a research part. The theoretical part defines the notion of "big data", what is big data and what is not, ethics and ethics being used in using the data, ethical questions in industrial big data, ethics in business, data protection law and information availability law. a presentation of data for the greater good or information about breaking the privacy, and lastly, the case of Edward Snowden and information on journalism and ethics that are appropriate for that kind of journalism.

The research part contains the results of the contents of the Croatian portal 24sata and Tportal.hr portal as well as the foreign news portals of BBC and The Guardian.

There are many examples of big data used in practice. Businesses are expanding their traditional datasets with social media, browser and similar data to better understand their customers, their behaviors and preferences. Traders can optimize stocks based on the predictions generated from big data, and more cities are using data from road sensors and cameras to optimize traffic flow. These are just some of the examples where we find big data and how it can be used. Since big data has already taken root in our lives and the systems of business itself, the question is, how ethical is their use and is it ethical to ban the use of big data if they can help the public good?

With the help of content analysis, we will explore what big data is written by Croatian and foreign portals and then compare their reporting. Equally, we will investigate whether newspaper articles alert users to their harmfulness, that is, whether they offer protection tips to the public and whether they show the articles and their positive side of use.

Keywords: big data, data, ethics, ethics privacy, data protection, data theft, personal data, gdpr, internet portals

1. Uvod

Veliki podaci objedinjuju se iz različitih izvora, uključujući društvene medije, povijest pretraživanja interneta, zapise o online transakcijama, mobilnim tehnologijama i senzorima koji prikupljaju informacije o lokaciji i bilo kojem drugom izvoru na kojemu su svjesno ili nesvjesno ostavljeni digitalni tragovi. Neke od tih podataka aktivno i dobrovoljno daju korisnici i potrošači na mrežnim stranicama, dok se drugi prikupljaju različitim sredstvima i tehnologijama ugrađenim u rutinske aktivnosti svakodnevnog života. S obzirom na širenje društvenih mreža, mobilnih tehnologija i sve veću digitalizaciju, količina podataka koja se danas generira dosegla je jako visoku razinu. Upravo bi se iz tih razloga trebala razmotriti i etička pitanja, u smislu krši li nam određena uporaba velikih podataka temeljna građanska, društvena, politička i zakonska prava.

Još unazad pet do deset godina bili smo svjedoci nerazmjernih uporaba velikih podataka koji su se iskorištavali na neetičan način. Kord Davis (2012: 10) navodi kako je 16. veljače 2012. New York Times objavio članak o sposobnosti Targeta [1] da identificira kada je klijentica trudna. Target je odbio komentirati i sudjelovati u toj „priči“, ali članak je svejedno napisan i objavljen. Niz komentara i naknadnih vijesti izazvali su brojna pitanja, od zakonitosti djelovanja Targeta do šire javne zabrinutosti o osobnim informacijama koje su postale javne. Dalje navodi drugi slučaj, iz kolovoza 2011. kada se Facebook suočava s kritikom da izlaže imena i telefonske brojeve svih u kontaktima na mobilnim uređajima koji koriste opciju „kontakti“ na Facebook mobilnoj aplikaciji (2012: 11).

Nedjeljko Frančula (2018: 69) navodi da ako se u Google Znalac[2] upiše 'Big Data' (veliki podaci) i traže svi radovi koji u naslovu sadrže tu sintagmu, pojavljuje se 59 900 radova. Od 2005. do 2011. objavljena su ukupno 593 rada, a od 2015. do 2017. objavljeno je svake godine više od 8000 radova. To pokazuje u kolikoj su mjeri posljednjih godina veliki podaci (Big Data) postali predmet znanstvenih i stručnih istraživanja.

Velike podatke (Big Data) koriste velike organizacije i marketinške agencije te se tako približavaju potrošačima i pomoću njihovih osobnih podataka mogu saznati njihove potrebe.

[1] Target Corporation je četvrti najveći trgovac u Sjedinjenim Američkim Državama, koji posluje u 1.556 trgovina u 47 država (https://www.encyclopedia.com/social-sciences-and-law/economics-business-and-labor/businesses-and-occupations/target-corp, stranica posjećena: 10. svibnja 2019.).

[2] Google Scholar ili Google Znalac je slobodno dostupan web pretraživač znanstvenih radova, u vlasništvu tvrtke Google.

Postavlja se pitanje je li takav način primjene etičan, pa dolazimo do činjenice da se regulacija razvija sporije od tehnologije.[3] Kada se raspravlja o toj temi onda su vrlo važna pitanja koju kontrolu imamo nad tim podacima, gubi li se naša autonomija i tko donosi odluke o nama. Javnost teško dolazi do odgovora na ta pitanja te su upravo zato u prenošenju informacija i tumačenju ljudskih prava na tom području jako bitni mediji. Internet postaje sve važniji izvor dnevnog informiranja, a vodeću ulogu u tome imaju dnevno-informativni portali.

Osim što podatke povezujemo s velikim korporacijama koje koriste naše podatke u komercijalne svrhe, važno je rasvijetliti i njihovu potencijalnu stranu. Veliki podaci naime pokazuju velike i bitne promjene na području medicine, genetike i astronomije.

S ciljem dobivanja što jasnije slike o ranije spomenutim otvorenim pitanjima i dvojbama, u ovome će radu teorijsku osnovu pratiti i istraživački dio. Glavni cilj ovog istraživanja je analizom sadržaja odabranih hrvatskih i inozemnih internetskih portala (24 sata, Tportal.hr, BBC i The Guardian) utvrditi koliko se često na internetskim portalima spominje etičnost u korištenju velikih podataka od 1. siječnja 2016. do 10. svibnja 2019. godine. Unutar glavnog cilja, specifično ćemo se usmjeriti na usporedbu o načinu izvještavanja na internetskim portalima u inozemstvu i u Hrvatskoj, a cilj je i utvrditi stupanj pozitivnog i negativnog načina izvještavanja o korištenju velikih podataka (informiranje, analiza, kritika). Hipoteze koje ćemo potvrditi ili opovrgnuti su sljedeće:

H1: Tekstovi na internetskim portalima u Hrvatskoj u manjoj mjeri pružaju informacije o zaštiti i savjetima o upotrebi velikih podataka.

H2: Tekstovi na internetskim portalima u Hrvatskoj u manjoj mjeri izvještavaju o upotrebi osobnih podataka u komercijalne svrhe.

H3: Tekstovi na internetskim portalima u većoj mjeri predstavljaju podatke kao prijetnju za osobnu sigurnost.

H4: Mediji u Hrvatskoj u manjoj mjeri od inozemnih izvještavaju o pozitivnim stvarima upotrebe velikih podataka.

[3] U svibnju 2018. je na snagu stupila odredba o zaštiti osobnih podataka - GDPR, no ona vrijedi samo za europsko područje, dok su Sjedinjene Američke Države slabije regulirane.

2. Definiranje pojma veliki podaci (Big Data)

Veliki podaci su pojam koji opisuje veliki broj strukturiranih i nestrukturiranih podataka koji se prikupljaju iz više izvora. Većina radnika informatičke tehnologije (IT) se koristi strukturiranim podacima. Strukturirani podaci mogu se lako pretraživati pomoću osnovnih algoritama, npr. proračunske tablice i podaci iz senzora stroja, dok nestrukturirani podaci više nalikuju ljudskom jeziku.

Kada pak govorimo o pravoj definiciji velikih podataka, nužno je napomenuti kako se radi o pojmu koji je iz godinu u godinu poprimao nova i nadograđena obilježja onoga što su veliki podaci. Tako Kristina Kocijan u svom članku *Big Data: kako smo došli do Velikih podataka i kamo nas oni vode* (2014: 2 prema: Francis Xaver Diebold, 2000: 1) navodi da prvu definiciju velikih podataka kao fenomena nalazimo kod Francisa Xaver Diebolda koji piše:

„U zadnje je vrijeme dosta dobre znanosti, bez obzira je li u pitanju fizika, biologija ili sociologija, bilo prisiljeno suočiti se – od čega je često i profitirala – s fenomenom Velikih podataka. Veliki podatci odnose se na eksploziju u količini (a katkad i kvaliteti) dostupnih i potencijalno relevantnih podataka, uglavnom kao posljedica skorih i besprimjernih napredaka u tehnologiji zapisivanja i pohranjivanja podataka.“ (Diebold 2000: 1)

Kirsten. E. Martin objašnjava da veliki podaci imaju sposobnost predviđanja budućih ishoda i da je to proces koji koristi masovne skupove podataka i algoritamsku analizu za izdvajanje novih informacija i značenja. Također ih opisuje kao sredstvo i trenutak u kojem obujam podataka i brzina ograničavaju uporabu tradicionalnih alata (2015: 69).

Velike podatke obilježava se V-ovima te se smatra da je volumen njihov najveći dio. Kocijan (2014 prema: Reeve, 2013; Simon, 2013; Berman, 2013) piše o 3V u kojeg se ubrajaju: volumen (velika količina podataka), varijantnost (raznolikost tipova podataka: tradicionalne baze podataka, fotografije, dokumenti) i velocitet, to jest brzina kojom se akumuliraju novi podaci (iz sličnih izvora podataka, iz prethodno arhiviranih podataka, iz podataka koji stalno pridolaze iz različitih izvora – engl. *streamed* data), ali i brzina kojom se očekuje da podaci budu dostupni za analizu. Postojanje takvih vrijednosti razlikuje „velike podatke“ od „puno podataka“. Ripon Patgiri i Arif Ahmed (2016: 17) slažu se da je volumen najveći dio velikih podataka, ali i ukazuju na postojanje kontroverzi i zbunjenosti oko V-ova velikih podataka. Navode da je 3V najprije definirao Doug Laney u 2001. kako bi opisao upravljanje podacima u 3 dimenzije što označava volumen, brzinu i raznolikost. U zadnjih nekoliko godina Velikim podacima dodani su mnogi V-ovi, a radi se o 7V i 1C. U 7V se

ubrajaju vjerodostojnost, vrijednost, valjanost, varijabilnost, nestalnost, virtualnost, vizualizacija/vidljivost i 1C kao složenost[4].

Kocijan (2014: 2) smatra da su veliki podaci svoj uspon doživjeli tek 2008. godine, no termin 'Big Data' se pojavljuje u akademskim i neakademskim radovima od sredine '90-ih godina 20. stoljeća. Ista autorica navodi da je teško utvrditi tko ga je prvi osmislio i počeo upotrebljavati, ali zasad zasluge idu Johnu Masheyu, kao najvjerojatnijemu autoru termina, koji je krajem '90-ih *„samo tražio jednostavan naziv za cijeli niz tema koji bi najbolje prenio sliku da se granice računala mijenjaju i napreduju.“* On je u to vrijeme radio za Silicon Graphics, veliku kompaniju koja se bavila računalnom grafikom, a njezinim uslugama se koristio Hollywood za izradu specijalnih efekata, ali i agencije za video nadzore. Bila je to upotreba novih tipova podataka u velikim razmjerima (Kocijan, 2014: 2 prema: Lohr, 2013).

Veliki podaci imaju veliki potencijal utjecanja na društvo i znanost, uključujući način na koji pristupamo i provodimo istraživanja. Tijekom proteklih nekoliko godina došlo je do dramatične evolucije naših tehnoloških sposobnosti prikupljanja i razmjene informacija što je omogućilo prikupljanje, distribuciju i analizu velikih količina podataka na načine koji nikada prije nisu bili mogući. S obzirom na brzinu kojom se stvari odvijaju, može se reći da se radi o svojevrsnoj podatkovnoj revoluciji (eng. data revolution) koja ima potencijal da promijeni način na koji živimo, radimo i razmišljamo. Stoga ne čudi stvaranje novih zanimanja kao što su podatkovni znanstvenik (eng. data scientist), menadžer velikih podataka, analitičar velikih podataka ili podatkovni arhitekt. Prema izvještaju tvrtke McKinsey & Company[5] iz 2011. godine (Manyika i suradnici), njihova predviđanja su bila da će do 2018. godine samo na američkom tržištu nedostajati preko milijun i pol podatkovnih znanstvenika.

International Dana Corporation (IDC)[6] navodi da se veličina podataka udvostručuje svake dvije godine, a do 2020. digitalni svemir – podaci koje stvaramo i kopiramo godišnje, doseći će 44 zetabajta ili 44 trilijuna gigabajta.

---

[4] veracity (vjedostojnost), value (vrijedsnot), validity (vrijednost), variability (varijabilnost), volatity (nestalnost), virtual (virtualnost), visualization/visibility (vizualizacija/vidljivost), complexity (složenost)

[5] McKinsey & Company je američka svjetska konzultantska tvrtka za upravljanje. Provodi kvalitativnu i kvantitativnu analizu za procjenu upravljačkih odluka u javnom i privatnom sektoru. Smatra se jednom od najprestižnijih konzultantskih tvrtka za upravljanje. McKinseyjeva klijentela uključuje 80% najvećih svjetskih korporacija i opsežan popis vlada i neprofitnih organizacija.

[6] Vodeći globalni pružatelj marketinških informacija, savjetodavnih usluga i događanja za tržišta informacijske tehnologije, telekomunikacija i potrošačke tehnologije. S više od 1.100 analitičara diljem svijeta, IDC nudi globalno, regionalno i lokalno znanje o tehnologiji, mogućnostima i trendovima u industriji u više od 110 zemalja (stranica posjećena: 10. 5. 2019., https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm).

2.1. Što su veliki podaci?

Digitalni podaci su posvuda, u svakom sektoru, u svakoj ekonomiji i organizaciji, a mi kao korisnici digitalne tehnologije stvaramo i ostavljamo velike količine podataka iza sebe.

Google, Amazon, Netflix, Starbucks, Apple samo su neke od kompanija koje se koriste velikim podacima i to u svrhe privlačenja korisnika, oglašavanja, ali i za osobne istraživačke potrebe. Također, veliki podaci se koriste za poboljšanje mnogih aspekata gradova i zemalja. Na primjer, veliki podaci omogućuju gradovima optimiziranje prometnih tokova na temelju informacija o prometu u stvarnom vremenu.

Podaci su se u prošlosti mogli prikupljati samo iz proračunskih tablica i baza podataka, a danas se podaci nalaze u raznim oblicima kao što su e-pošta, PDF dokumenti, fotografije, videozapisi, audio zapisi i još mnogo toga.

Kord Davis (2012: 10) u svojoj *knjizi Ethics of Big Data: Balancing Risk and Innovation* objašnjava da su kompanije kao što su Target, Apple, Linkedin, Facebook i Twitter, ali i političke stranke i veliki broj organizacija koje stvaraju veliku ekonomsku, društvenu i političku vrijednost, svjesne da je došlo doba velikih podataka i da je tu da ostane. Dalje navodi primjer iz travnja 2011., kada su dvojica istraživača sigurnosti priopćila da iPhone uređaji redovito bilježe položaj svakog uređaja u skrivenu datoteku. Apple je potvrdio da je tvrdnja istinita što je jasan pokazatelj da se radi o metodi putem koje je datoteka generirana i pohranjena, a što je uzrokovalo sigurnosne probleme. Odluka o korištenju te tehnološke metode imala je jasne i izravne etičke posljedice u stvarnom svijetu.

Autor se pita kako možemo odrediti koji su podaci pouzdani? Bez obzira radi li se o nama osobno ili o drugima, veliki podaci eksponencijalno povećavaju količinu informacija i načine na koje možemo komunicirati s njima. I pojedinci i organizacije mogu imati legitiman interes za razumijevanje kako se postupa s tim podacima (Kord Davis, 2012).

Annika Richterich (2018: 5) objašnjava da se podaci smatraju vrijednim komercijalnim resursima te da postoje brojni načini i razlozi za prikupljanje velikih podataka u komercijalne svrhe. Društvena mreža kao što je Facebook dokumentira digitalne interakcije korisnika (prema: Geerlitz and Helmond, 2013), a mnogi programi za razmjenu izravnih poruka i davatelji usluga e-pošte skeniraju poruke korisnika u reklamne svrhe ili koriste ključne riječi koje su vezane uz sigurnost (prema: Gibbs, 2014; Wilhelm, 2014; Godin, 2013). Ne bilježe se samo digitalne interakcije i komunikacije korisnika, već i njihovi fizički pokreti

koji se pretvaraju u digitalne podatke. Za primjer uzmimo Apple Watch[7] koji prati, arhivira i analizira korake i broj otkucaja srca vlasnika uređaja.

Kristina Kocijan (2014: 6) navodi da podatke prikupljaju i čitači e-bookova koji tako dobivaju uvid u navike ljudi koji ih čitaju (koliko dugo čitaju stranicu, gdje čitaju, kako brzo okreću stranicu, jesu li odustali od čitanja, jesu li napravili bilješku na margini ili nešto podcrtali ...). Ovi su podaci od velike važnosti i izdavačima i autorima knjiga. Tako su npr. Barnes&Noble prikupljali podatke preko svojih Nook čitača e-knjiga i otkrili kako ljudi odustaju na pola čitanja dugih knjiga koje nisu fikcija. Zato su ponudili serijal *Nook snaps* s kratkim osvrtima na teme poput zdravlja i trenutnih događaja (prema: Mayer-Schönberger, Cukier, 2013).

Jedna od zajedničkih značajki svih ovih primjera je pojava kontinuiranog povećavanja i širenja baza podataka. Takve baze podataka omogućuju uvid u fizičko stanje korisnika (sadašnje ili buduće); učestalost i (jezične) kvalitete njihovih društvenih kontakata; njihove postavke i obrasce pretraživanja; i njihovu geografsku pokretljivost. Općenito govoreći, korporativni podaci su namijenjeni prodaji ili korištenju tih podataka kako bi se omogućila prilagodba iskustva korisnika, a prije svega kako bi generirali profit (Richterich, 2018: 6).

2.2. Što sve nisu veliki podaci?

Prema Kristini Kocijan (2014, prema: Berman, 2013) postoji 10 razlika između standardnih podataka i velikih podataka:

1. ciljevi – standardni podaci daju odgovor na specifično pitanje s unaprijed određenim ciljem; veliki podaci imaju odgovore na raznovrsna pitanja s prilagodljivim ciljem;
2. lokacija – standardni podaci se uglavnom nalaze unutar jedne organizacije; veliki podaci se mogu nalaziti raspršeni na različitim lokacijama;
3. struktura i sadržaj podataka – standardni podaci su strukturirani podaci s domenom iz jednoga područja, ujednačene forme; veliki podaci su nestrukturirani podaci (tekstni dokumenti, slike, filmovi, zvučni zapisi itd.) koji mogu dolaziti iz različitih domena s dodatnim vezama na podatke iz drugih izvora;

[7] Apple Watch je linija pametnih satova koje je dizajnirala i razvila tvrtka Apple Inc. Povezan je s iOS-om i drugim Appleovim proizvodima i uslugama. Može pratiti zdravstvenu aktivnost, komunicirati s prijateljima i pokrenuti široku paletu aplikacija.

4. priprema podataka – standardne podatke priprema uglavnom korisnik tih podataka; velike podatke priprema mnogo ljudi jer su i podaci iz različitih izvora, dok su korisnici podataka rijetko ljudi koji su podatke pripremili;

5. životni vijek – standardni podaci imaju ograničen vijek postojanja (u prosjeku 7 godina po završetku projekta); veliki podaci sadrže podatke bez ograničenog životnog vijeka jer se većina integrira u nove projekte koji koriste velike podatke;

6. mjerenja – standardni podaci se uglavnom mjere s pomoću jednoga protokola dok se veliki podaci mjere različitim protokolima (upravo je utvrđivanje kvalitete podataka kod VP najzahtjevniji posao);

7. reproduciranje – projekti koji koriste standardne podatke daju se lako reproducirati; projekti koji upotrebljavaju velike podatke mogu se rijetko kad reproducirati;

8. financijsko ulaganje – financije uložene u projekte sa standardnim podacima su relativno male, za razliku od financija uloženih u projekte s velikim podacima koje mogu dovesti i do bankrota;

9. introspekcija – pojedinačni standardni podaci mogu se identificirati s pomoću njihove lokacije određene retkom i stupcem unutar tablice, međutim, kod velikih podataka procedura za identifikaciju puno je složenija i ona se može ostvariti s pomoću tehnike introspekcije;

10. analiza – kod standardnih podataka analiza se može vršiti nad svim podacima istovremeno, kod velikih podataka analiza se odvija u koracima (osim u paralelnoj analizi koja se istodobno odvija na više računala) na način da se podaci izvlače, pregledavaju, smanjuju, normaliziraju, transformiraju, vizualiziraju, interpretiraju i ponovo analiziraju različitim metodama.

3. Etika i etičnost korištenja Velikih Podataka

Unutar industrije velikih podataka, podaci kao što su oni o potrošačima na mreži ili podaci o lokaciji prenose se s jedne tvrtke na drugu u lancu opskrbe informacijama. Unutar lanca opskrbe potrošači pružaju informacije tvrtkama, koje ih zatim prosljeđuju tvrtkama za praćenje, a također ih mogu proslijediti agregatorima podataka. Pitanje privatnosti i osjećaj sigurnosti i moć nad podacima koje kao korisnici ostavljamo, sigurno su neka od važnijih pitanja. Kod privatnosti je važno naglasiti da bi bilo dobro imati više znanja o kontroli nad

svojim podacima te znanja o osobnim podacima koji se prikupljaju o nama prilikom korištenja internetskih usluga.

Jeff Collman i Sorin Adam Matei (2016: 44) objašnjavaju da privatni i javni subjekti prikupljaju, analiziraju i ponekad šire velike podatke. Primjerice, prikupljanje podataka o istraživanju tržišta putem društvenih medija omogućilo je oglasnim tvrtkama kao što su Facebook i Google stvaranje ogromnih digitalnih carstava pomoću poslovnih modela koji bi bili nemogući prije deset godina. Autori navode da osim toga, vladine agencije prikupljaju ogromne količine podataka putem telekomunikacija u svrhu provedbe zakona, nadzora, epidemiologije i niza drugih pitanja i problema. Tehnološki alati koji su omogućili organizacijama da prikupljaju i analiziraju te podatke jednostavno su previše sofisticirani i prebrzo se mijenjaju kako bi javnost dala pristanak za prikupljanje i korištenje takvih informacija.

U sljedećim će poglavljima rada stoga biti predstavljeni i primjeri pozitivnih primjena velikih podataka na području medicine i genetike te što Zakon o zaštiti podataka (GDPR) govori o iskorištavanju podataka u takve svrhe na području Europe. Autorica Annika Richterich (2018: 34 prema: Hedgecoe, 2016: 578) navodi da su biomedicinska istraživanja regulirana tijekom posljednjih 50 godina i da to područje ima dugu tradiciju u uspostavljanju etičkih načela. Dalje objašnjava da se u kontekstu etike takva istraživanja ne provode samo unutar sveučilišta, već internetske i tehnološke korporacije same provode istraživanja, zaobilazeći etički nadzor koji se primjenjuje na sveučilišta.

Uravnoteženje privatnosti i sigurnosti usko je povezano s istraživanjima javnog zdravlja i biomedicinskim istraživanjima: zaštita pojedinca i građanskih prava stoji nasuprot javnom zdravlju i dobrobiti kao zajedničkom/javnom dobru.

Kirsten E. Martin (2015) donosi primjere korisnih i upitnih upotreba velikih podataka. Jedan od primjera se odnosi na prepoznavanje lica, koje može omogućiti prepoznavanje potencijalnih terorista na nekim društvenim događanjima, dok se u svrhu toga upitna namjena odnosi na društvene mreže koje koriste fotografije za identifikaciju članova. Drugi primjer se odnosi na strojno učenje prepoznavanja provalnika i mogućnost pristupa telefonskim zapisima za identifikaciju, a upitna primjena se odnosi na mogućnost da se pametnom telefonu pristupi bez naloga.

U demokratskim društvima privatnost se smatra građanskim pravom. Pravo na privatnost (implicitno ili eksplicitno) usidreno je u mnogim nacionalnim ustavima (Richterich, 2018: 35 prema: González Fuster, 2014; Glenn, 2003). Zaštita osobnih podataka obično se smatra

proširenjem prava na privatnost. Autorica navodi da ih povelja o temeljnim pravima Europske unije tretira odvojeno što se vidi iz članka 8 s naglaskom na zaštitu podataka i poštivanje privatnog i obiteljskog života obuhvaćenog člankom 7, a oni glase (Povelja Europske unije o temeljnim pravima, 2016: 202/395):

Članak 7. ***Poštovanje privatnog i obiteljskog života***

Svatko ima pravo na poštovanje svojeg privatnog i obiteljskog života, doma i komuniciranja.

Članak 8. ***Zaštita osobnih podataka***

1. Svatko ima pravo na zaštitu osobnih podataka koji se na njega ili nju odnose.

2. Takvi podaci moraju se obrađivati pošteno, u utvrđene svrhe i na temelju suglasnosti osobe o kojoj je riječ, ili na nekoj drugoj legitimnoj osnovi utvrđenoj zakonom. Svatko ima pravo na pristup prikupljenim podacima koji se na njega ili nju odnose i pravo na njihovo ispravljanje.
3. Poštovanje tih pravila podliježe nadzoru neovisnog tijela.

Abraham L. Newman (2015: 507) govori da privatnost nije mrtva i da je Europska unija po pitanju privatnosti u suprotnosti s politikom privatnosti u SAD-u. U odluci iz 2014. Sud pravde Europske unije odlučio je da „pojedinci imaju pravo - pod određenim uvjetima - tražiti od tražilica da uklone veze s osobnim podacima o njima“ (Annika Richterich, 2018: 36). Ovo je samo jedna od razlika između Europe i Amerike kada je riječ o privatnosti te zato Newman ističe kako privatnost nije mrtva.

Kako se digitalne komunikacije i dalje povećavaju, ljudi dijele još više podataka, uključujući i osobne podatke. Globalni digitalni izvještaj 2019 (Global Digital Report) pokazuje da ima da ima 4,39 milijardi internetskih korisnika, što je povećanje za 366 milijuna (9 posto) u odnosu na siječanj 2018. godine, 5,11 milijardi jedinstvenih mobilnih korisnika, 3,48 milijardi korisnika društvenih medija s ukupnim rastom u svijetu za 288 milijuna (9 posto) od ovog doba prošle godine te 3,26 milijardi ljudi koristi društvene medije na mobilnim uređajima s rastom od 297 milijuna novih korisnika, što predstavlja međugodišnji rast veći od 10 posto. Zbog velikog broja podataka koje stvaraju korisnici, ti podaci postaju javno dostupni što otvara mogućnost za „rudarenje podataka“[8] i razvijanje algoritama. Osobne demografske informacije, kao i ideje i misli (*tweets* ili poruke) koje bi se nekada dijelile u

[8] Rudarenje podataka ili podatkovno rudarenje (eng. *Data mining*) je sortiranje, organiziranje ili grupiranje velikog broja podataka i izvlačenje relevantnih informacija.

privatnom okruženju s grupama poznanika sada su dostupne bilo kome tko ima računalo i internet.

3.1. Etička pitanja u industriji velikih podataka

Budući da u poslovanju raste sve veće zadovoljstvo zbog onog što veliki podaci mogu učiniti za biznis, takve novonastale okolnosti otvaraju nova pitanja vezana uz poslovanje i etiku. Neka od tih pitanja odnose se na stjecanje, pohranjivanje i korištenje velikih količina podataka o osobinama ljudi, njihovom ponašanju, sklonostima, odnosima i lokacijama.

Iako je Kord Davis knjigu *Ethics of Big Data: Balancing Risk and Innovation* napisao 2012., od tog se doba neke stvari nisu mnogo promijenile. On navodi (2012: 17) da je uz pomoć velikih podataka moguće rekonstruirati cijelu povijest našeg putovanja bilo gdje na planeti, jer dolazi do povezivanja disperzivnih aspekata naše osobnosti na načine koje ponekad ne možemo u potpunosti kontrolirati. Naše slike na odmoru vjerojatno nisu zamišljene kao relevantan materijal prilikom prijave za posao, ali veliki podaci značajno mijenjaju način upravljanja reputacijom u takvim situacijama. Ovo je samo jedan od primjera kako veliki podaci omogućuju širi i dublji uvid u ljudsko ponašanje i aktivnosti nego ikada prije. Poslovni svijet je jako brzo uočio njihov potencijal, a uvide u naše ponašanje pretvaraju u proizvode i usluge.

Davis (2012) također smatra da sve dok tehnologija velikih podataka nudi mogućnost povezivanja informacija i inovacije novih proizvoda i usluga, kako za profit, tako i za veće društveno dobro, ono je, kao i sva tehnologija, etički neutralna. To znači da ne dolazi s ugrađenom perspektivom o tome što je ispravno ili pogrešno ili što je dobro ili loše u njihovom korištenju. Tehnologija velikih podataka nema okvir vrijednosti. Autor nastavlja da pojedinci i korporacije imaju vrijednosne sustave, a jedino traženjem odgovora na etička pitanja možemo osigurati da se veliki podaci koriste na način koji je usklađen s tim vrijednostima. Smatrajući da su veliki podaci neutralni, Davis (2012) objašnjava da pravi problem leži u tome što sama upotreba velikih podataka nije etična, napominjući da iako veliki podaci predstavljaju ogromnu priliku (u obliku novih proizvoda i usluga) za široku poslovnu i društvenu korist, s druge strane ona također predstavlja ozbiljan rizik.

Kocijan (2014: 15) također smatra da veliki podaci sami po sebi nisu štetni, loši ni opasni. Napominje da ono što ih čini takvima jest sposobnost čovjeka da zbog očuvanja moći nekolicine, koristi sve dostupne resurse (u ovom slučaju velike podatke) na štetu većine.

Budući da veliki podaci izazivaju zabrinutost zbog osobne privatnosti, Davis (2012) ukazuje na to da se stvaraju nova pitanja o osobnom identitetu, a osobito o tome tko posjeduje naše osobne podatke i kako povećana prisutnost i dostupnost velike količine podataka utječe na naš život. On nas upućuje na četiri zajednička elementa koja pripadaju organizacijama, ali i nama kao pojedincima, a ti elementi definiraju što se može smatrati okvirom za etiku velikih podataka. Ta četiri elementa su: 1) *Identitet* - kakav je odnos između našeg offline i online identiteta?; 2) *Privatnost* – tko bi trebao kontrolirati pristup podacima?; 3) *Vlasništvo* - tko posjeduje podatke, koje su obveze ljudi koji koriste te podatke?; 4) *Ugled* - kako možemo odrediti koji su podaci pouzdani? (Davis, 2012: 9)

3.2. Poslovna etika

Kao što je već navedeno, rast velikih podataka tijekom posljednjeg desetljeća otvorio je vrata mnogim mogućnostima i prijetnjama. Veliki podaci nisu samo veliki i moćni, već mogu biti podložni i pogreškama. Trenutačno možemo obraditi veliku količinu podataka munjevitom brzinom. To predstavlja puno prilika, ali također znači da smo u opasnosti donijeti loše odluke u kratkom vremenskom razdoblju, s utjecajem većim od onoga što je čovječanstvo ikada zamišljalo u prošlosti. Tvrtke i organizacije koje koriste velike podatke i druge oblike tehnologije moraju se pridržavati etičkih načela koja su povezana s njima. U tom kontekstu možemo i govoriti o poslovnoj etici kao načinu poslovanja. Kristijan Krkač i Đuro Njavro (2016: 13) u knjizi *Poslovna etika, korporacijska društvena odgovornost i održivost*, u predgovoru pišu da je etika poslovanja kao etika pokera, a u pokeru je blefiranje i varanje dopušteno. A Borna Jalšenjak i Kristijan Krkač (2016: 7) objašnjavaju da je etika poslovanja eksplicitno, teorijsko-praktično i kritičko promišljanje morala u sferama poslovanja ili poslovne administracije. Martina Gregorić (2015: 8 prema: Bebek, 2005) govori da etika u poslovanju može biti raznolika i da se ona odnosi na sam posao u interakciji sa svijetom u cjelini, ali i na interakciju i poslovanje jedan-na-jedan s kupcima. Dalje navodi da je temelj poslovne etike ponašanje u skladu sa zakonom.

Jalšenjak i Krkač (2016: 448) navode da se etika kao filozofska disciplina koja proučava načela moralno dobrog djelovanja čovjeka izučava odavno, ali etičnost u organizaciji, na radnom mjestu i općenito u poslovanju tek se u novije vrijeme promatra kao respektabilni činitelj uspješnosti i profitabilnosti. Također tvrde da država pozitivnim zakonskim propisima unosi etičke normative u poslovanju radi zaštite svojih građana od neetičkih poslovanja.

U jednome se članku Instituta poslovne etike iz Londona (2016) navodi da je IoT[9] omogućio tvrtkama da prikupljaju podatke na način koji korisnici ne mogu u potpunosti razumjeti. Kao rezultat toga, neki korisnici se osjećaju stalno pod kontrolom „Velikog brata“, koji služi gospodarskom interesu poduzeća i nad kojim oni imaju malo ili nimalo kontrole. Ova percepcija dovela je do onog što je označeno kao „deficit povjerenja u podatke“: istraživanje pokazuje da je povjerenje javnosti u tvrtke niže od općeg povjerenja što može negativno utjecati na ugled poduzeća i osiguravajućih društava. Jalšenjak i Krkač (2016) govore o važnosti društvenih medija i prikupljanja podataka od korisnika, što zapravo prerasta u istraživanje tržišta. Također napominju da postoji sve veći broj znanstvenih istraživanja koja se bave negativnim aspektom personalizacije putem interneta i etičnosti zadiranja u privatnost korisnika. Iako personalizacija može povećati korist prikazivanog oglasa, ona može također izazvati osjećaj ranjivosti potrošača koji iskorištavanje osobnih podataka promatraju kao zadiranje u njihovu privatnost, te u konačnici zbog toga mogu razviti negativne osjećaje prema samom oglašivaču.

Posljednjih godina raste potreba za velikim podacima, bilo u poslovanju ili zdravstvu. Kako smo naveli, važno je govoriti o njihovoj etičnosti u poslovanju i o štetnosti, odnosno koristi i pitanjima mogu li nama ljudima veliki podaci pomoći i odmoći. Oni su moćni izvori za potencijalno golemu ekonomsku i društvenu vrijednost. Varon Grover, Roger H. L. Chiang, Ting-Peng Liang i Dongsong Zhang (2018: 390, prema: PromptCloud 2016) navode da su veliki podaci narasli iz industrije vrijedne 6,8 milijardi dolara na nevjerojatnih 32 milijarde dolara u samo tri godine. S obzirom na visoki rast, razne tvrtke i informacijske tehnologije (IT) prihvaćaju jedinstvene prilike da kapitaliziraju velike podatke kako bi stekli konkurentsku prednost. Autori upućuju na to da su tvrtke izvijestile kako troše 10 posto svog IT proračuna samo na podatke i tako prolaze revoluciju iskorištavanja velikih podataka i analitike kao stratešku imovinu, koja će voditi njihovo donošenje odluka i poboljšati poslovne procese i rezultate.

Već smo pisali o problemu velikih podataka i privatnosti korisnika, kao neetičnom poslovanju, ali nažalost postoje razni primjeri istraživača i tvrtki koje koriste podatke bez obzira na privatnost pojedinca. Jeff Collman i Sorin Adam Matei (2016: 83) objašnjavaju da mnoge tvrtke koriste podatke o kupnji za slanje ciljanog oglašavanja i da se u načelu

[9] Internet stvari (engl. *Internet of things*) označava povezivanje uređaja putem interneta. Spajanje uređaja može biti bežično i omogućava nove mogućnosti za međusobnu interakciju ne samo između različitih sustava i donosi nove mogućnosti njihove kontrole, kao i praćenje i pružanje naprednih usluga. To mogu biti, na primjer, hladnjaci, perilice ili razni sigurnosni sustavi.

korištenjem internih podataka o klijentima ne krši zakon o privatnosti već samo ciljanje može biti neetično. Kao što smo naveli u uvodu, primjer takvog ciljanja je izvršio Target. Autori navode da je Target mogao kombinirati demografske podatke s podacima o kupnji za oko 25 proizvoda te su na taj način vršili identifikaciju trudnica. Njihova analitika je bila dovoljno precizna za predviđanje termina rođenja djeteta, ali i samog tržišta na temelju tog zaključka. Dok se na površini čini da je znanje nečega privatnog o nama kao klijentima pohvalno u smislu odnosa proizvod-kupac, u ovom slučaju Target je odabrao plasirati kupone za bilo koju ženu za koju su predvidjeli da je trudna, uključujući i tinejdžerice. Prema standardima većine ljudi, to nije etički prikladno i nije prihvatljivo.

Prije deset godina, online praćenje provedeno je pomoću jednostavnih "kolačića" koji bi bilježili kada je korisnik posjetio neku web lokaciju i što je na njoj tražio. Sada napredni alati ne samo da mogu pratiti ponašanje pregledavanja, nego mogu povezati to ponašanje s osobnim korisničkim podacima, uključujući lokaciju, demografske podatke, pa čak i zdravstvene podatke (Collman i Matei, 2016: 85 prema: Valentino-Devries, 2010; Olsen, 2002). Praćenje, ciljano oglašavanje, narušavanje privatnosti i prava na pristup informacijama Jalšenjak i Krkač (2016: 476) smatraju gorućim pitanjima današnjeg etičnog marketinga. Među ključnim etičkim dilemama ističu se:

- pitanja vezana uz privatnost umreženih potrošača,
- obmanjivanje i lažno predstavljanje,
- distribucija nepoželjne pošte i sadržaja,
- zlouporaba otvorenih inovacija i *crowdsourcinga*[10],
- intelektualno vlasništvo i društveni mediji.

Marketinške teorije uvijek će počivati na dva postulata – zadovoljavanju potreba potrošača s jedne strane i ostvarivanju profita za tvrtku s druge strane – dok se digitalne tehnologije nastoje dokazati kao idealna platforma za postizanje oba marketinška cilja (Jalšenjak, Krkač, 2016).

[10] Otvorene inovacije ističu se prije svega boljim razumijevanjem potrošačevih potreba, prikupljanjem važnih informacija itd. Kada se na sudjelovanje u otvorenim inovacijama poziva mnoštvo ljudi izvan same kompanije, a proces povezivanja se odvija putem *online* medija, onda je riječ o novijem trendu poznatijim pod nazivom *crowdsourcing.* Ovaj se pojam prvi put pojavljuje 2006. godine u članku časopisa Wired, u kojem autor Jeff Howe *crowdsourcing* definira kao „novi *online* poslovni model koji nastoji iskoristiti prednosti kreativnih rješenja velikog broja umreženih pojedinaca putem svojevrsnog otvorenog poziva na suradnju.“

3.3. Zakon o zaštiti osobnih podataka

Kada se spomene privatnost prije svega mislimo na privatnost pojedinca, kao i na otkrivanje osobnih podataka i identiteta. Uredba Europske unije, to jest *Zakon o zaštiti osobnih podataka,* koji je stupio na snagu u svibnju 2018. godine određuje da svatko ima pravo na privatnost. Opća uredba o zaštiti podataka ili GDPR skup je propisa koji želi zaštititi osobne podatke pojedinca, a zakon se odnosi na cijelu Europu. Cilj zakona je omogućiti pojedincu veću kontrolu nad osobnim podacima. Svatko ima pravo u svakom trenutku dobiti informaciju koje podatke neka organizacija o njemu ima i dati privolu da se ti podaci obrađuju ili zatražiti „zaborav" tih podataka.

Jozo Čizmić i Marija Boban (2018: 379) navode da su i ranije postojale direktive i zakonodavni i regulativni okvir zaštite osobnih podataka, no nova Uredba o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom kretanju takvih podataka bitan je napredak u području zaštite osobnih podataka, budući da se njom osigurava ujednačeno i jednoobrazno postupanje nadzornih tijela za zaštitu osobnih podataka, što će imati za posljedicu jednostavniju i jednaku zaštitu prava svih pojedinaca u Europskoj uniji. Također, uvode se nove i pojednostavljuju se već postojeće definicije, određuju se biometrijski i genetski podaci, preciznije opisuju postojeći pojmovi, jačaju prava ispitanika te se smanjuju i pojednostavljuju pojedine administrativne obveze voditelja zbirke osobnih podataka, jačaju nadzorne ovlasti te mogućnost izricanja kazni od (strane) tijela za zaštitu osobnih podataka. Autori također naglašavaju da je jako važno istaknuti kako GDPR zamjenjuje Direktivu 95/46/EZ Europskog parlamenta i Vijeća Europske unije o zaštiti pojedinaca u vezi s obradom osobnih podataka i o slobodnom protoku takvih podataka te stupa na snagu danom donošenja i izravno se primjenjuje u svim državama članicama EU-a.

Direktiva 95/46/EZ	GDPR
• Identitet kontrolora • Svrhe obrade • Obaveza odgovoriti na subjekta podataka • Pravo pristupa, ispravaka i prigovora • Primatelji • Prijenos podataka	• Identitet kontrolora i DPO-a • Svrha • Razdoblje čuvanja podataka • Pravo pristupa, ispravaka, ograničenja i prigovora • Pravo na podnošenje žalbe • Primatelji • Prijenosi • Pravo povlačenja suglasnosti u bilo kojem trenutku • Legitiman interes kontrolora ili treće osobe (ako je relevantno) • Informacije o profiliranju • Sve ostale informacije koje jamče zakonitost prerade

Tablica 1. Usporedba razlika između Direktive 95/46/EZ i nove Opće uredbe o zaštiti podataka (GDPR)11

(J. Čizmić i M. Boban, 2018: 381)

Ključan dio za obradu osobnih podataka jest „privola“ osoba na korištenje osobnih podataka koja se smatra jasnim činom odobrenja. Naime, „privola“ ispitanika znači svako dobrovoljno, posebno, informirano i nedvosmisleno izražavanje želja ispitanika kojim on izjavom ili jasnom potvrdnom radnjom daje pristanak za obradu osobnih podataka koji se odnose na njega (Čizmić, Boban, 2018: 382 prema: GDPR, čl. 4. st. 11.). Novost je činjenica da su u slučaju proboja sigurnosti podataka tvrtke dužne obavijestiti nadležne službe, ali i pojedinca čiji su osobni podaci povrijeđeni što ranije nije bio slučaj. Najvažniji iskorak je pravo na brisanje, poznato i kao pravo na zaborav (engl. Right to be forgotten), sukladno čl. 17. GDPR-a. Načelo ovog prava je omogućiti pojedincima da zatraže brisanje ili uklanjanje osobnih podataka ako nema uvjerljivog razloga za njihovu obradu. Sukladno čl. 17. st. 1.

Čizmić i Boban (2018: 387) objašnjavaju da je najvažniji naglasak na kaznenim odredbama. Dalje navode da nepoštivanje odredbi Uredbe povlači kazne i to do 4 % ukupnog

godišnjeg prometa na svjetskoj razini. Za razliku od ranije, odnosit će se na sve tvrtke koje posluju na području Europske unije (a ne samo one koje su registrirane u EU). Europska komisija je također jasno dala do znanja da za ozbiljno kršenje Uredbe neće biti milosti. Kaznu od 110 milijuna eura koju je poznata društvena mreža Facebook dobila početkom 2018. godine, propisala je izravno Europska komisija. Iako na temelju potpuno druge regulative, ova kazna je propisana upravo zbog kršenja privatnosti građana i pružanja lažnih informacija o spajanju s WhatsAppom. Tom su prilikom predstavnici Facebooka izjavili kako se osobni podaci korisničkih računa ne mogu spojiti, ali Facebook je učinio upravo to. Spojio je račune WhatsAppa s računima Facebooka. Europska komisija je reagirala i propisala kaznu od 110 milijuna eura; 0,5 % ukupnih prihoda Facebooka na globalnoj razini i 50 % maksimalne moguće zarade koju propisuje regulativa o spajanju kompanija. Italija je za isti prekršaj kaznila WhatsApp i time popunila državni proračun za tri milijuna eura (Čizmić, Boban, 2018: 387).

Collman i Sorin Adam (2016: 26) navode da je transparentnost jedan od faktora u određivanju etičnosti pri prikupljanju i analizi Velikih Podataka. U komercijalnom kontekstu transparentnost od korisnika zahtijeva da bude u potpunosti obaviješten i detaljno informiran kako se podaci prikupljaju, što je učinjeno ili što će biti učinjeno s njima. Također, otkrivanje trećim stranama, kao što su vladine agencije, slučajevi koji uključuju kaznena djela (droga) trgovanja ili seksualnog iskorištavanja na otvorenim društvenim mrežama, materijalna podrška terorističkim organizacijama kroz prikupljanje sredstava, itd., kršenje intelektualnog vlasništva, zlostavljanje ili nasilje, treba smatrati legitimnim vrstama otkrivanja. Transparentnost bi trebala uključivati više od informiranja korisnika o metodama koje će se koristiti za zaštitu identiteta, ali i uvjeravanje da njegov identitet neće biti otkriven čak i u situacijama koje su trenutno u okviru kaznenog prava.

U radu je već ranije bilo naglašeno da po pitanju regulative, a vezano uz zaštitu podataka, postoji razlika između EU i Amerike. Collmann i suradnici (2016) pišu da američki državljani imaju snažan osjećaj da mnogi entiteti prikupljaju informacije o njima. U doba interneta, prikupljanje podataka je trajna i sveprisutna praksa. Nakon pojave društvenih medija poput e-pošte, blogova, Facebooka i Twittera, građani su sve više postali svjesni da i oni sami imaju sposobnost praćenja aktivnosti moćnijih, čime se podupiru prijedlozi da velike organizacije budu otvorenije i odgovornije.

No, brzina kojom organizacije prikupljaju i analiziraju informacije dovodi do toga da se stvari prebrzo mijenjaju i da javnost uopće ne stigne dati informirani pristanak o prikupljanju. To od javnosti zahtijeva veću razinu povjerenja u ove institucije nego ikada

prije, uključujući mnoge slučajeve implicitnog odobrenja ili jednostavne slijepe vjere u najbolje namjere korporacija i vladinih organizacija. Autori napominju da gotovo nijedan korisnik ne čita softversku licencu (prema: Bakos et al., 2014). Taj je problem očigledan deset godina i rezultirao je ozbiljnim kršenjima privatnosti od strane korisnika koji su napravili neopravdane pretpostavke o tome jesu li pružatelji softvera dobronamjerni. Nerazumno je očekivati da će prosječni korisnik imati istinsko razumijevanje o tome kakve se informacije prikupljaju o njemu i kako će se one koristiti (Collman Matei i Matei, 2016: 44).

Varon Grover, Roger Chiang, Ting-Peng Liang i Dongsong Zhang (2018: 415) objašnjavaju da veliki podaci moraju biti zaštićeni kako bi se spriječile zlonamjerne povrede podataka. Navode da rastuća kibernetička sigurnost znači da organizacije trebaju očekivati i pripremiti se za kršenje podataka te osmisliti i provesti mjere za pravovremeno otkrivanje takvih kršenja, ali i minimizirati njihove negativne učinke. Od organizacija se očekuje da poštuju propise o sigurnosti podataka i da kada dođe do povrede podataka izvijeste potrošače o kojima je riječ.

Kada se govori o GDPR-u, treba spomenuti dio zakona koji govori o obradi genetskih podataka. Tako Masha Sabani i Pascal Borry (2018: 149) objašnjavaju da genetski podaci sadrže osjetljive zdravstvene i nezdravstvene informacije o pojedincima i članovima njihovih obitelji, pa je stoga usvajanje odgovarajućih mjera zaštite privatnosti od najveće važnosti pri obradi genetskih podataka u istraživačke ili kliničke svrhe. Jedan od glavnih pravnih instrumenata za zaštitu osobnih podataka u EU je nova Opća uredba o zaštiti podataka (GDPR), koja je ukinula Direktivu 95/46/EZ, s krajnjim ciljem povećanja učinkovitosti i usklađivanjem zaštite osobnih podataka. Čizmić i Boban (2018: 382) navode da *genetski podaci* predstavljaju osobne podatke *„koji se odnose na naslijeđena ili stečena genetska obilježja pojedinca koja daju jedinstvenu informaciju o fiziologiji ili zdravlju tog pojedinca, i koji su dobiveni osobito analizom biološkog uzorka dotičnog pojedinca“* (prema: GDPR, čl. 4. st. 13.) dok se pod pojmom *biometrijski podaci* podrazumijevaju *„osobni podaci dobiveni posebnom tehničkom obradom u vezi s fizičkim obilježjima, fiziološkim obilježjima ili obilježjima ponašanja pojedinca koja omogućuju ili potvrđuju jedinstvenu identifikaciju tog pojedinca, kao što su fotografije lica ili daktiloskopski podaci”* (prema: GDPR, čl. 4. st. 14.). Masha Sabani i Pascal Borry u svom članku *Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation* (2018) navode da su izmjene i dopune Nacrta Komisije na polju genetskih istraživanja, od strane Europskog parlamenta dovele do velike zabrinutosti među biomedicinskim i zdravstvenim istraživačkim zajednicama koje su predložene zahtjeve smatrali preprekom za istraživanja. Autori

nastavljaju da treba napomenuti kako GDPR priznaje izazove dobivanja konkretnog pristanka za istraživanje u svrhu prikupljanja podataka te da bi subjektima trebalo dopustiti da daju pristanak određenim područjima znanstvenog istraživanja. Uredba navodi da *„države članice mogu zadržati ili uvesti dodatne uvjete, uključujući ograničenja, u odnosu na obradu genetskih podataka, biometrijskih podataka ili podataka koji se odnose na zdravlje"* (čl. 9.4). Države članice mogu tražiti jaču zaštitu genetskih podataka i strože uvjete za obradu genetskih podataka za svrhe istraživanja. Autori dalje navode da jedan dio zakona upućuje na to da se istraživanja u znanstvene i povijesne svrhe ili u statističke svrhe treba smatrati kompatibilnom obradom. To znači da će retrospektivna uporaba genetskih baza podataka biti dopuštena, čime se optimizira korištenje već prikupljenih podataka za buduće istraživačke svrhe. Obrada osobnih podataka treba biti transparentna i poštena, a ispitanici trebaju unaprijed dobiti relevantne informacije o daljnjoj obradi.

Što se tiče definicije znanstvenog istraživanja, treba napomenuti da Uredba favorizira široko tumačenje prema kojemu se *„obrada osobnih podataka u znanstveno-istraživačke treba tumačiti na širok način, uključujući, na primjer, tehnološki razvoj i demonstracije, temeljna istraživanja, primijenjena istraživanja i privatno financirana istraživanja."* Zabrinutost zbog moguće zlouporabe izuzeća istraživanja od strane komercijalnih aktera dovelo je do rasprave između Biobanking i BioMolecular resources Research Infrastruktura - Konzorcija europske istraživačke infrastrukture (BBMRI-ERIC) koji zahtijevaju da se precizira znanstveno istraživanje koje traži javni interes. Doprinos javnih i privatnih subjekata u unapređenju biomedicinskih istraživanja je esencijalan, te se oba privatna i javno financirana istraživanja mogu smatrati korisnima, sve dok su u skladu s javnim interesima i dok je postupak obrade podataka transparentan.

4. Veliki podaci za veće dobro (Data for good) ili podaci kao kršitelji privatnosti?

U raznim znanstvenim člancima, medijima i knjigama pisalo se o benefitima, to jest koristima u korištenju velikih podataka, ali i o njihovoj lošoj strani koja je vezana uz korisnika i osobne podatke. Kaže se da veliki podaci imaju dvojnost, dobru stranu i lošu stranu. Iako imaju nevjerojatan potencijal za poboljšanje svijeta, lako može doći do njihove

zloupotrebe, isključivo u svrhu praćenja ponašanja kako bi se zaradio novac. Isto tako, može doći do manipulacije ljudi putem velikih podataka.

U pogledu korištenja velikih podataka treba krenuti od 'manjih stvari', od ciljanog marketinga od kojeg nitko od nas ne može pobjeći. U tom smislu Davis (2012) raspravlja da su mnogi ljudi zabrinuti, jer ih njihova povijest pretraživanja „slijedi“ putem tražilica, no on govori da postoje realni scenariji koji pružaju izravnu korist za nas kao potrošače. Budući da će korisnici vidjeti digitalno oglašavanje, u tom bi slučaju mogli vidjeti i oglase za stvari koje ih mogu zainteresirati, kao što je na primjer pregledavanje aktivnosti odmora na Bermudskim otocima. U tom slučaju nam ciljano oglašavanje može pomoći kako bismo vidjeli niže cijene hotela ili nekih drugih aktivnosti te navodi da svatko od nas sigurno želi uštedjeti novac.

Collman i Matei (2016: 85) pišu da je uznemirujuće to što oglašivači plaćaju tvrtkama da prate ljude koji se koriste internetom kako bi bolje razumjeli koje web stranice posjećuju ili koje aplikacije koriste (prema: EPIC 2015). Iako se te informacije mogu koristiti za ciljane oglase, to se također može koristiti za ciljane oglase na pristran način, što znači da se koriste demografski podaci, a cijene se prilagođavaju za različite podskupine. Časopis Wall Streeta je 2010. godine proveo studiju o tehnologijama praćenja i otkrio da su na 50 najboljih internetskih stranica prosječno instalirana 64 razna tehnološka alata za praćenje, bez ikakvog upozorenja. Tako zdravstvene i farmaceutske tvrtke, primjerice, koriste zdravstvene pojmove i web lokacije koje korisnici pretražuju, a odnose se na zdravlje (Collmann i Matei, 2016: 86, prema: Angwin 2010).

Kada je riječ o većim mogućnostima vezanim uz iskorištavanje velikih podataka, Kristina Kocijan (2014) navodi da je Engleska vlada 2013. dodijelila 189 milijuna funti za istraživanja u području velikih podataka. Sintetička biologija, kao prva sljedeća disciplina po količini dodijeljenih novaca, dobila je 88 milijuna funti. Maris Mirović, Mario Miličević i Ines Obradović (2018: 57) navode da se veliki podaci sve češće koriste u medicini, pa su za primjer uzeli snimanje i analizu otkucaja srca i uzorak disanja pacijenata, nakon kojih su predvidjeli infekciju 24 sata prije nego je došlo do pojave fizičkih simptoma. Smatraju da ne postoji područje koje ne bi imalo koristi od korištenja velikih podataka te da svijest o njihovoj vrijednosti i potencijalu i dalje raste. Autori navode još primjera, kao što su podaci dobiveni prometnim senzorima koji se mogu koristiti za identifikaciju problema u stvarnom vremenu, što znači da vozači mogu donositi odluke na račun informacija, a sve kako bi uštedjeli vrijeme, dok nadležna tijela za ceste mogu kontrolirati promet i brzo intervenirati kada je to potrebno. Los Angeles, na primjer, koristi podatke za kontrolu semafora, što je smanjilo zagušenje prometa za procijenjenih 16%. U Dublinu se podaci prikupljeni od prometnih

senzora, GPS uređaja i drugih izvora koriste za izgradnju digitalne karte gradske prometne mreže u stvarnom vremenu. Kao rezultat, poboljšan je prometni tijek u gradu (Mirović, Miličević, Obradović, 2018: 57).

Tehnike velikih podataka mogu se također koristiti za praćenje plovila u stvarnom vremenu i za analizu performansi broda i navigacijskih podataka prikupljenih iz senzora i sustava za prikupljanje podataka. Dobiveni uvidi omogućuju razvoj strategija optimalnog dizajna logističke mreže i energetske učinkovitosti. Uz smanjenje troškova i koristi za okoliš, također se može poboljšati i sigurnost na moru. Otkrivanje anomalija omogućuje identificiranje problema u realnom vremenu kako bi se oni mogli riješiti što je prije moguće. Međutim, bavljenje velikim podacima znači i suočavanje s problemom količine i kvalitete podataka. Osim toga, postoje sigurnosni i privatni izazovi koji zahtijevaju uspostavu zakonodavnog okvira koji će diktirati upravljanje podacima. Kao primjer, recimo da pomorska industrija može imati velike koristi od velikih podataka, ali je prije nego što ih je moguće iskoristiti do punog potencijala potrebno prevladati puno izazova (Mirović, Miličević, Obradović, 2018).

Kada je riječ o korištenju velikih podataka u zdravstvene svrhe Silvije Vuletić u svom članku *Što očekuje javno zdravstvena profesija od „Big Data" tehnologije?* (2018: 27) smatra da je potrebno prije svega postići konsenzus profesionalnih i edukacijskih institucija (npr. medicinski fakulteti i zdravstvena veleučilišta te zavodi za javno zdravstvo) oko intersektorskih i interdisciplinarnih djelovanja (npr. specijalizacije unutar javnog zdravstva, te preventivne i kurativne medicine). Edukacijske ustanove trebaju ugraditi osnove za buduće podatkovne znalce ili znanstvenike u svoje kurikule i kroz specijalizacije omogućiti njihovo usavršavanje. Također navodi da u neposrednoj praksi treba upoznavati javnozdravstvene profesionalce s projektima izabranih zemalja Europske unije. Hrvatsko javno zdravstvo treba početi s definiranjem potencijalnog programa u području *„Big Data"* te pripremiti odgovarajuću znanstvenu aplikaciju. Osnovni edukacijski cilj bio bi upoznavanje i razumijevanje novog modela procesiranja javnozdravstvenih podataka.

Collmann i Matei (2016: 54) drže da istraživanje velikih genomskih podataka može izložiti subjekte psihološkim, socijalnim i ekonomskim štetama, osobito ako istraživanje otkriva osjetljive informacije o ponovno identificiranim pojedincima ili rasnim / etničkim / geografskim skupinama s kojima se identificiraju. Psihološka oštećenja uključuju neželjene promjene u misaonim procesima i emocijama (npr. epizode depresije, zbunjenosti, osjećaja stresa, krivnje i gubitka samopoštovanja). Društvene i ekonomske štete mogu uključivati neugodnost unutar poslovne ili društvene skupine sudionika, gubitak zaposlenja ili kazneni

progon uzrokovan, na primjer, napadima na privatnost i kršenjem povjerljivosti. Osim toga, neka društvena i bihevioralna istraživanja mogu dati informacije o pojedincima koji bi mogli „obilježiti“ ili „stigmatizirati“ subjekte, bilo kao pojedince ili kroz udruživanje s određenom skupinom. Kako ističu autori, iako se ove štete često navode kao razlozi za zabrinutost zbog genomskih istraživanja, dokazi o tim štetama do sada su prilično niski.

4.1. Edward Snowden i novinarstvo

U svibnju 2013. godine Edward Snowden u javnost je pustio najveću zbirku strogo povjerljivih dokumenata, što se nije dogodilo nikada ranije. Vodi se kao bjegunac kojeg traže Sjedinjene Države, a trenutno se nalazi u Rusiji. Curenje podataka je dovelo do novog razumijevanja do koje mjere je američka vlada prikupljala podatke o svojim građanima što je potaknulo međunarodni razgovor o digitalnoj privatnosti. Nedavna otkrića o Facebooku i Cambridge Analytici pokazuju koliko su ljudi daleko od toga da znaju kako se koriste njihovi podaci, a još manje da se kontroliraju.

U svom intervjuu za britanski The Guardian 04.06.2018. godine, Snowden ističe: *„Vlada i korporativni sektor su se bavili našim neznanjem. Ali sada znamo - kaže Snowden. Ljudi su sada svjesni. Ljudi su još uvijek nemoćni da ga zaustave, ali mi pokušavamo.“*

Paul Lashmar (2016: 3) u svom članku *No More Sources? The impact of Snowden's revelations on journalists and their confidential sources*, navodi da je tadašnji urednik britanskih novina The Guardian Alan Rusbridger upozorio kako ovakva razina nadzora ima ozbiljne posljedice za novinarstvo: *„Svaki novinar bi trebao razumjeti da ne postoji takva stvar kao što je povjerljiva digitalna komunikacija. Nitko od nas nema povjerljive izvore. Svi ćemo morati raditi na tome, u ovom svijetu gdje ljudi mogu presresti sve.”*

Edward Snowden (rođen 1983.) američki je računalni stručnjak, bivši djelatnik američke Središnje obavještajne agencije (CIA) i izvođač radova Agencije za nacionalnu sigurnost (NSA). On je od kraja 2012. godine neovlašteno kontaktirao novinare Glenna Greenwalda i Lauru Poitras (Lashar, 2016). Autor dalje piše da su od lipnja 2013. dokumenti objavljivani u različitim medijima diljem svijeta, a najpoznatiji su The Guardian (Velika Britanija), Der Spiegel (Njemačka), The Washington Post i New York Times (SAD), O Globo (Brazil), Le Monde (Francuska), kao i mediji u Švedskoj, Kanadi, Italiji, Nizozemskoj,

Norveškoj, Španjolskoj i Australiji. Dokumenti otkrivaju operativne detalje globalnog aparata za nadzor koji zajednički vode Five Eyes[11] u uskoj suradnji sa svojim komercijalnim i međunarodnim partnerima. Američki istraživački novinar Seymour Hersh siguran je da je Edward Snowden „promijenio prirodu rasprave“ o nadzoru. Hersh je rekao da su on i drugi novinari pisali o nadzoru, ali Snowden je bio važan jer je pružio dokumentirane dokaze. Glenn Greenwald, bivši novinar The Guardiana, koji je surađivao s Edwardom Snowdenom, sumirao je njegovu percepciju djelovanja mreže NSA, ističući da je sada očito da se metapodaci[12] većine poruka e-pošte, brojni telefonski pozivi i još mnogo toga kopiraju u velike pohranjene podatke, koji agencijama dopuštaju da pregledaju korisne informacije. Snowden je također otkrio da su agencije tajno pregovarale za „backdoors“ u sigurnosti mnogih računalnih programa, društvenih mreža, web stranica i pametnih telefona (Lashar 2016, prema: Greenwald, 2013).

Kada se govori o etičnosti zaštite izvora, Lashar (2016: 5) objašnjava da je zaštita izvora ključni koncept u novinarstvu. Svi glavni etički kodeksi za novinare širom svijeta pokrivaju zaštitu izvora. Također, dalje piše da medijski komentator Tim Luckhurst govori da se povjerljivi izvori nikada ne smiju identificirati dok su živi (Lahar, 2003: 6). Autor napominje da ova dužnost podjednako vrijedi i za druge zemlje. Dalje navodi da je većina izvora otvorena, ali neki budu opisani kao anonimni ili povjerljivi. To se može ili ne mora spominjati u novinarskoj priči, ali nikad po imenu. Lashar nastavlja te upućuje na mišljenje Matt Carlsona, koji iako je kritičan prema onome što vidi kao prekomjerno korištenje neimenovanih izvora, smatra da anonimni izvori dopuštaju novinaru da se suprotstavi informaciji koja dolazi od poznatih izvora te im dozvoljava da tajnu informaciju plasira u javnost. To dopušta novinarima da na drugačiji način postupe s informacijom koja može biti važna u određenom procesu. Jednako tako, upućuje na to da postoje brojni trenutci u kojima su novinari pustili vijest značajne političke važnosti korištenjem neimenovanih izvora (Lashar, 2016, prema: Carlson, 2012, 6-7).

Emily Bell i Taylor Owen (2017: 22) navode da je prije slučaja sa Snowdenom određena povlaštena manjina sigurno znala za metapodatke, a sada nakon Snowdena nema

[11] The Five Eyes, često skraćeno FVEY, anglofonski je obavještajni savez koji obuhvaća Australiju, Kanadu, Novi Zeland, Ujedinjeno Kraljevstvo i Sjedinjene Države.

[12] Metapodatci, podaci koji pomažu u identifikaciji, opisu i lociranju nekih podataka, napose umreženih elektroničkih izvora. Naziv metapodaci (engl. *metadata*) skovao je Jack Myers potkraj 1960-ih i registrirao ga kao robni znak istoimene tvrtke *(Metadata Company)*. Polovicom 1990-ih u informatičkim se krugovima naziv počeo koristiti za podatke kojima se omogućuje lakše snalaženje među računalnim datotekama, posebno onima na područjima prirodoslovlja, društvenih znanosti i geoinformacijskih sustava (stranica posjećena: 1. 6. 2019., http://www.enciklopedija.hr/Natuknica.aspx?ID=40384).

izgovora za bilo koga u novinarstvu (ili zakonu, ili medicini, ili bilo kojoj profesiji). Bivši glavni savjetnik NSA izjavio je u raspravi što je u organizaciji Guardiana održana 2013. u New Yorku da metapodaci apsolutno govore sve o nečijem životu (Emily Bell i Taylor Owen, 2017: 22).

5. Medijsko pravo na pristup informacijama

U hrvatskom *Vodiču Pravo na pristup informacijama: za medije i udruge* (2018) stoji da je „*pravo na pristup informacijama jedno je od ključnih instrumenata svakodnevne borbe protiv korupcije putem transparentnosti te povećanja razine integriteta i odgovornosti demokratskih institucija svih razina vlasti.“* To znači da javnost posjeduje pravo da zna kako demokratska vlast u njezino ime odlučuje i troši javni novac i da to podrazumijeva postojanje efikasnog sustava medijskog izvještavanja i informiranja javnosti o pitanjima koja su od važnosti za društvenu zajednicu i građane. Mediji bez pristupa informacijama ne mogu efikasno obavljati svoj posao te funkcionirati kao čuvari demokracije.

Pravo na pristup informacijama je Ustavom Republike Hrvatske i Zakonom o pravu na pristup informacijama zajamčeno pravo svakoga građanina i pravne osobe na pristup informacijama koje posjeduju tijela javne vlasti u Republici Hrvatskoj, neovisno o svrsi u koju se informacija želi koristiti. To znači da svaki građanin (fizička osoba, u bilo kojem svojstvu), kao i svaka pravna osoba, pa tako i mediji, novinari, udruge, savezi udruga, imaju pravo na informaciju koja je javnog karaktera. To pravo možete ostvariti bez obzira jeste li državljanin ili stranac, odnosno imate li kao pravna osoba sjedište u RH ili u inozemstvu (Pristup na Informacije: za medije i udruge 2018: 6). Ovdje je važno napomenuti kako korisnik nije dužan navesti svrhu zbog koje traži pristup informaciji te da korisnik može s informacijom slobodno raspolagati, odnosno informaciju može javno objaviti.

Informacije koje su tijela javne vlasti dužna objavljivati na svojim stranicama propisano je u članku 10. Zakona o pravu na pristup informacijama, a mogu se svrstati u četiri kategorije:

- Informacije vezane za donošenje odluka: opći akti i odluke, nacrti zakona i drugih propisa te općih akata za koje se provodi postupak savjetovanja s javnošću, zaključci sa službenih sjednica i na njima usvojeni dokument;
- Informacije vezane za okvir rada i planiranje: zakoni i drugi propisi, godišnji planovi, programi, strategije, upute, izvješća, natječajna dokumentacija i ishodi natječajnih postupaka, informacije o unutarnjem ustrojstvu te imena i kontakti čelnika/voditelja;

- Informacije vezane za financije: izvor financiranja, proračun i njegovo izvršenje, informacije o dodijeljenim sponzorstvima i donacijama, informacije o postupcima javne nabave;
- Informacije vezane za pružanje usluga i komunikaciju s korisnicima: informacije o javnim uslugama, registri i baze podataka iz njihove nadležnosti, obavijesti o načinu i uvjetima ostvarivanja prava na pristup i ponovnu uporabu informacija, odgovori na najčešće postavljena pitanja, informacije o načinu podnošenja upita građana i medija, vijesti, priopćenja, itd. (Pristup na Informacije: za medije i udruge 2018: 8).

U nekim slučajevima novinar može odabrati podnošenje zahtjeva za pristup informacijama, osobito u slučaju kad neku informaciju u smislu Zakona o pravu na pristup informacijama (dokument ili drugu vrstu zapisa) ne može dobiti putem Zakona o medijima. Drugi najčešći slučaj je istraživačko novinarstvo, kada će novinar posvetiti dulje vrijeme i više energije prikupljanju različitih informacija i dokaza koji su mu potrebni za izradu neke novinarske priče i želi biti siguran da će ih dobiti, osobito ako su te informacije ‘osjetljive’, u smislu da raskrinkavaju nedozvoljeno ponašanje.

5.1. Data Novinarstvo (podatkovno novinarstvo)

U prethodnom poglavlju govorilo se o pravu na pristup informacijama unutar Republike Hrvatske. To znači da mi kao građani imamo pravo na javne informacije, ali i svaka pravna osoba, pa tako mediji i novinari. Ako je novinaru potrebna neka javna informacija, on ima pravo na nju. Pristup informacijama u sferi novinarstva važan je za dio rada koji upravo slijedi.

Naime, *Data Novinarstvo* je metoda i tehnologija crpljenja informacija, zaključaka i pokazatelja iz niza podataka uz pomoć različitih softvera. *Data Novinarstvo* na određeni način predstavlja budućnost i budućnost novinarstva. Novinar koji bi se želio baviti takvim novinarstvom trebao bi razumjeti kvantitativne podatke.

S obzirom na pravo na pristup informacijama, uzmimo za primjer dokumente o kaznenim djelima. Novinar koji zatraži pristup takvim informacijama, može ustanoviti određene trendove u okvirima rada *Data Novinarstva*. To znači da primjerice takav dokument možemo 'prebaciti' u Excel te putem programskih radnji tražiti i dobivati rezultate i usporedbe koje nas zanimaju.

Tako u vodiču za medije i udruge (2018: 10) stoji da su tijela javne vlasti dužna objaviti svoje registre i baze podataka (ako im pristup nije zaštićen posebnim propisima) kao otvorene podatke, koji su slobodno dostupni svima na korištenje i ponovnu uporabu u strojno čitljivom obliku (otvoreni, strojno čitljivi formati su npr. XLS; CSV, RDX, JSON i dr.)

Korisni otvoreni podaci za novinare:
• izborni rezultati • proračuni lokalnih jedinica • podaci o porezima • podaci o kaznenim djelima • podaci o radu institucija statistički podaci (npr. socio-ekonomski pokazatelji)

Pravo na pristup informacijama: vodič za medije i udruge (2018:10)

Data.Journalism.com [13] postavlja pitanje zbog čega se podatkovno novinarstvo razlikuje od ostatka novinarstva? Odgovor je kombinacija tradicionalnog 'nosa za vijesti' i sposobnost da ispričate uvjerljivu priču, s velikom količinom i rasponom digitalnih informacija koje su sada dostupne. Na stranicama su postavljeni neki od primjera vezanih uz *Data Novinarstvo*, a jedan dio se odnosi na projekt „Tajne ubojstava“ Toma Hargrovea iz Scripps Howard News Servicea. Iz vladinih podataka i javnih evidencija izradio je zahtjev za demografski detaljnu bazu podataka s više od 185.000 neriješenih ubojstava, a zatim je osmislio algoritam za pretraživanje uzoraka koji ukazuju na moguću prisutnost serijskih ubojica. Radi se o projektu koji koristi tehnike društvenih znanosti i interaktivnu prezentaciju podataka na internetu kako bi ih čitatelji mogli sami istražiti.

Colin Porlezza sa Sveučilišta u Londonu (2019: 189) navodi da za razliku od tradicionalnog novinarstva koje je tek nedavno usvojilo koncepte kao što su otvorenost i odgovornost, podatkovno novinarstvo je duboko ukorijenjeno u kulturi otvorenog koda koja

[13] DataJournalism.com je izradio Europski centar za novinarstvo i podržava ga Google News Initiative (stranica posjećena 5. 6. 2019., https://datajournalism.com/).

dolazi s povećanom transparentnošću, podjelom i sudjelovanjem. Podaci o društvu – pokušaj da se sve pretvori u format podataka kako bi se kvantificirali, temeljan je proces jer zamjenjuje, kako u opsegu tako i stilu, način na koji se stvarnost vidi i način na koji je izgrađena. Autor nastavlja te govori da takav tip novinarstva promatra podatke u društvu i povezuje s njima pitanja iz javne sfere i da sve više obuhvaća instrumente koji se temelje na procesima koji uključuju podatke i algoritme. To vodi do novih oblika u novinarstvu koji nude nove mogućnosti, kao što je „češljanje“ podataka iz javnih uprava ili otkrivanje protuzakonitih djela. *Data novinarstvo* uživa sve veću pozornost i uspjeh u medijskoj praksi.

Jednako tako, sve je više znanstvenih članaka koji donose mnoge definicije podatkovnog novinarstva. Tako Porlezza (2019: 190) podatkovno novinarstvo shvaća kao hibridni oblik [novinarstva] koje obuhvaća statističke analize, informatiku, vizualizaciju, web dizajn i izvješćivanje. Hibridnost podatkovnom novinarstvu daje kombinacija različitih uloga i zadataka koje često izvode različiti akteri, a ne samo novinari. Porlezza objašnjava da mnogi kvantitativno orijentirani novinari moraju imati znanje iz informatike i tek nakon toga mogu početi s učenjem u praksi i postati dijelom tima za podatkovno novinarstvo (Porlezza, 2019: 192).

5.2. Etičnost u podatkovnom novinarstvu

Kuutti (2016: 2) objašnjava da se usprkos svom digitalnom karakteru etika podatkovnog novinarstva ne odnosi na tehnička pitanja, nego na izbor novinara u njihovom radu. Podaci su samo alat u novinarstvu i ne uključuju etički sadržaj. Glavno pitanje je sposobnost novinara za ispitivanje novinske teme pomoću podataka i njegove stručnosti za tumačenje podataka koje su prikupili. Autori dalje navode da se pouzdanost podatkovnog novinarstva odnosi na pouzdanost podataka i da se etička razmatranja u podatkovnom novinarstvu ne razlikuju puno od onih u bilo kojem drugom području novinarstva. U načelu, podatkovno novinarstvo susreće iste etičke probleme kao i tradicionalno novinarstvo, kao što su na primjer ugrožavanje privatnosti i neistinito prikazivanje. Međutim, informacijska tehnologija koja se koristi u ovoj vrsti novinarske prakse stvorilo je dodatnu zabrinutost u vezi s pristupom materijalima podataka, korištenjem podataka u analizi, transparentnošću u objavljivanju rezultata i otvorenošću povratne informacije od javnosti i ispravljanjem pogrešaka u priči. Lawrie Zion i David Craig (2014: 203 prema: Bradshaw, 2013) navode da je točnost podatkovnog novinarstva najvažniji etički kriterij, to jest da je najvažnije pružanje

odgovarajućeg konteksta pričama koje se ispričaju. To može utjecati na način na koji analiziramo podatke i na koji izvještavamo na temelju tih podataka. Autori nas upućuju na primjer novinara Nilsa Mulvada koji je krajem 2012. dobio veterinarske recepte, ali on ih je odlučio ne objavljivati, jer je shvatio da su puni pogrešaka. Objavljivanje bi značilo izravno identificiranje određenih liječnika veterine i optužbu za loše postupanje koje nisu učinili. Stoga su koristili podatke kako bi istaknuli loše primjere te na taj način mogli započeti novinarsko istraživanje te ispravnost informacija.

Pristup	Analiza	Objavljivanje	Povratna informacija
• odabir teme • tumačenje • podrijetlo podataka • dovoljno podataka	• zaključak • tumačenje • metoda analize • dovoljnost metode	• drugi izvori • tumačenje • pitanja privatnosti • prozirnost • kontekst • vizualizacija	• vanjske povratne informacije • otvorenost za pogreške • povezanost

Svaka od četiri faze zahtijeva određenu etičku refleksiju
(Heikki Kuutti, 2016: 5)

Jedan od primjera u kojima se etički kodeks razlikuje ovisno o državi i zanimanju, kako navodie Lawrie Zion i David Craig (2014: 208, prema: Bradshaw 2013), istraga je oko dobivanja EU poticaja za stočarstvo. Istraga je bila provedena u suradnji s novinarima iz različitih zemalja koji su imali drugačije koncepte i način gledanja na osobne podatke. Oni su koristili zakone o pristupu informacijama svojih zemalja kako bi identificirali tko su pojedinci koji primaju milijunske iznose. Nakon toga je Europski sud pravde ograničio tiskanje novina na temelju zaštite privatnosti. Također navode drugi primjer koji se odnosi na Ameriku, gdje je pristup osobnim podacima lakši, što ne znači da njihovo objavljivanje ne izaziva etičke probleme. Nakon masovnih ubojstava u Connecticutu, Journal News iz New Yorka je odlučio objaviti interaktivnu mapu o osobama koje imaju dozvolu za korištenje oružja, koja se temeljila na zakonu o pravu na pristup informacijama. Mapa je dovela do smanjenja broja

čitatelja i javnih poziva na bojkot novina (Lawrie Zion i David Craig (2014: 208, prema: Bradshaw, 2013).

6. Metodologija istraživanja

6.1. Ciljevi i hipoteze istraživanja

Glavni cilj ovog istraživanja je analizom sadržaja hrvatskih i inozemnih internetskih portala utvrditi koliko se često na internetskim portalima spominje etičnost u korištenju velikih podataka od 1. siječnja 2016. do 10. svibnja 2019. Specifični cilj se odnosi na usporedbu o načinu izvještavanja na internetskim portalima u inozemstvu i u Hrvatskoj. Također, cilj je utvrditi stupanj pozitivnog i negativnog načina izvještavanja o korištenju velikih podataka.

Hipoteze koje smo postavili, a koje ćemo istraživanjem potvrditi ili opovrgnuti su sljedeće:

H1: Tekstovi na internetskim portalima u Hrvatskoj u manjoj mjeri pružaju informacije o zaštiti i savjetima o upotrebi velikih podataka.

H2: Tekstovi na internetskim portalima u Hrvatskoj u manjoj mjeri izvještavaju o upotrebi osobnih podataka u komercijalne svrhe.

H3: Tekstovi na internetskim portalima u većoj mjeri predstavljaju podatke kao prijetnju za osobnu sigurnost.

H4: Mediji u Hrvatskoj u manjoj mjeri od inozemnih izvještavaju o pozitivnim stvarima upotrebe velikih podataka.

U ovom diplomskom radu kao metoda istraživanja koristit će se analiza sadržaja hrvatskih i inozemnih internetskih portala prema samostalno izrađenoj analitičkoj matrici. Kako bi se odgovorilo na istraživačka pitanja te opovrgnule ili potvrdile postavljene hipoteze, analizirali smo sadržaj hrvatskih portala – 24 sata i Tportal.hr te inozemne portale – BBC i The Guardian. Istraživanje obuhvaća razdoblje od 1. siječnja 2016. do 10. svibnja 2019. godine. Istraživanje će se provesti na uzorku od 198 internetskih članaka koji su objavljeni na hrvatskim i inozemnim internetskim portalima. Riječ je o primarnom izvoru podataka. U radu

se koriste metode grafičkog prikazivanja u kojemu se prezentira struktura promatranih članaka na navedena četiri portala prema odabranim obilježjima. Za prikaz strukture je korišten MS EXCEL 2016.

Almir Elezović (2012: 62) navodi da ideja o praćenju medija (*media monitoring*) predstavlja proširenje metodološkog pristupa u komunikološkom istraživanju poznatom i kao analiza sadržaja. Dodaje da su je klasični teoretičari analize sadržaja, osobito Bernard Berelson i Harold Lasswell, uveli kao instrument pomoću kojeg se može odrediti što mediji doista rade te definirati uređivačka politika prema različitim aspektima društvenog života – uključujući i same medije.

Istraživanje u ovome radu provedeno je metodom kvantitativne analize sadržaja, više od dva desetljeća najčešće korištenom metodom prilikom promatranja zastupljenosti određenih tema ili problema u medijima, to jest trendova u masovnoj komunikaciji. Kvantitativna istraživanja su ona koja pružaju brojčani opis istraživane pojave, redovito putem statističke analize i sažimanjem izvornih podataka (Milas, 2009: 47). Goran Milas dalje navodi da se kvantitativna istraživanja često koriste kao početni korak u upoznavanju i definiranju problema ili kao nadopuna kvantitativnim istraživanjima, jer ona pomažu ispitivaču da se usmjeri na ono što ispitanicima doista znači kritičnu točku koju valja ispitati.

Jednako tako, Milas (2009: 500) ističe da pretpovijest analize sadržaja seže u kasno 17. stoljeće. U tom su dobu zabilježene teološke studije koje su pratile širenje nereligijskih sadržaja putem novina (prema: Krippendorff, 1980). U jednom od prvih takvih napisa pokazalo se kako su religijski, znanstveni i literarni sadržaji od 1881. u vodećim njujorškim novinama počeli sve više ustupati mjesto skandalima i drugim temama. Autor dalje navodi da je slična studija provedena 1910., a ona je upozorila na velik novinski prostor posvećen trivijalnostima u odnosu na ozbiljne sadržaje.

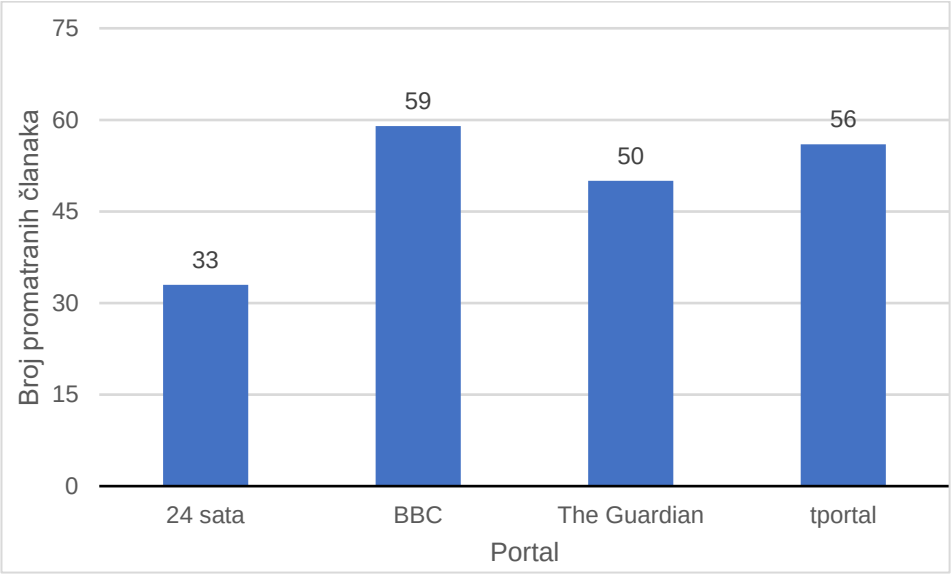
Valja dodati da su masovni mediji važni instrumenti za informiranje širokog auditorija te istovremeno i okvir javnog i privatnog mnijenja, a znanstveno opisivanje i procjenjivanje medijskog djelovanja koje traga za trendovima ideološkog pripovjedačkog diskursa u medijima itekako je potrebno (Elezović, 2012: 64).

Milas (2009) navodi obilježja analize sadržaja, a to su: objektivnost, sustavnost i općenitost. Objektivnost je nezavisnost rezultata istraživanja od istraživača koji ih je dobio, a postiže se praćenjem jasnih i propisanih pravila koja svakom istraživaču omogućuju da, na temelju istog materijala, dođe do sukladnih zaključaka. Sustavnost podrazumijeva pravila prema kojima se izvaci materijala uključuju ili isključuju iz analize pri čemu istraživač ne

smije dokazivati vlastite pretpostavke nego se mora držati unaprijed utvrđenih načela. Općenitost se ogleda u teorijskom značenju koji bi trebao nadići opisivanje sadržaja, a pokazuje znanstvenu vrijednost analize sadržaja (Milas, 2009: 501).

Analiza sadržaja pokazuje se posebno korisnom kada je u pitanju velika opsežnost materijala i kada nije moguće izvršiti cjelovitu analizu i kada istraživaču nisu dostupni nikakvi drugi podaci osim zabilježenih poruka, a njezinom namjenom žele se opisati obilježja komunikacije, zaključivati o događajima koji su joj prethodili kao i o njezinim učincima, prethodno odgovarajući na tri ključna pitanja (Milas, 2009: 503 prema: Holsti, 1968: 599): što komunikacija donosi, kako to čini i kome se obraća. Na pitanje „što“ može se odgovoriti usporedbama kroz vrijeme, između situacija ili između različitih varijabli. Pitanje „kako“ najčešće se dovodi u svezu s promidžbom dok se pitanjem „kome“ nastoji „povezati znana obilježja primatelja i poruke kojom im se obraća“ (Milas, 2009: 504). Lamza Posavec (2004: 171) navodi da se, u najširem smislu, analizom sadržaja smatra postupak proučavanja i raščlanjivanja nekog pisanog ili slikovnog materijala kojim se nastoje uočiti neke osobine odnosno poruke toga sadržaja.

Uzorak istraživanja je odabran pomoću tražilice portala (24 sata, Tportal.hr, BBC i The Guardian) koristeći ključne riječi: „veliki podaci“, „big data“, „etičnost“, „etika“, „etika privatnosti“, „zaštita podataka“, „krađa podataka“, „osobni podaci“, „gdpr“.

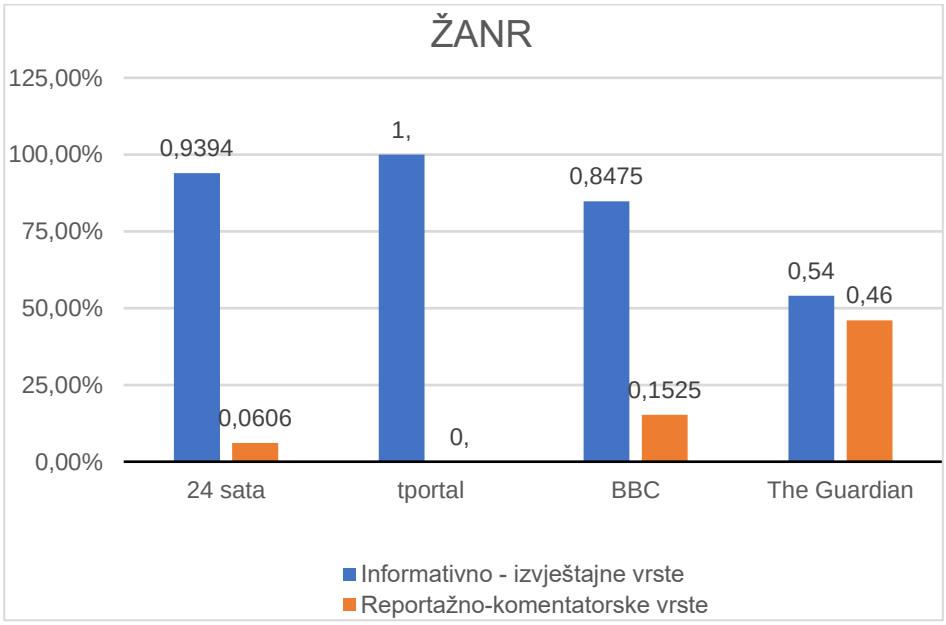


Grafikon 1. Broj objavljenih priloga (N=198)

U istraživanju je analizirano ukupno 198 novinskih priloga sa četiri internetska portala (Grafikon 1): dva hrvatska (24 sata i Tportal.hr) i dva inozemna (The Guardian i BBC). Najviše novinskih priloga je pronađeno na inozemnom portalu BBC (59 novinskih priloga; 29,80 %), dok je najmanje s hrvatskog portala 24 sata (33 novinska priloga; 16,57 %).

Jedan od razloga odabira navedenih portala je njihova pristupačnost tražilici ključnih riječi i mogućnost pregledavanja u većem vremenskom razdoblju. Odabir The Guardiana je bio temeljen na situaciji Edwarda Snowdena te činjenici da je bio prvi časopis koji je objavio dokumente u javnosti i čiji su novinari surađivali sa Snowdenom. Andrew Chadwick i Simon Collister (2014: 4) pišu: *„Tijekom 2013. i početkom 2014. godine, 192-godišnja britanska novina, jedan od velikih starih igrača britanskog novinarstva, objavio je ambiciozan i vrlo kontroverzan niz istraživačkih članaka o američkoj Nacionalnoj sigurnosnoj agenciji (NSA)… Riječ je o 192-godišnjim novinama Guardian i izvor tih članaka bilo je izvješće o desetinama tisuća povjerljivih dokumenata koje se procurile od bivšeg insajdera NSA-e imenom Edward Snowden".* Hrvatski portali bili su izabrani temeljem podataka tvrtke Gemius[14], koja ističe da je 24 sata najčitaniji portal u Hrvatskoj, a Tportal.hr zauzima treće mjesto kao najčitaniji portal u Hrvatskoj.

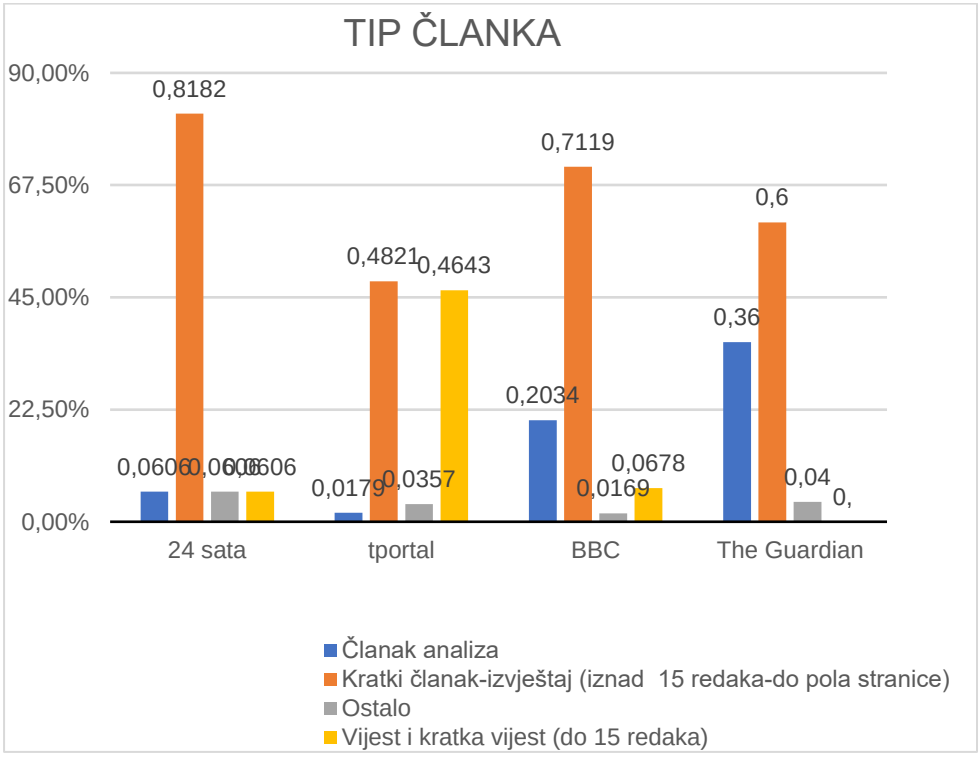
6.2. Analiza i interpretacija rezultata



[14] Gemius je tvrtka za poslovno savjetovanje, koja pruža podatke, rješenja i preporuke u području upravljanja internetskim oglašivačkim kampanjama.

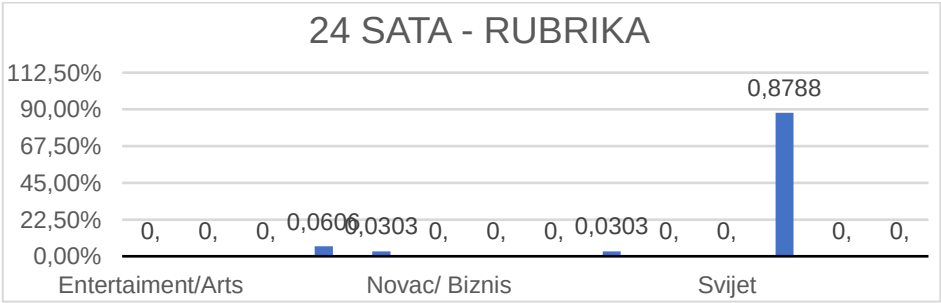
Grafikon 2. Vrste novinskih priloga (N=198)

Analizirani prilozi su većinom informativno-izvještajne vrste (Grafikon 2), pri čemu je veća zastupljenost utvrđena kod promatrana dva hrvatska portala (93,94 % - 24 sata i 100 % t-portal) u odnosu na inozemne portale (84,75 % BBC i 54,00 % The Guardian).



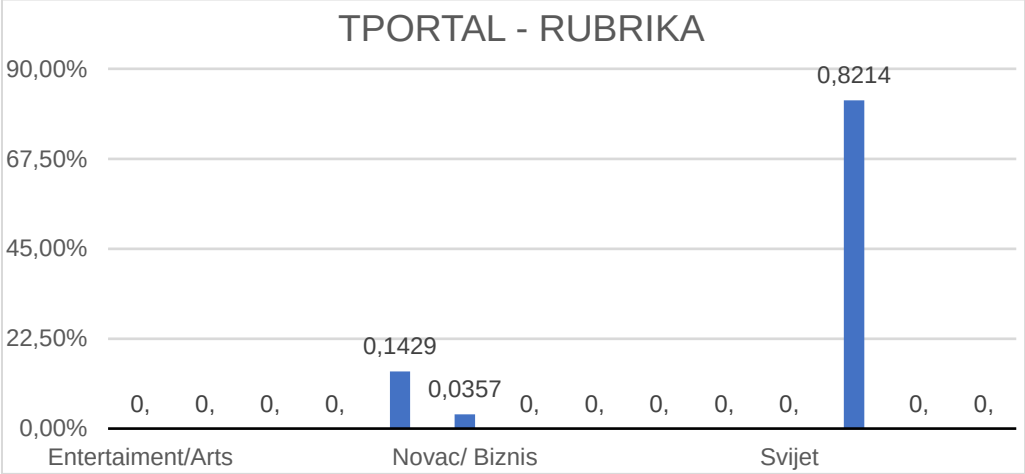
Grafikon 3. Tip članka (N=198)

Najveći broj priloga su bili kratki izvještaji, tj. prilozi iznad 15 redaka do pola stranice (Grafikon 3), pri čemu je najveća zastupljenost navedenog tipa članaka utvrđena kod 24 sata (81,82 %), dok je najmanja zastupljenost utvrđena kod Tportal.hr-a (48,21 %).



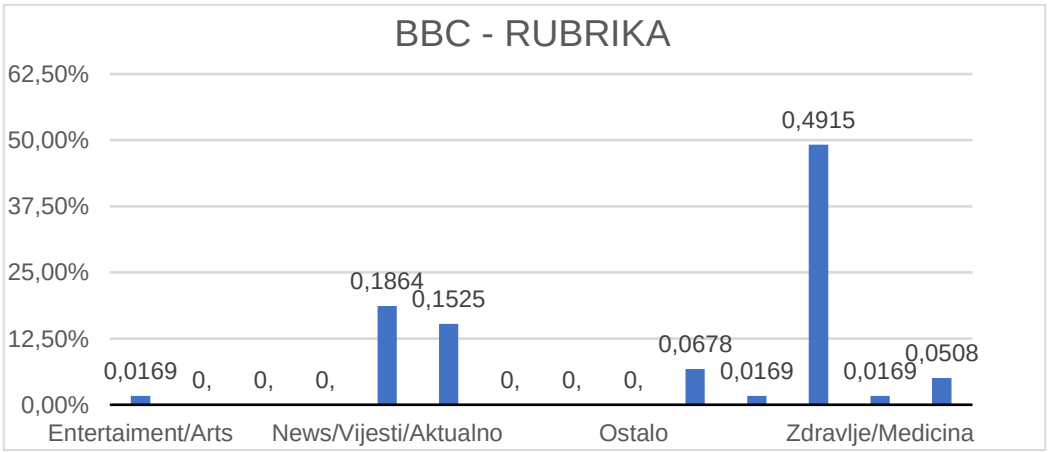
Grafikon 4. Tip rubrike na portalu 24 sata (N=33)

Najveći broj rubrika u kojima su objavljeni analizirani prilozi kod portala 24 sata je iz područja tehnologije (87,88 %), dok je 6,06 % članaka rubrike Life/style te 3,03 % članaka iz ostalih rubrika (Grafikon 4). U strukturi rubrika postoje 4 tipa rubrika.



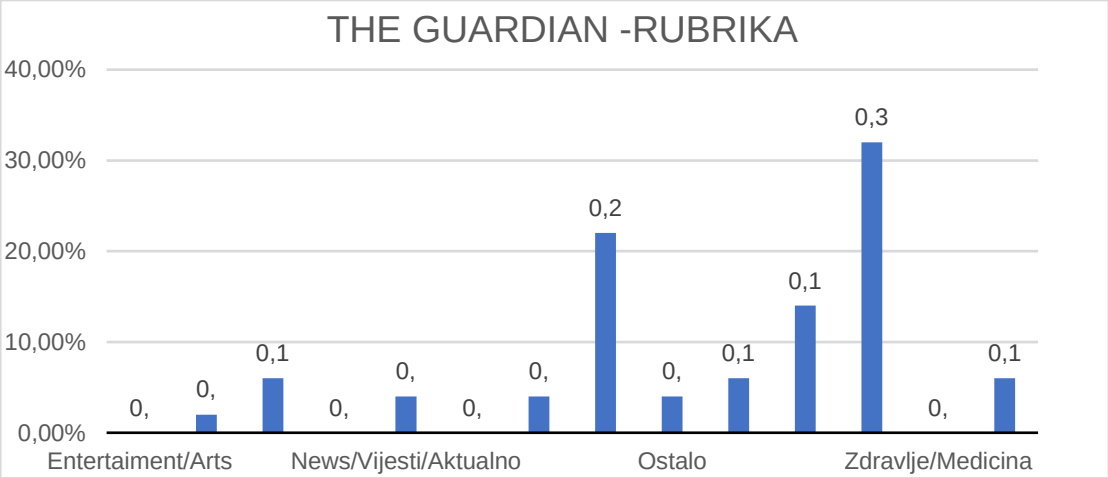
Grafikon 5. Tip rubrike na tportalu (N=56)

Najveći broj rubrika u kojima su objavljeni prilozi na portalu Tportal.hr je iz područja tehnologije (81,14 %), dok je 14,29 % priloga objavljeno u rubrici Novac/vijesti/aktualno, a 3,57 % članaka iz rubrike novac/biznis). U strukturi rubrika postoje 3 tipa rubrika (Grafikon 5).



Grafikon 6. Tip rubrike na portalu BBC-a (N=59)

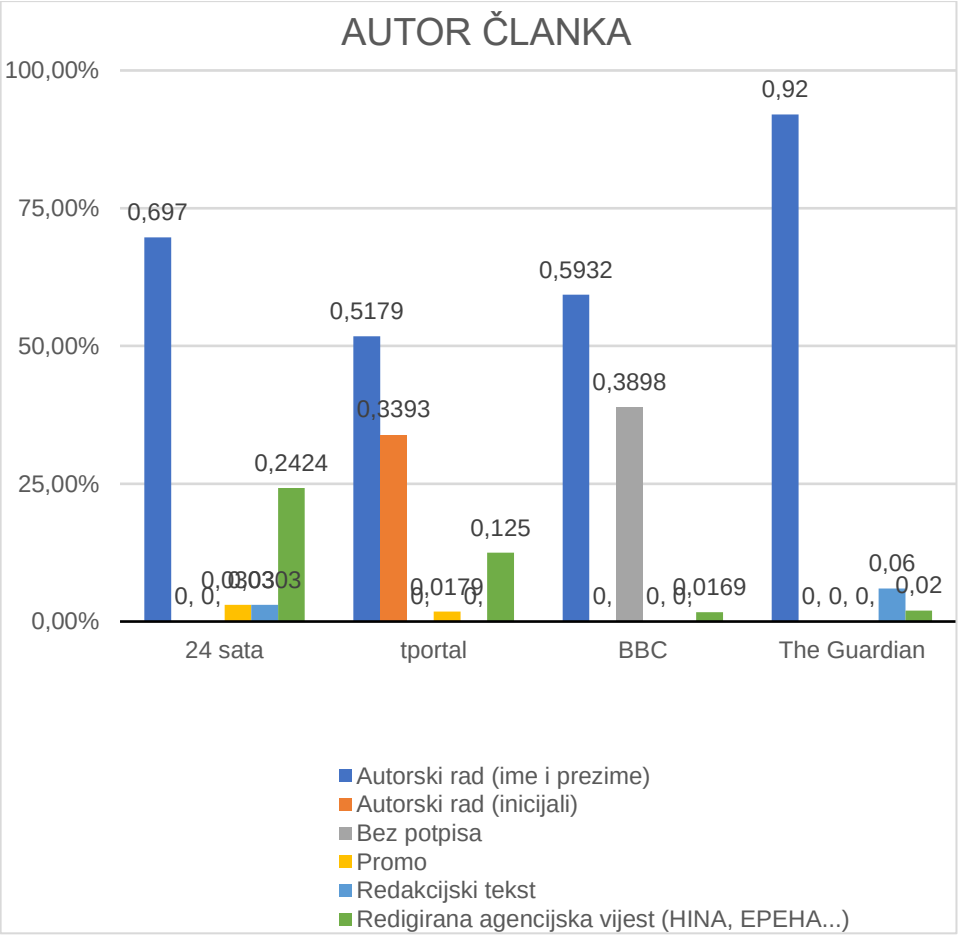
Najveći broj priloga na portalu BBC-a objavljen je u rubrici iz područja tehnologije (49,15 %), dok je zastupljeno 8 tipova rubrika (Grafikon 6).



Grafikon 7. Tip rubrike na portalu The Guardiana (N=50)

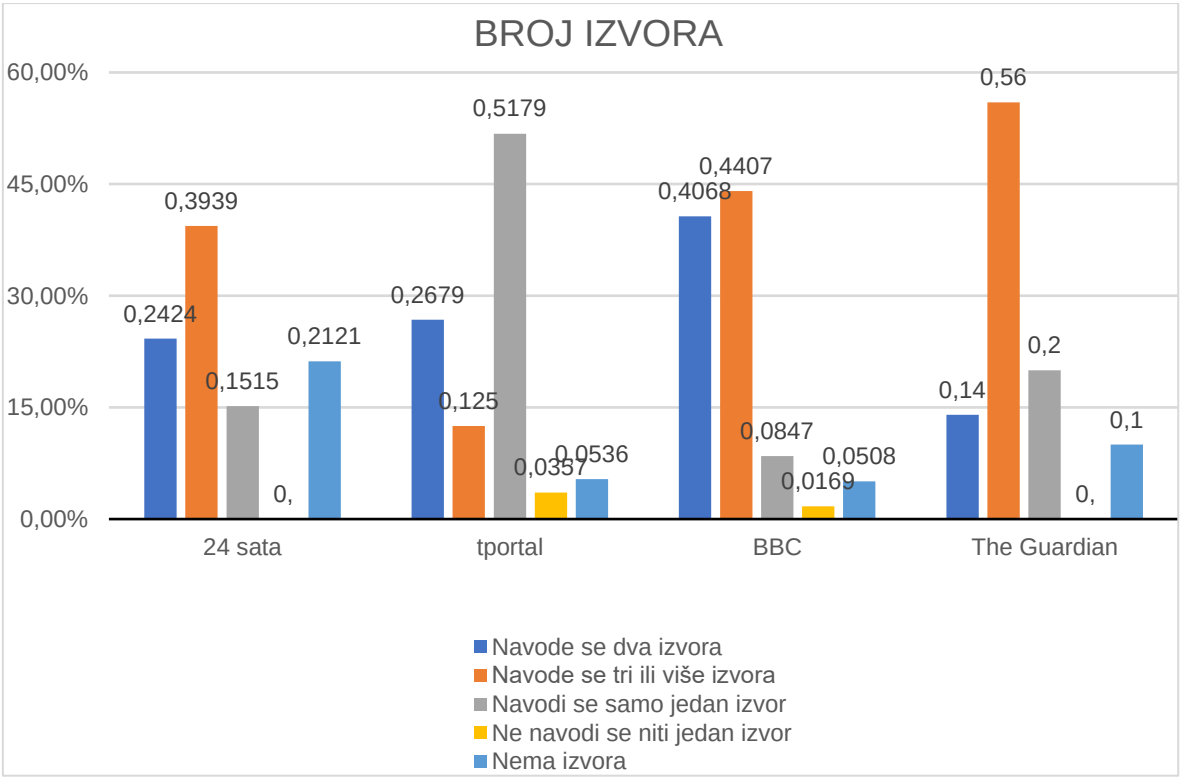
Najveći broj priloga na portalu The Guardiana objavljen je u rubrici iz područja tehnologije (32,00 %), dok je zastupljeno 10 tipova rubrika (Grafikon 7).

Dakle, zaključno možemo reći da hrvatski portali priloga uglavnom objavljuju u rubrici tehnologije, dok strani portali priloga objavljuju u nešto raznovrsnijim rubrikama.



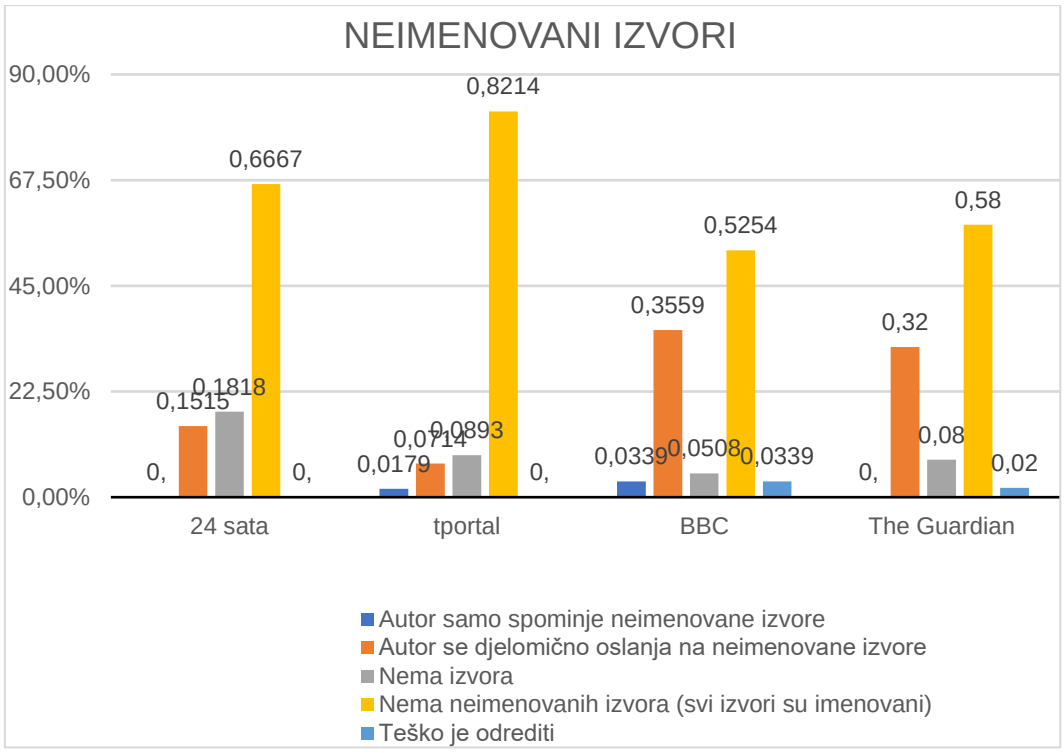
Grafikon 8. Autorstvo priloga

Više od 50 % priloga na sva četiri analizirana portala su autorski radovi s objavljenim punim imenom i prezimenom autora (Grafikon 8). Najveća zastupljenost navedenog tipa priloga je utvrđena kod The Guardiana (92,00 %), dok je najmanja zastupljenost istog tipa priloga utvrđena na hrvatskom portalu Tportal.hr (51,79 %). Za razliku od hrvatskih portala, kod inozemnih portala nije utvrđeno postojanje promotivnih priloga.



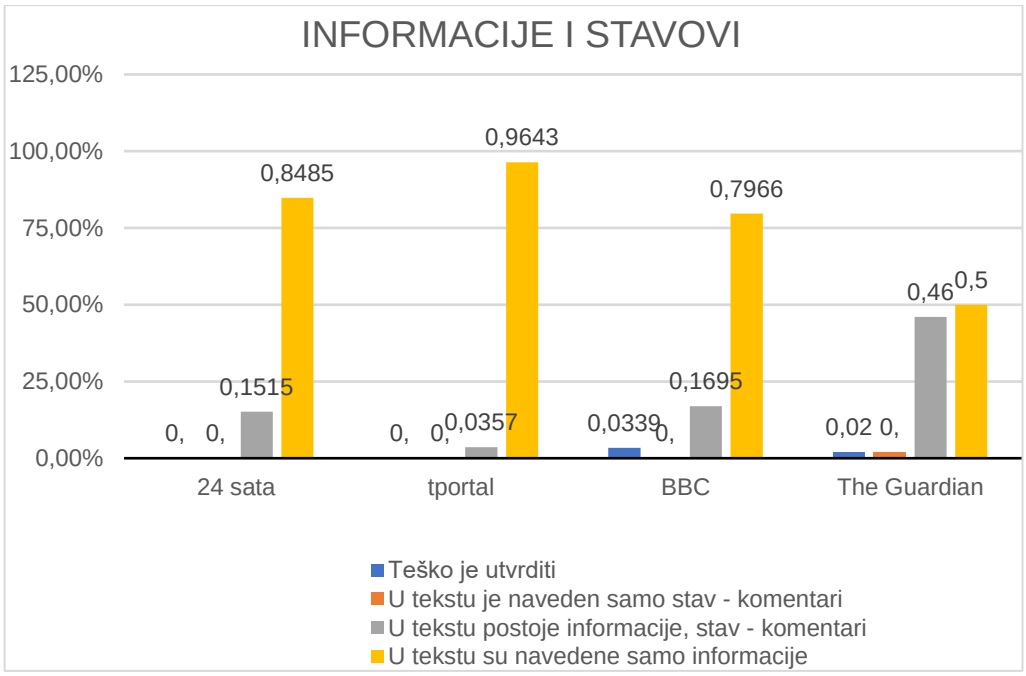
Grafikon 9. Izvori u prilogu

Navođenje 3 ili više izvora je u većoj mjeri zastupljeno kod inozemnih portala (44,07% - BBC i 56,00 % The Guardian), dok je navođenje 3 ili više izvora kod hrvatskih portala zastupljeno sa 39,39 % kod 24 sata, te 12,50 % kod Tportal.hr-a (Grafikon 9).



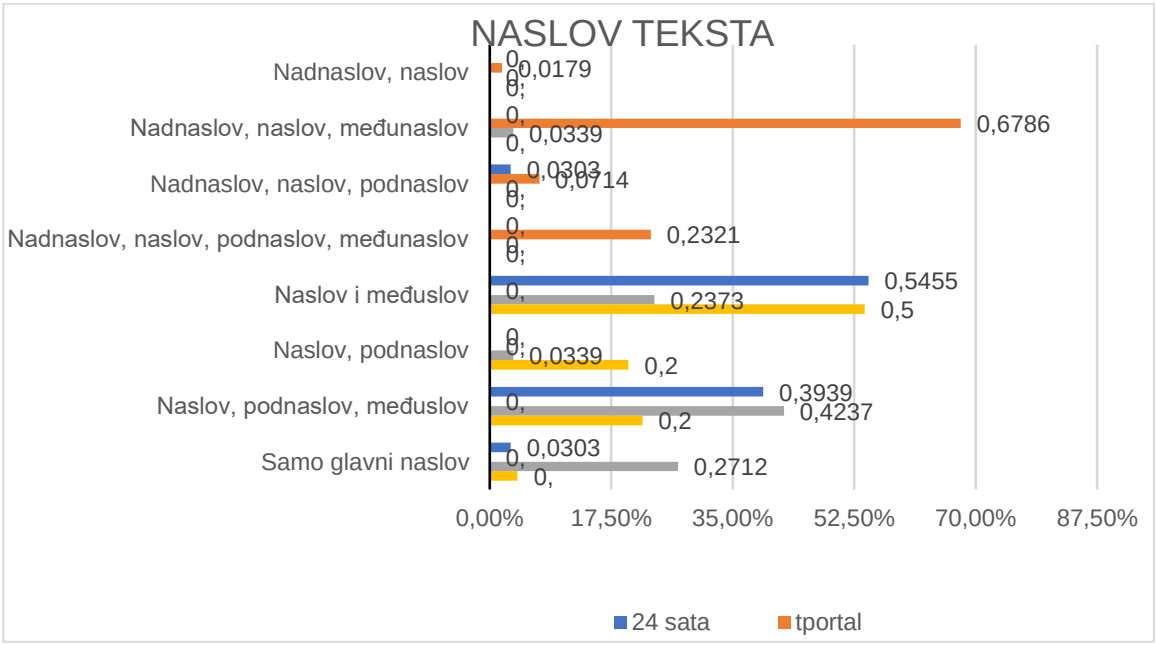
Grafikon 10. Neimenovani izvori u prilozima

U velikom broju priloga autori točno navode izvore informacija. Imenovanje izvora je utvrđeno kod 66,67 % priloga portala 24 sata i kod 82,14% priloga portala Tportal.hr. Također, utvrđena je veća zastupljenost imenovanih izvora u prilozima kod hrvatskih portala u odnosu na inozemne gdje je utvrđeno potpuno imenovanje izvora kod 52,54 % priloga BBC-a i 58,00 % priloga The Guardiana (Grafikon 10). Inozemni portali učestalije objavljuju priloge u kojima se autori djelomično oslanjaju na neimenovane izvore u odnosu na hrvatske portale.



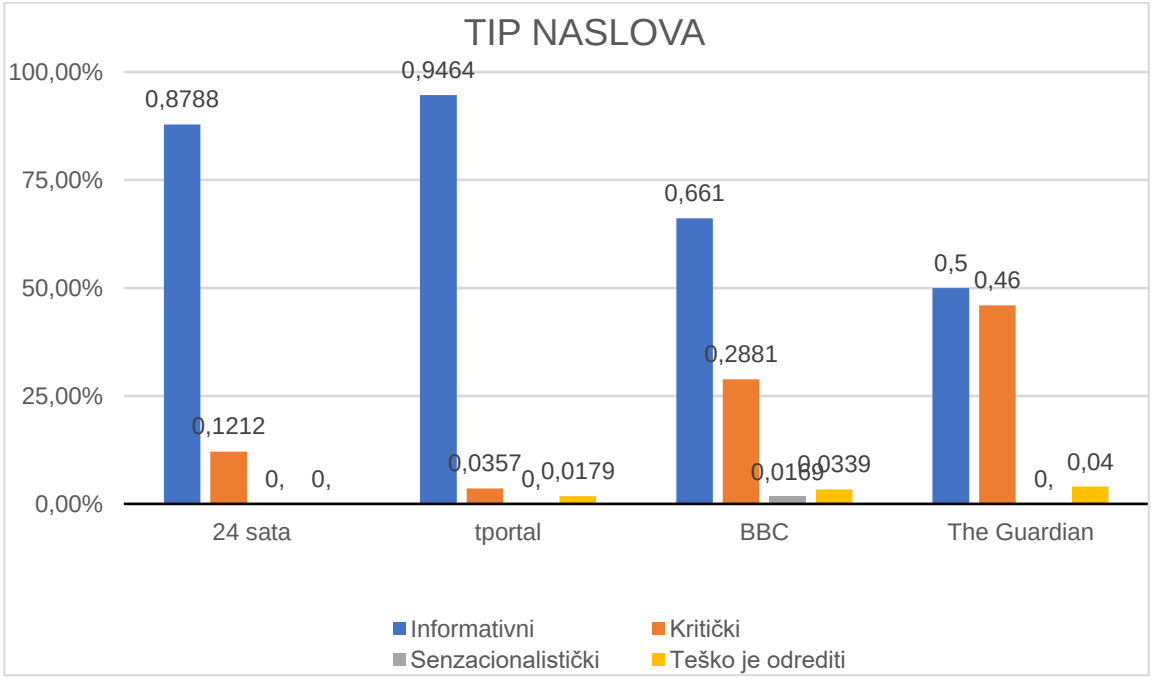
Grafikon 11. Informacije i stavovi novinara u prilozima

Kod sva četiri analizirana portala prevladavaju prilozi kod kojih su u tekstu navedene samo informacije te je utvrđeno da je veća zastupljenost navedenog tipa priloga kod hrvatskih portala (84,85 % kod portala 24 sata, te 96,43 % kod Tportal.hr-a) u odnosu na inozemne portale gdje je zastupljenost navedenog tipa priloga 79,66 % kod portala BBC te The Guardiana 50,00 % (Grafikon 11).



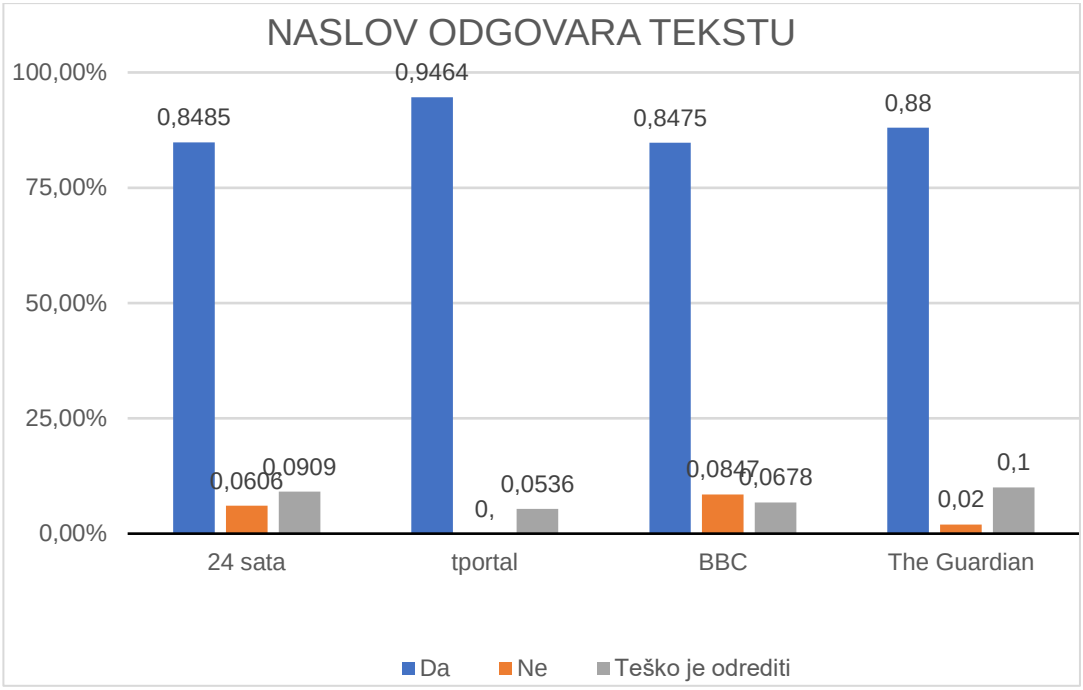
Grafikon 12. Naslovi kao oprema priloga

Kod Tportal.hr-a najčešće se navodi nadnaslov, naslov i podnaslov (67,86 % priloga), dok hrvatski portal 24 sata najčešće navodi naslov i međunaslov. Među inozemnim portalima najčešće se navodi naslov i međunaslov (portal The Guardian; 54,00 %) te naslov, podnaslov i međunaslov kod inozemnog portala BBC (42,37 % priloga) (Grafikon 12).



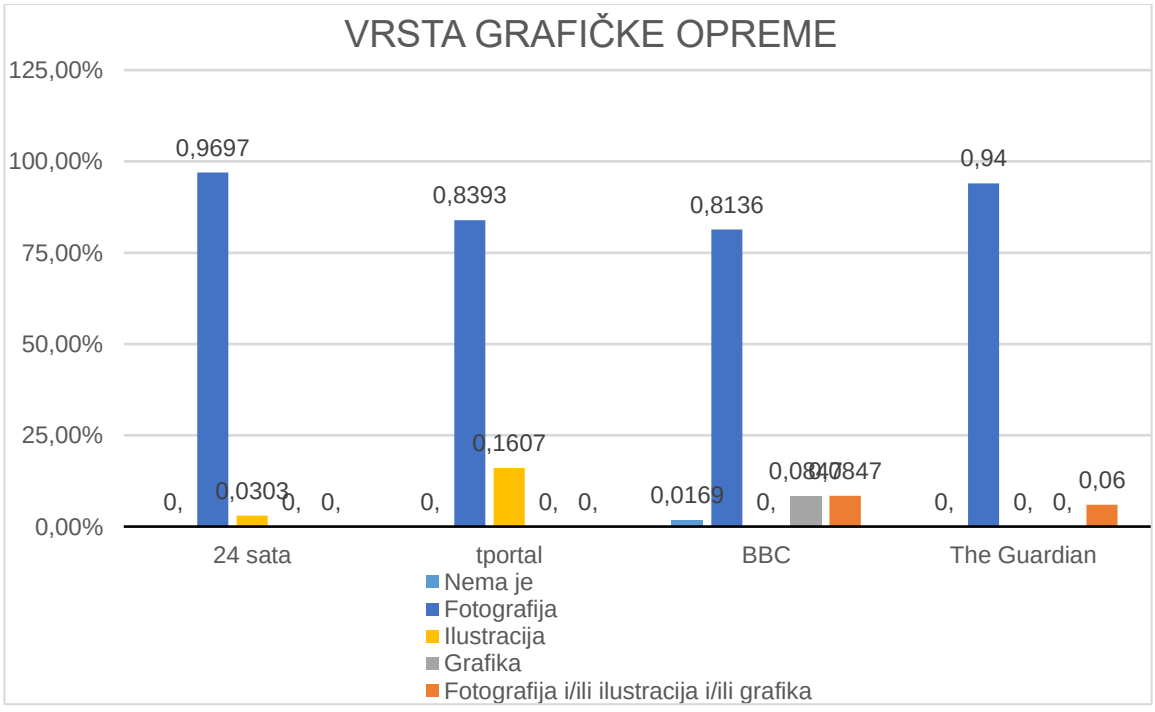
Grafikon 13. Tip naslova

Svi promatrani portali najčešće imaju informativan naslov priloga, a taj tip naslova češće je utvrđen kod hrvatskih portala (87,88 % kod portala 24 sata te 94,64 % kod portala Tportal.hr), dok su informativni naslovi utvrđeni kod inozemnih portala BBC (66,10 %) i The Guardian (50,00 %) rjeđi u odnosu na hrvatske portale (Grafikon 13).



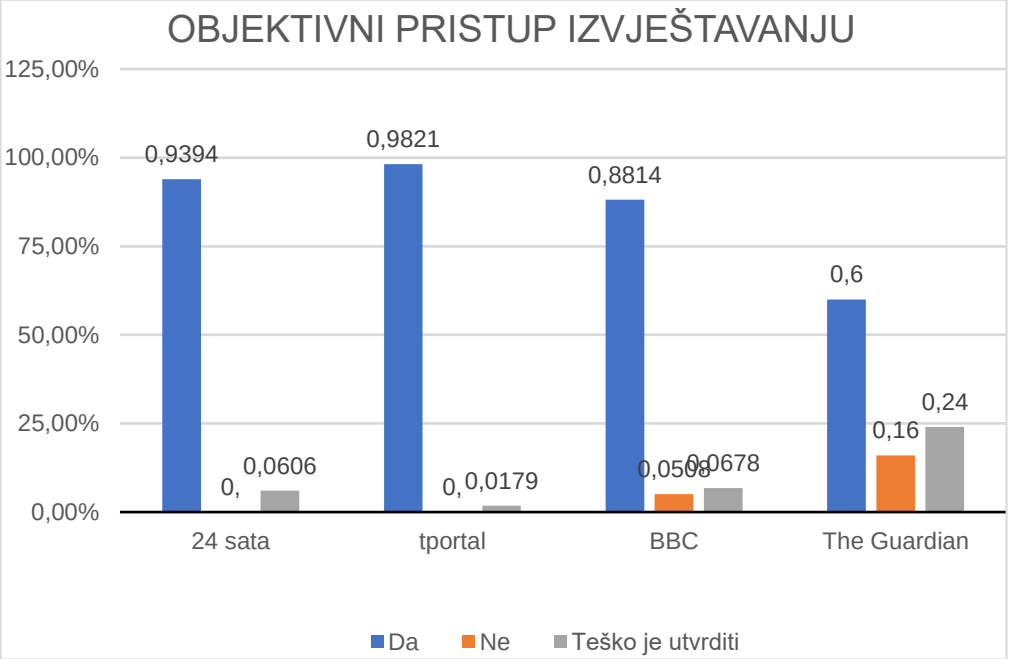
Grafikon 14. Odgovara li naslov tekstu

Kod većeg dijela priloga svih analiziranih portala naslov najčešće odgovara tekstu, pri čemu je kod hrvatskog portala 24 sata utvrđena usklađenost kod 84,85 % članaka, a kod portala Tportal.hr 94,64 %, dok je kod inozemnih portala utvrđena usklađenost 84,75 % kod portala BBC, te 88,00 % kod portala The Guardiana (Grafikon 14).



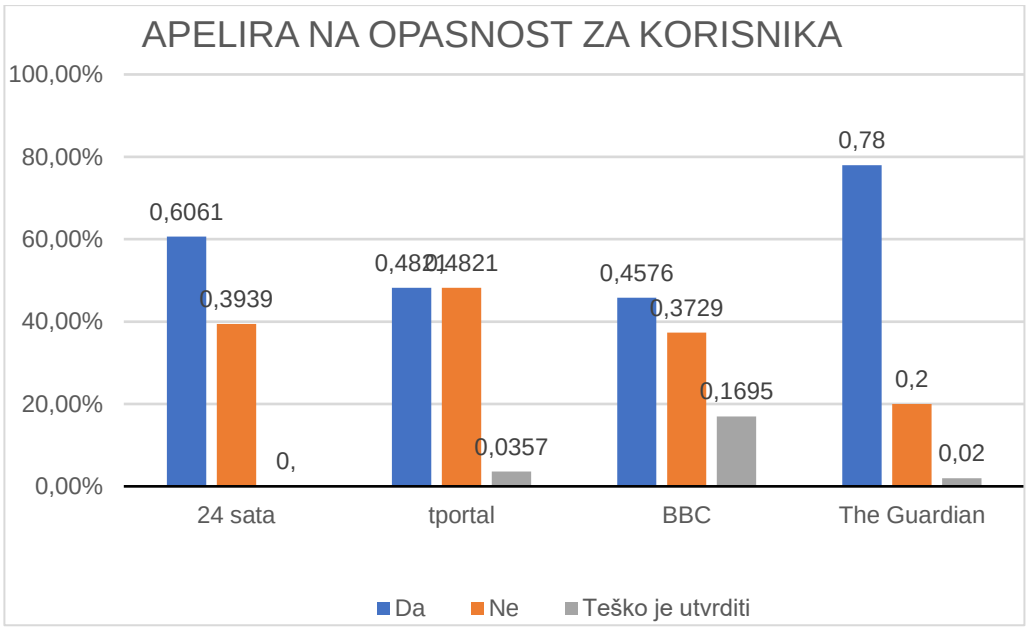
Grafikon 15. Vrsta grafičke opreme

Većina priloga na sva četiri analizirana portala donosi i fotografije. Najveća zastupljenost fotografije kao vrste grafičke opreme je utvrđena kod hrvatskog portala 24 sata (96,97 %), dok je najmanja zastupljenost istog tipa grafike opreme utvrđena kod BBC-a (81,36 %). Ilustracija kao vrsta grafičke opreme je zastupljena najviše kod hrvatskog portala Tportal.hr (16,07 %), dok kod inozemnih portala uopće nije zastupljena. Grafika se javlja u 8,47 % priloga s portala BBC, dok se kod ostala tri portala uopće ne javlja. Fotografija i/ili ilustracija i/ili grafika se javljaju kod inozemnih portala, dok kod hrvatskih portala nisu zastupljene (Grafikon 15).



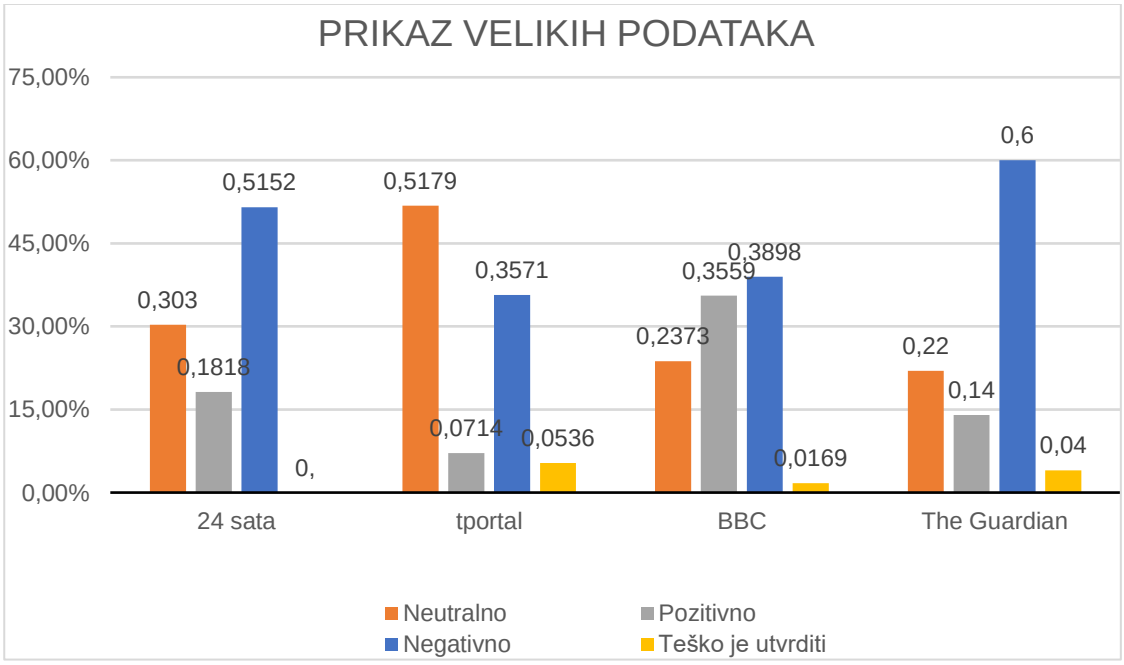
Grafikon 16. Objektivnost u izvještavanju

Objektivni pristup izvještavanju u većoj mjeri otkrili smo kod hrvatskih portala (24 sata - 93,94 %; Tportal.hr - 98,21 %), dok smo ga kod inozemnih portala pronašli rjeđe (BBC - 88,14 %; The Guardian - 60,00 %) (Grafikon 16). U 24 % priloga s portala The Guardiana je teško utvrditi primjenjuje li se objektivni pristup izvještavanju.



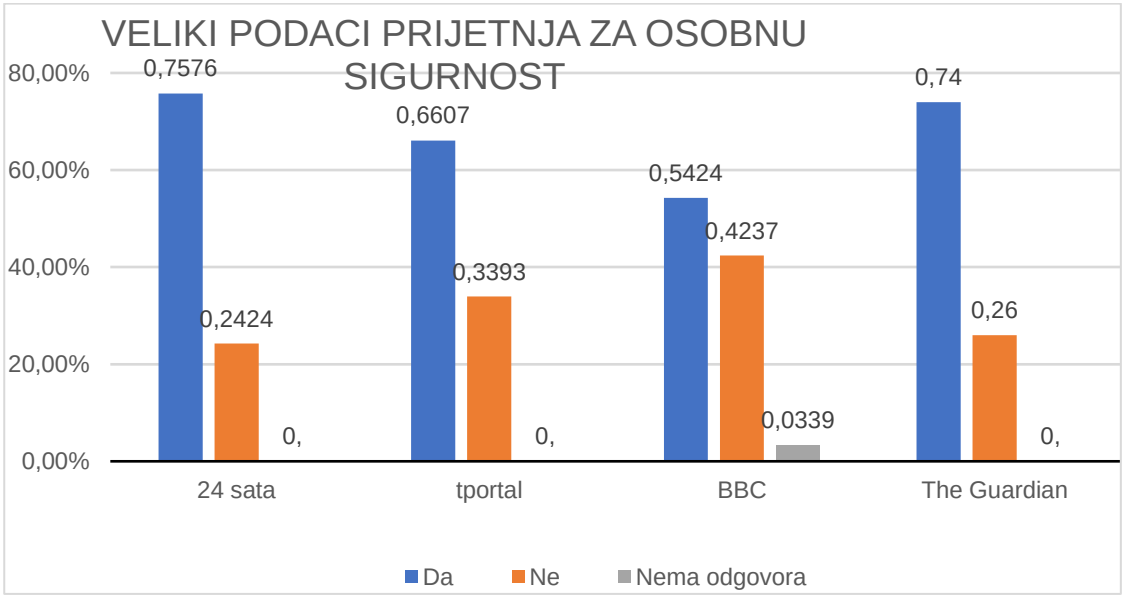
Grafikon 17. Apel na opasnost za korisnika

Prilozi s portala The Guardiana u najvećoj mjeri apeliraju na opasnost za korisnike (78,00 %), dok oni s BBC-a to čine u najmanjoj mjeri (45,76 %). U 16,95 % priloga s portala BBC-a je teško utvrditi apeliraju li na opasnost za korisnike (Grafikon 17).



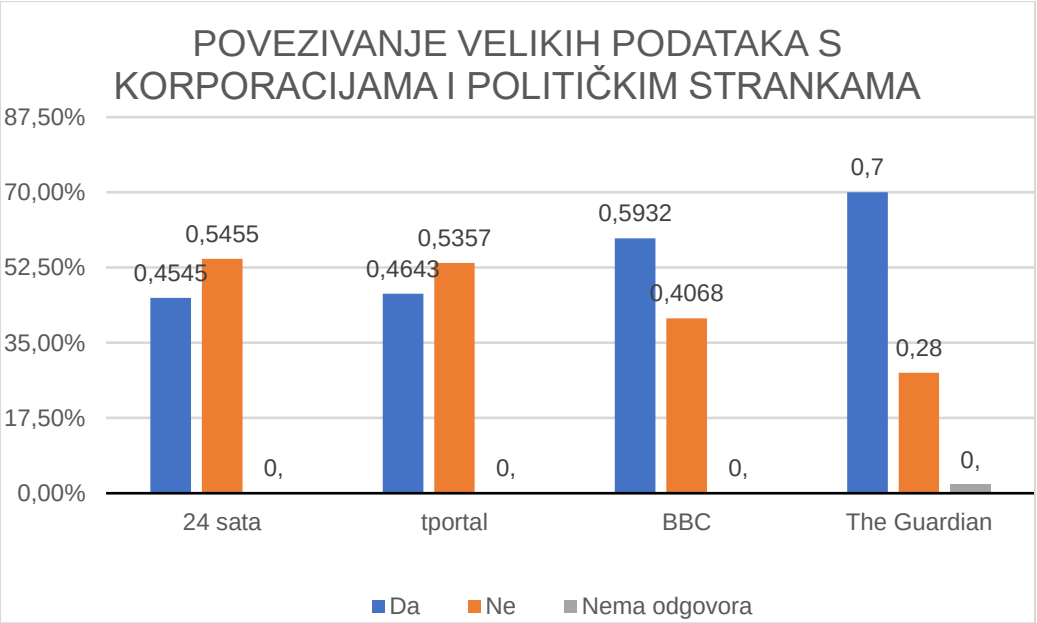
Grafikon 18. Prikazivanje velikih podataka u tekstu

Kod tri portala (24 sata – 51,52 %, BBC – 38,98 %, The Guardian – 60,00 %) su veliki podaci prikazani u najvećoj mjeri negativno, dok su kod Tportal.hr-a u najvećoj mjeri prikazani neutralno (51,79 %) (Grafikon 18).



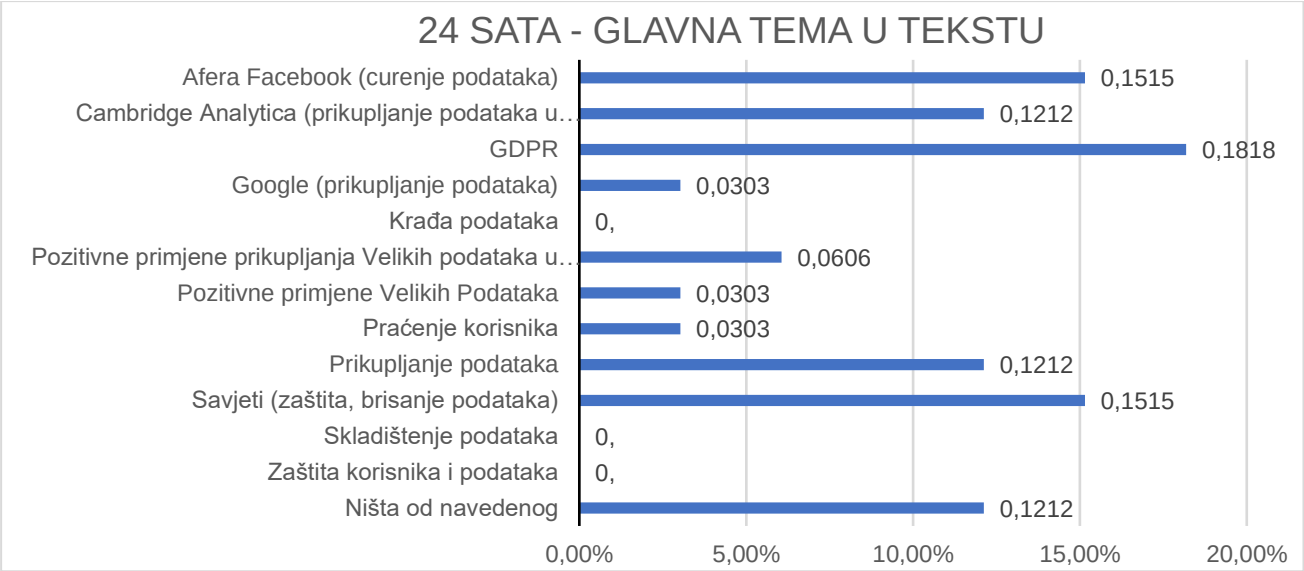
Grafikon 19. Veliki podaci kao moguća prijetnja za osobnu sigurnost

Svi analizirani portali najčešće predstavljaju velike podatke kao prijetnju za osobnu sigurnost, a takav je način predstavljanja u najvećoj mjeri utvrđen kod portala The Guardiana (74,00 %), dok je u najmanjoj mjeri utvrđen kod portala BBC (52,24 %) (Grafikon 19).



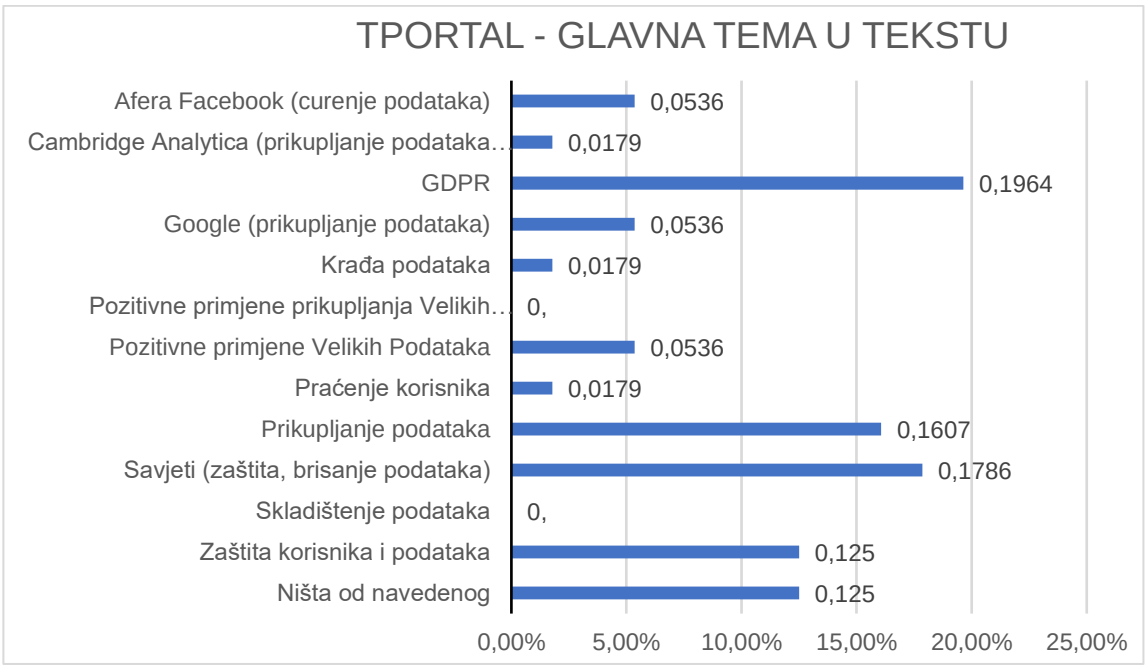
Grafikon 20. Povezivanje velikih podataka s korporacijama i političkim strankama

Povezivanje velikih podataka s korporacijama i političkim strankama koje koriste osobne podatke u komercijalne svrhe je u većoj mjeri utvrđeno u prilozima s inozemnih portala (Grafikon 20). Tako je povezivanje velikih podataka s korporacijama i političkim strankama koje koriste osobne podatke u komercijalne svrhe utvrđeno kod 59,32 % priloga portala BBC, te kod 70,00 % priloga portala The Guardiana. Veća zastupljenost povezivanja velikih podataka s korporacijama i političkim strankama utvrđena je kod inozemnih portala u odnosu na hrvatske. Povezivanje velikih podataka s korporacijama i političkim strankama koje koriste osobne podatke u komercijalne svrhe je utvrđeno kod 45,45 % priloga s portala 24 sata, te kod 46,43 % priloga s Tportal.hr.



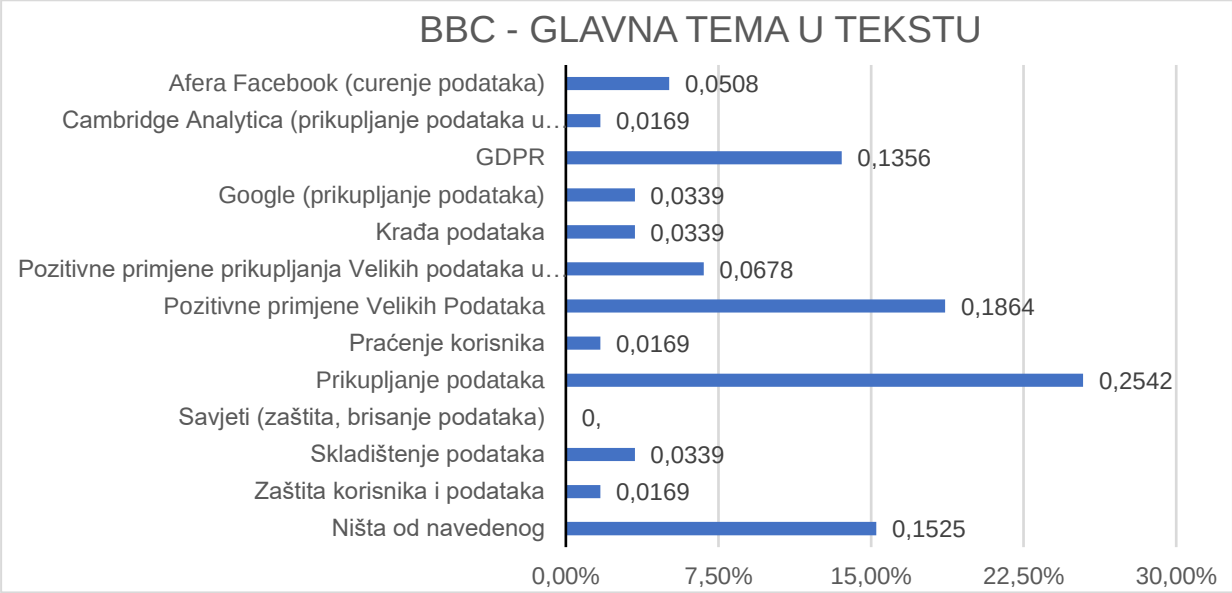
Grafikon 21. Zastupljenost tema u prilozima 24 sata

Najveći broj priloga s portala 24 sata odnosi se na temu GDPR-a (18,18 %), dok je jednak broj priloga (3,03 %) posvećen temi praćenja korisnika, pozitivne primjene velikih podataka te Googlea (prikupljanje podataka). Niti jedan članak s portala 24 sata nije bio na temu zaštite korisnika i podataka, skladištenja podataka i krađe podataka (Grafikon 21).



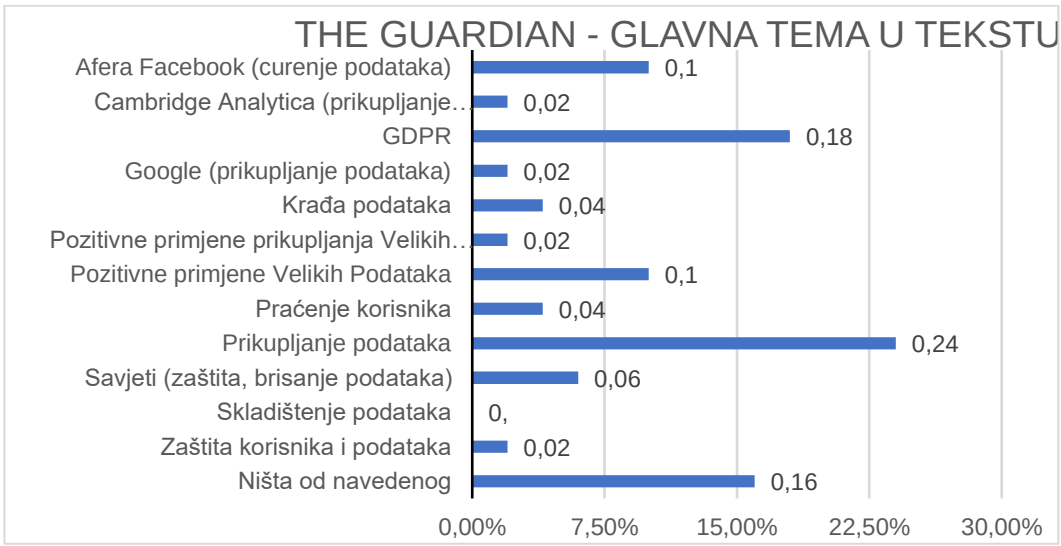
Grafikon 22. Zastupljenost tema u prilozima Tportala.hr

Najveći broj priloga s portala Tportal.hr odnosi se na temu GDPR-a, dok je jednak broj priloga (1,79 %) posvećen temi praćenja korisnika odnosno krađe podataka. Nijedan prilog s portala Tportal.hr nije obradio teme skladištenja podataka i pozitivne primjene prikupljanja velikih podataka u medicini (Grafikon 22).



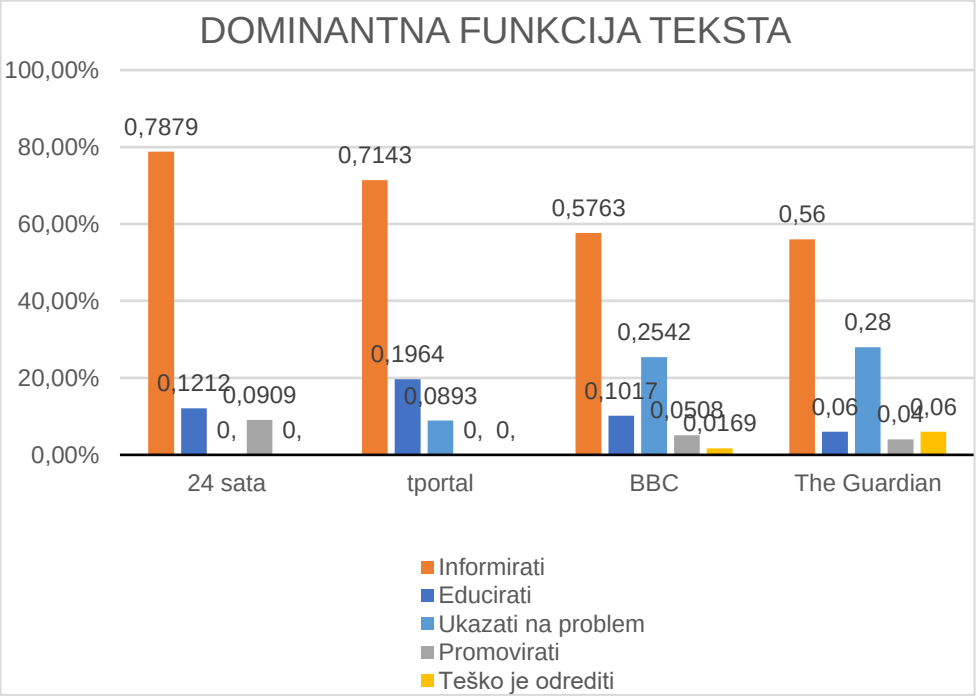
Grafikon 23. Zastupljenost tema u prilozima portala BBC-a

Najveći broj priloga s portala BBC posvećen je temi prikupljanja podataka (25,42 %) dok je jednak broj priloga (1,69 %) obradio temu zaštite korisnika i podataka, praćenja korisnika te Cambridge Analytica (prikupljanje podataka u političke svrhe, izbore). Niti jedan članak s portala BBC nije bio na temu savjeti (zaštita, brisanje podataka) (Grafikon 23).



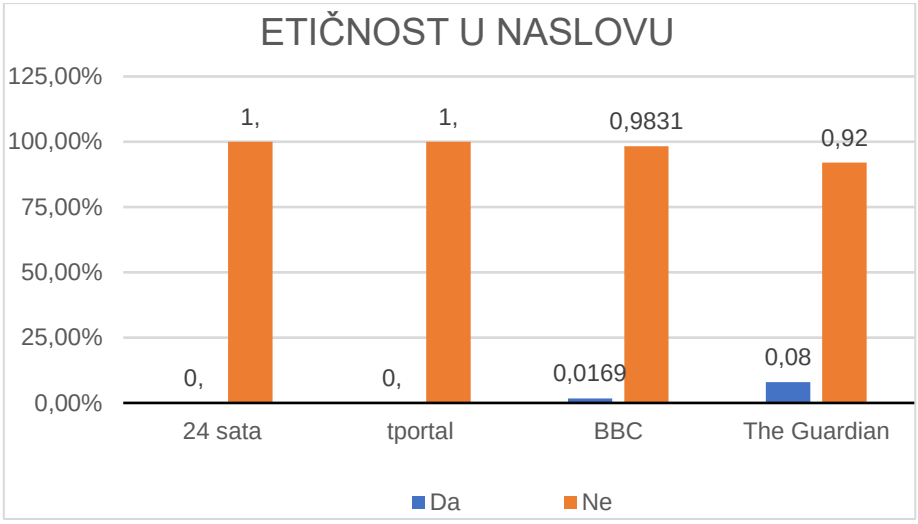
Grafikon 24. Zastupljenost tema u prilozima portala The Guardiana

Najveći broj priloga s portala The Guardian je na temu prikupljanja podataka (24,00 %), dok je jednak broj priloga (2,00 %) na temu zaštite korisnika i podataka, pozitivne primjene prikupljanja velikih podataka u medicini, Google (prikupljanje podataka) te Cambridge Analytica (prikupljanje podataka u političke svrhe, izbore). Niti jedan članak s portala The Guardian nije bio na temu skladištenja podataka (Grafikon 24).



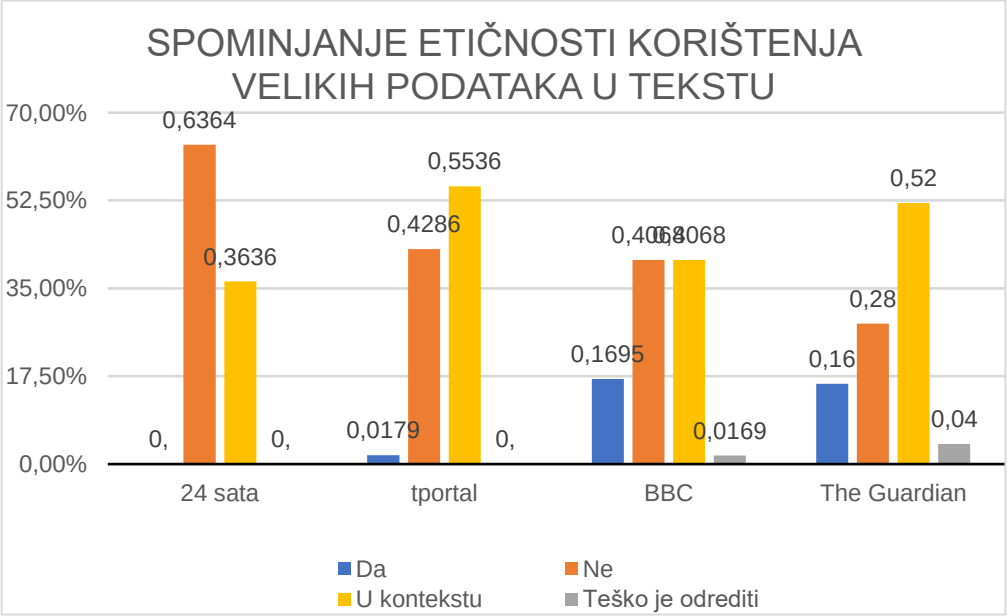
Grafikon 25. Funkcija teksta

Kod većeg dijela priloga svih analiziranih portala najčešća dominantna funkcija teksta je informirati, pri čemu je ta funkcija na hrvatskom portalu 24 sata utvrđena kod 78,79 % priloga, a kod portala Tportal u 71,43 %, dok je kod inozemnih portala ta funkcija utvrđena u 57,63 % priloga portala BBC te 56,00 % portala The Guardian (Grafikon 26). Kod inozemnih portala značajan je bio i broj članaka s funkcijom ukazivanja na problem (BBC – 24,42 %, The Guardian – 28,00 %), u odnosu na hrvatske portale kod kojih se navedena funkcija ne javlja.



Grafikon 26. Etičnost u naslovu

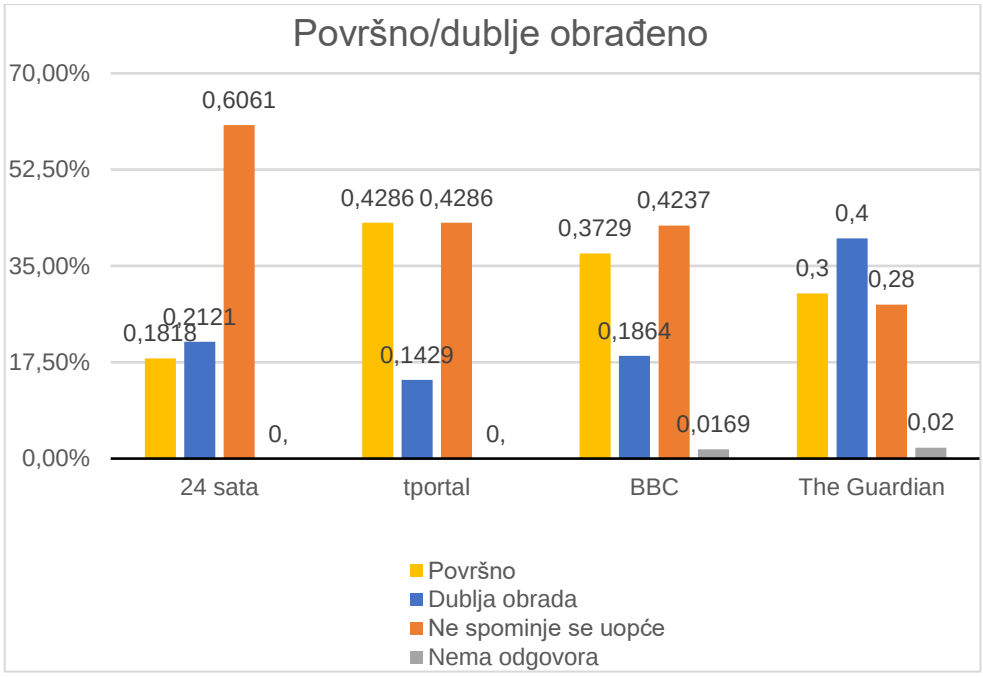
Etičnost u naslovu u većoj smo mjeri otkrili kod inozemnih portala, pri čemu je kod 9,69 % priloga s oba portala utvrđeno spominjanje etičnosti u naslovu (Grafikon 26). Kod portala 24 sata i Tportal.hr etičnost ili općenito neka naznaka o potrebi etičnoga korištenja velikih podataka u naslovu se uopće ne spominje.



Grafikon 27. Etičnost korištenja velikih podataka u tekstu

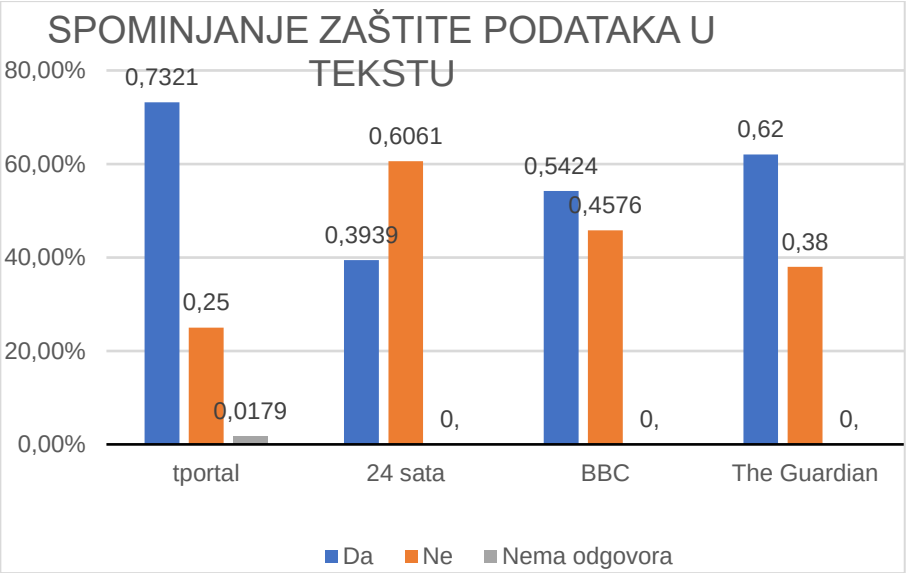
U 63,64 % priloga s portala 24 sata se ne spominje pitanje ili problematika etičnosti korištenja velikih podataka u tekstu, dok se u 55,36 % priloga s portala Tportal.hr spominje etičnost korištenja velikih podataka u kontekstu. U jednakom broju priloga (40,68%) s portala

BBC se ne spominje. odnosno u kontekstu se spominje etičnost korištenja velikih podataka u tekstu. U više od 50% priloga s portala The Guardian se u kontekstu spominje etičnost korištenja velikih podataka u tekstu (Grafikon 27).



Grafikon 28. Obrada teme

Na portalu Tportal.hr u jednakom broju priloga (42,86 %) površno se, odnosno nikako ne spominje etičnost korištenja velikih podataka u tekstu. Kod inozemnih portala su rezultati različiti. Tako se u 42,37 % priloga s portala BBC etičnost korištenja velikih podataka ne spominje u tekstu, a u 40,00 % priloga s portala The Guardian se etičnost korištenja velikih podataka u tekstu dublje obrađuje (Grafikon 28).



Grafikon 29. Spominjanje zaštite podataka u tekstu

U većem broju priloga portala Tportal.hr (73,21 %), BBC (54,24 %) odnosno The Guardian (62,00 %) se zaštita podataka spominje u tekstu, dok se kod većeg broja (60,61 %) članaka s portala 24 sata zaštita podataka u tekstu ne spominje.

6.2. Diskusija

Ovim istraživanjem se namjeravalo utvrditi kako su pojedini hrvatski i inozemni portali izvještavali o velikim podacima i etičnosti u njihovim korištenju. Sama pak tema o velikim podacima je kompleksna i komplicirana, jer još uvijek ne postoji suglasje oko tih pitanja, pa tako postoji i mogućnost da svaka osoba ima svoje mišljenje, a zemlje različite zakone vezane uz ove velike podatke kao informacije.

Naša prva hipoteza se odnosila na to da tekstovi na internetskim portalima u Hrvatskoj u manjoj mjeri pružaju informacije o zaštiti i savjetima o upotrebi velikih podataka. Kod 24 sata, uzorak koji je ušao u istraživanje je bio gotovo upola manji nego kod ostalih istraživanih portala, jer smo mogli izdvojiti i analizirati samo 33 priloga iz razdoblja koje smo unaprijed utvrdili za analizu. Istraživanje je pokazalo da nijedan prilog od njih 33 kao glavnu temu nije imao pružanje prijedloga i savjeta kako da se korisnici zaštite ili načine na koje mogu obrisati (uvjetno rečeno) svoje podatke kako ne bi bili dostupni drugima. Tema koja je najviše bila

zastupljena na portalu 24 sata bila je vezana uz GDPR. Tportal.hr je imao više priloga od 24 sata, pa je tako analizom obuhvaćeno 56 priloga. Istraživanje je pokazalo da njih 17% govori o savjetima i zaštiti, dok je najzastupljenija tema i na tom portalu bila vezana uz GDPR.

Istraživanje je pokazalo da ni portal BBC-a, kao ni onaj 24 sata, u navedenom razdoblju nije imao nijedan prilog koji je upućivao na zaštitu korisnika i savjete, dok je The Guardian imao samo 6 % priloga u kojima je bila zastupljena tema vezana uz savjete, što je opet znatno manje nego kod Tportal.hr-a. Tportal.hr je svojim čitateljima nudio članke kao što su: *Ne želite da Apple, Facebook i Google prate vaš svaki korak? Evo što morate poduzeti*. U tekstu su ponuđeni jednostavni savjeti i preporuke kako smanjiti razinu praćenja koje velike tehnološke tvrtke obavljaju putem pametnih telefona i srodnih uređaja. Zbog toga na temelju dobivenih rezultata zaključujemo da smo istraživanjem opovrgnuli prvu hipotezu (H1) da hrvatski portali u manjoj mjeri pružaju informacije o savjetima i zaštiti u odnosu na strane portale.

Druga hipoteza kojom smo pretpostavili da tekstovi na internetskim portalima u Hrvatskoj u manjoj mjeri izvještavaju o upotrebi osobnih podataka u komercijalne svrhe, u većoj je mjeri utvrđena u prilozima s inozemnih portala. Povezivanje velikih podataka s korporacijama i političkim strankama koje koriste osobne podatke u komercijalne svrhe utvrđeno je kod 59,32 % članaka portala BBC-a i kod 70,00 % članaka portala The Guardian. Ovaj rezultat je potvrdio našu hipotezu, da internetski portali u Hrvatskoj u manjoj mjeri izvještavaju o upotrebi podataka u komercijalne svrhe. Portal 24 sata imao je 45.45 % priloga o upotrebi podataka u komercijalne svrhe, dok je kod Tportal.hr-a zabilježeno 46.43 % priloga koji govore o ovakvoj temi. Kod The Guardiana je mnogo priloga bilo na temu Facebooka, kao što je sumnja da je podatke 87 milijuna korisnika Facebooka, od kojih je 2,7 milijuna Europljana, nezakonito pribavila i koristila tvrtka Cambridge Analytic.

Wolfie Christl u svom članku *How Companies Use Personal Data Against People* (2017: 5) navodi da je ne tako davno, razmjer i dubina osobnih podataka u rukama trgovačkih subjekata bila prilično ograničena i prilično jednostavna za nadzor. Kreditni biroi, tvrtke za izravni marketing i tvrtke koje prodaju proizvode i usluge potrošačima počele su prikupljati, upravljati i razmjenjivati podatke o ljudima prije nekoliko desetljeća. Prijašnje baze podataka o potrošačima bile su izolirane, sporo ažurirane i obuhvaćale su samo djelić života tipične „prosječne“ osobe. U zadnjih nekoliko godina stanje se dramatično promijenilo. Od društvenih mreža, pametnih telefona i online oglašavanja, širok raspon tvrtki počeo je pratiti ljude u gotovo svim aspektima njihovih života. Ponašanja, kretanja, društveni odnosi, interesi,

slabosti i većina privatnih trenutaka milijarda ljudi stalno se bilježe, procjenjuju i analiziraju u realnom vremenu. Autorica nadalje navodi da velika i jako neobuzdana komercijalna upotreba osobnih podataka izaziva velike probleme i zabrinutost o budućnosti autonomije, jednakosti, ljudskog dostojanstva i demokracije svakog pojedinca.

Treća hipoteza se odnosi na to da internetski portali u većoj mjeri predstavljaju podatke kao prijetnju za osobnu sigurnost. Svi analizirani portali najčešće predstavljaju velike podatke kao prijetnju za osobnu sigurnost, a takav način predstavljanja u najvećoj mjeri utvrđen je kod portala The Guardian (74,00 %), dok je u najmanjoj mjeri utvrđen kod portala BBC-a (52,24 %). Hrvatski portali pokazali su veći stupanj prikazivanja velikih podataka kao prijetnje za sigurnost od BBC-a i to 24 sata u 75.76 %, dok je kod Tportal.hr-a zabilježeno 66.07 % priloga koji prikazuju velike podatke kao prijetnju za osobnu sigurnost. S ovim rezultatima možemo reći da smo potvrdili treću hipotezu, a činjenica je da mediji korisnicima daju do znanja da njihovi podaci nisu uvijek sigurni. Osiguranje velike sigurnosti podataka dolazi s njegovim brigama i izazovima, zbog čega je više nego korisno upoznati se s njima. Tijekom istraživanja susretali smo se s prilozima koji su upućivali na: rudarenje podataka, hakiranje podataka i profila na društvenim mrežama, istraživanje nad korisnicima, a da toga nisu ni svjesni. Svaki korisnik interneta trebao bi biti svjestan sigurnosnih prijetnji zbog velikih podataka, a trebao bi imati i svijest o tome koristiti kao primarnu zaštitu. Jedan od primjera priloga The Guardiana koji želimo izdvojiti naslovljen je *Are you ready? Here is all the data Facebook and Google have on you.* Tekst govori o tome da „berba“ naših osobnih podataka daleko nadilazi ono što mnogi od nas mogu zamisliti. Prilog pokazuje koliko informacija se o nama čuvaju Facebook i Google, a da toga uopće nismo svjesni. U njemu se također napominje da Google zna koje aplikacije svatko od nas koristi, da posjeduje svu našu YouTube povijest, da imaju oglašivački profil o svim korisnicima i da su podaci kojima raspolaže Google dovoljni za „popunjavanje“ milijuna Wordovih dokumenata.

Zadnja se hipoteza odnosi na to da mediji u Hrvatskoj u manjoj mjeri od inozemnih izvještavaju o pozitivnim stvarima upotrebe velikih podataka. Ta hipoteza je potvrđena, jer inozemni portali BBC i The Guardian u većoj mjeri prikazuju pozitivnu stranu korištenja podataka. Već smo u radu pisali o benefitima, to jest prednostima i koristima u korištenju velikih podataka za javno dobro. Kod The Guardiana i BBC-a veći broj priloga govori o korištenju velikih podataka u medicinske svrhe. Također, jedno od pitanja u matrici se odnosilo na prevladavanje teme u tekstu, a jedna od mogućnosti je bila i pozitivna primjena velikih podataka. Kod 24 sata takva tema je bila prisutna samo u 3.03 %, a kod Tportal.hr-a 5.36 % .

Kod pitanja autorstva priloga – a ono je i pokazatelj vjerodostojnosti napisanoga teksta i odgovornosti uredništva – uočeno je da ni BBC ni The Guardian nisu imali ni jedan promo prilog, dok su oba hrvatska portala u tome ipak različita. Također, u većoj je mjeri zabilježeno da su objavljivani prilozi čiji su autori agencije. Kod sva četiri portala prevladavao je informativni naslov, a zatim kritički kojeg je najviše imao The Guardian s 46 %.

Konačno, kao što je već navedeno, u ovom istraživanju korištena je kvantitativna metoda analize sadržaja, a analiziran je sadržaj 4 internetska portala, dva hrvatska i dva inozemna. Budući da smo pretraživali portale putem tražilica i ključnih riječi, postoji mogućnost da neki prilozi koji su mogli biti relevantni nisu uključeni u istraživanje, pri čemu se pogrješka može odnositi na same pretraživače i kvalitetu web stranica. Našim je radom ipak otvorena mogućnost za daljnja dublja i opsežnija istraživanja ove teme koja je relevantna kako za sadašnjost tako i za budućnost, osobito s komunikološkoga aspekta posvećenoga pravu svakoga pojedinca na privatnost i etici u komuniciranju.

7. Zaključak

U tijeku pisanja rada, pretraživanja literature i samih portala, susreli smo se sa brojnim istraživanjima i znanstvenim radovima koji se općenito odnose na velike podatke, ali i na njihovu primjenu, a sve se češće postavlja i pitanje je li ona etična i kako se koristi velike podatke. Ono što je odmah bilo uočljivo činjenica je da je većina znanstvenika i stručnjaka suglasna kako podaci sami po sebi nisu štetni, a neki ih nazivaju i neutralnima. Ono što je bitno je njihova upotreba i način na koji će korisnicima biti predstavljeni, odnosno jesu li korisnici upoznati s činjenicom da netko ima kontrolu nad njihovim podacima koje ostavljaju na internetu. Veliki podaci povećavaju sigurnosne izazove te usklađenost i upravljanje koji se primjenjuju na normalne podatke. Prilikom prikupljanja i obrade podataka, organizacije moraju poštivati propise i zadane norme. Iako se zakoni o zaštiti podataka širom svijeta razlikuju u određenim aspektima, svi oni imaju ista osnovna načela. Riječ je o čuvanju osobnih podataka, privatnosti podataka i kontroli upotrebe podataka. Korisnici bi trebali razumjeti koji se podaci prikupljaju, a obradu samih podataka morali bi odobriti upravo korisnici jer su oni, istini za volju, i vlasnici i autori tih podataka. No, isto tako treba upozoriti

na činjenicu da zbog velike količine podataka koje organizacije moraju obraditi, zaštita i upravljanje podacima postaju sve složeniji.

Nadalje, u pisanju i istraživanju za ovaj naš rad susreli smo se s brojnim djelima i znanstvenim člancima i različitim pozitivnim i negativnim primjerima u korištenju podataka, a najviše zanimanja je izazvao dio koji se odnosi na podatke koji se koriste za ono što se naziva „veće dobro“. Budući da se ovaj rad odnosio na etičnost u korištenju velikih podataka, zanimljivo je upitati se - je li etično braniti korištenje velikih podataka u zdravstvu, ako ti podaci mogu predvidjeti neke buduće negativne situacije? No, na to pitanje naš rad nije dao odgovor i ono može biti predmet budućih istraživanja, osobito s etičkoga aspekta koji je neizostavni dio svakodnevne ljudske komunikacije putem novih informacijskih i komunikacijskih tehnologija.

Istraživanje nam je pokazalo da mediji u Hrvatskoj više daju savjete o zaštiti i raznim mogućnostima uz pomoć kojih korisnici mogu obrisati svoje podatke. Osim toga, sva četiri portala obuhvaćena istraživanjem su u zadnje tri godine pisala kako su veliki podaci prijetnja za osobnu sigurnost te su isto tako korisnike educirali o štetnosti, ali i ukazivali na otvorene probleme vezane uz ta pitanja. Što se tiče pozitivnog prikaza velikih podataka, ono je najviše zabilježeno kod inozemnih medija.

Na kraju želimo istaknuti da svi veliki podaci nisu loši, kao što po sebi nije loše ni njihovo prikupljanje, ali se mogu koristiti za pogrešne, neetične pa čak i zle namjene. Osobno, optimistično gledamo na korištenje velikih podataka i na ono što oni mogu učiniti za društvo, ali također smo realni u pogledu ozbiljnih pitanja povezanih s kršenjem etičkih načela i prava na privatnost koje oni predstavljaju. Kao društvo, na nama je da provjeravamo korištenje velikih podataka. Također smatramo da je ključni faktor obuka naših inženjera i znanstvenika o podacima i etičkim standardima te da bi trebala biti slična pravilima koja postavljamo u naše pravne i zdravstvene sustave.

Popis literature

1. Berman, J. (2013). *Principles of Big Data: preparing, sharing, and analyzing complex information*, Amsterdam
2. Bell, E., Owen T., Khorana S., Henrichsen J. (2017). *Journalism After Snowden: The Future of the Free Press in the Surveillance State*, New York
3. Collmann, J., Matei, S. (2016). *Ethical Reasoning in Big Data: An Exploratory Analysis*, Springer, Washington
4. Čizmić, J., Boban, M. (2018). Učinak nove EU uredbe 2016/679 (GDPR) na zaštitu osobnih podataka u Republici Hrvatskoj, *Zbornik Pravnog fakulteta Sveučilišta u Rijeci*, vol. 39(1), 377-410
5. Davis, K. (2012). *Ethics of Big Data: Balancing Risk and Innovation*, O'Reilly Media, Sebastopol
6. Diebold, F. X. (2000). Big Data: Dynamic Factor Models for Macroeconomic Measurement and Forecasting, *Cambridge University Press*, vol. 3, 115-122
7. Elezović, A. (2012). O čemu pišu novine? Analiza sadržaja novinskih naslovnica (siječanj – lipanj 2011.), *Medijska istraživanja*, vol.18 (1), 61-88
8. Frančula, N. (2018). Big Data & Society, *Geodetski list*, 72(1): 69
9. Gerlitz, C., Helmond, A. (2013). The Like Economy: Social Buttons and the Data-Intensive Web, *New Media & Society*, vol. 15 (8), 1348-1365
10. Grover V., Chiang., Liang, T.,Zhang, D.(2018). Creating Strategic Business Value from Big Data Analytics: A Research Framework, *Journal of Management Information Systems*, vol. 35(2), 388-423
11. Glenn, R. (2003). *The Right to Privacy: Rights and Liberties Under the Law*, Abc-Clio, Santa Barbara
12. González, Gloria. (2014). *The Emergence of Personal Data Protection as a Fundamental Right of the EU*, Springer, Brussels
13. Jalšenjak, B.,Krkač, K. (2016). *Poslovna etika, korporacijska društvena odgovornost i održivost*, Mate d.o.o., Zagreb

14. Kocijan, K. (2014). Big Data: kako smo došli do Velikih podataka i kamo nas oni vode, *Zavod za informacijske studije*, 37-62

15. Krippendorff, K. (1980). *Content analysis: an introduction to its methodology*, Sage Publications, California

16. Kuutti, H. (2016). Ethics of data journalism: Four ethical phases in the working proces, A paper presented at Nordic Data Journalism Conference NODA 2016, 21 April 2016 in Helsinki, Finland

17. Lashar, P. (2016). No More Sources?: The impact of Snowden's revelations on journalists and their confidential sources, *Journalism Practice*, vol.11 (6), 1-24

18. Martin, K. (2015). Ethical Issues in the Big Data Industry, MIS Quarterly Executive, vol. 14 (2)

19. Mayer-Schönberger, V., Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, Boston

20. Milas, G. (2009). *Istraživačke metode u psihologiji i drugim istraživačkim znanostima*, Naklada Slap, Zagreb

21. Mirović, M., Miličević, M., Obradović, I. (2018). Big Data in the Maritime Industry Veliki skupovi podataka u pomorskoj industriji, *Naše More*, vol. 65 (1), 56-62

22. Musa, A., Bevandić D., Dvorski T., Jadrijević L. (2018). *Pravo na pristup informacijama: vodič za medije i udruge*, Zagreb

23. Newman, A. (2015). What the 'Right to be Forgotten' Means for Privacy in a Digital Age.' *Science*, vol. 347 (6221), 507–508

24. Patgiri, R., Ahmed, A. (2016). *Big Data: The V's of the Game Changer Paradigm*, 18th IEEE High Performance Computing and Communications, Sidney

25. Porlezza, C. (2019). Data journalism and the ethics of open source, Institute of Network Cultures, vol.29, 189-201

26. Reeve, A. (2013). *Managing Data in Motion: Data Integration Best Practice, Techniques and Technologies*, Waltham MA

27. Richterich, A. (2018). *The Big Data Agenda: Data Ethics and Critical Data Studies*, University of Westminster Press, London

28. Shabani, M., Pascal, B. (2017). Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation, *European Journal of Human Genetics*, vol. 26 (2), 149–156

29. Vuletić, S. (2017). Što očekuje javnozdravstvena profesija od „big data“ tehnologije, *Bilten Hrvatskog društva za medicinsku informatiku*, vol.23 (2), 26-27

30. Zion, Lawrie. (2015). *Ethics for Digital Journalists: Emerging Best Practises*

31. Wolfie, C. (2017). *How Companies Use Personal Data Against People*, Cracked Labs, Vienna

Internetski izvori

1. Opća uredba o zaštiti podataka, Zakon.hr objavljuje na stranici https://www.zakon.hr/, (stranica posjećena: 2. lipnja. 2019).

2. Macaskill, E., Hern, A. (2018). Edward Snowden: 'The people are still powerless, but now they're aware, The Guardian, https://www.theguardian.com/us-news/2018/jun/04/edward-snowden-people-still-powerless-but-aware (stranica posjećena: 11. lipnja 2019).

Prilozi

Istraživačka matrica

--1. Portali

1. 24 sata

2. Tportal

3. The Guardian

4. BBC

--2. Datum objave članka (napisati)

_____

--3. Glavni žanr

1. Informativno - izvještajne vrste (vijesti i izvještaji čija je osnovna namjera obavijestiti širi krug ljudi o određenim činjenicama).

2. Reportažno-komentatorske vrste (Reportaža je najsloženija novinarska vrsta koja događaj prenosi kroz priču koja ima uvod, zaplet, rasplet. Mora biti autentična, a tema joj je stvarni događaj. Osnovna funkcija je izazvati estetski doživljaj kod čitatelja. Osnovna značajka komentara je tumačenje, objašnjavanje onoga što se površinski ne vidi, a bitna je dobra argumentacija i snažan jezik).

3. Kolumne (Redovno ponavljajući članak u nekoj publikaciji kojeg pišu kolumnisti, a sadrži izričito gledište ili mišljenje).

--4. Tip članka

1. Vijest i kratka vijest (do 15 redaka)

2. Kratki prilog - izvještaj(iznad 15 redaka - do pola stranice)

3. Analitički prilog

4. Ostalo

--5. Rubrika

1. News/Vijesti/Aktualno
2. Svijet
3. Politika

4. Novac/ Biznis
5. Kultura
6. Znanost
7. Tehnologija
8. Popularno
9. Life/Style
10. Zdravlje/Medicina
11. Scena/Show/Spektakli
12. Sport
13. Komentari
14. Kultura
15. Mišljenje (Opinion)
16. Zabava/umjetnost (Entertainment/Arts)
17. Gradovi
18. Okolina
19. Ostalo

--6. Autor članka

1. Autorski rad (ime i prezime)
2. Autorski rad (inicijali)
3. Promo
4. Redigirana agencijska vijest (HINA, EPEHA ...)
5. Bez potpisa
6. Ostalo
7. Pismo čitatelja, autor: čitatelj
8. Redakcijski tekst

--7. Broj izvora[15]

[15] U izvore spadaju imenovani i neimenovani izvori.
Imenovani izvori (jasno identificirani) mora biti napisano ime i prezime ili inicijali osobe; ime agencije/institucije/državne službe/ministarstva/novine/nov. agencije/udruge; službeni dokumenti čiji je naziv napisan (zakoni, konvencije ...); policijske postaje/bolnice/hitne pomoći čija je adresa/ime napisano; popis stanovništva s napisanom godinom .
Neimenovani izvori: anonimne osobe; visoki dužnosnici čije se ime ne spominje; agencije/udruge/institucije… bez imena; dokumenti bez naziva; podaci uz koje ne piše točan naziv agencije/institucije koja je provela istraživanje (npr. Prema podacima iz 2016. godine; prema podacima iz policije; - ne piše tko je proveo istraživanje niti iz koje PU su podaci).
Nisu izvori: biografije.

1. Navodi se samo jedan izvor

2. Navode se dva izvora

3. Navode se tri ili više izvora

4. Ne navodi se niti jedan izvor

--8. Neimenovani izvori (Ako nema izvora, piše se 0)

1. Autor samo spominje neimenovane izvore

2. Autor se djelomično oslanja na neimenovane izvore

3. Autor se u potpunosti i isključivo oslanja na neimenovane izvore

4. Teško je odrediti

5. Nema neimenovanih izvora (svi izvori su imenovani)

--9. Informacija i stavovi (novinara)

1. U tekstu su navedene samo informacije
2. U tekstu je naveden samo stav - komentari
3. U tekstu postoje informacije, stav - komentari
4. Teško je utvrditi

--10. Kakvi su naslovi prisutni u tekstu

1. Samo glavni naslov
2. Nadnaslov, naslov, podnaslov, međunaslov
3. Nadnaslov, naslov, podnaslov
4. Naslov, podnaslov, međunaslov
5. Naslov i međunaslov
6. Nadnaslov, naslov
7. Naslov, podnaslov
8. Nadnaslov, naslov, međunaslov
9. Nema naslova

--11. Tip naslova (glavnog naslova)

1. Informativni: temelji se na podacima iz članka, uravnotežen, ne prenosi govor mržnje, ne donosi osude, doslovno prenosi citate.

2. Senzacionalistički: sve suprotno od informativnog, prozivanje, neujednačenost, ignoriranje etičkih pravila, prenošenje govora mržnje, kriva interpretacija podataka, sugestivna oprema, vrijeđa, proziva, opominje, manipulira, nema navodnih znakova.

3. Kritički: iz poznatih argumenata kritizira, ali ne vrijeđa, nije mu cilj destruktivnost afirmativnost, upućuje na propuste, poziva na akciju.

4. Teško je odrediti.

--12. Naslov članka odgovara tekstu članka
1. Da
2. Ne
3. Teško je odrediti

--13. Vrsta grafičke opreme

1. Nema je

2. Fotografija

3. Ilustracija

4. Grafika (tabele, grafovi, mape, info grafike)

5. Ostalo

6. Fotografija i/ili ilustracija i/ili grafika

--14. U tekstu se primjenjuje objektivni pristup izvještavanju
1. Da
2. Ne
3. Teško je utvrditi

--15. Tekst apelira[16] na opasnost za korisnike
1. Da
2. Ne
3. Teško je utvrditi

--16. U tekstu su veliki podaci prikazani (odnosi se na cijeli članak: na novinarov stav i citate)
1. Neutralno
2. Pozitivno
3. Negativno
4. Teško je utvrditi

--17. Velike podatke se predstavlja kao moguću prijetnju za osobnu sigurnost
1. Da
2. Ne

--18. Velike podatke se povezuje s velikim korporacijama i političkim strankama koje koriste osobne podatke u komercijalne svrhe?

[16] Tekst metaforički i asocijativno poziva na te emocije.

1. Da
2. Ne

--19. Koja je glavna tema zastupljena u tekstu?

1. GDPR
2. Afera Facebook (curenje podataka)
3. Krađa podataka
4. Prikupljanje podataka
5. Pozitivne primjene prikupljanja velikih podataka u medicini
6. Praćenje korisnika
7. Savjeti (zaštita, brisanje podataka)
8. Pozitivne primjene velikih podataka
9. Google (prikupljanje podataka)
10. Cambridge Analytica (prikupljanje podataka u političke svrhe, izbore)
11. Skladištenje podataka
12. Zaštita korisnika i podataka
13. Ništa od navedenog

--20. Dominantna funkcija teksta?

1. Informirati
2. Educirati
3. Zabaviti
4. Ukazati na problem
5. Promovirati
6. Teško je odrediti

--21. Spominje li se etičnost u naslovu?

1. Da
2. Ne

--22. Spominje li se u tekstu etičnost pri korištenju velikih podataka?

1. Da
2. Ne
3. U kontekstu[17]
4. Teško je odrediti

--23. Ako se etičnost spominje, je li spominjanje površno ili je dublje obrađeno?

[17] Okružje nekog teksta ili elementa teksta (riječi, fraze, rečenice) koje s njime tvori smislenu cjelinu

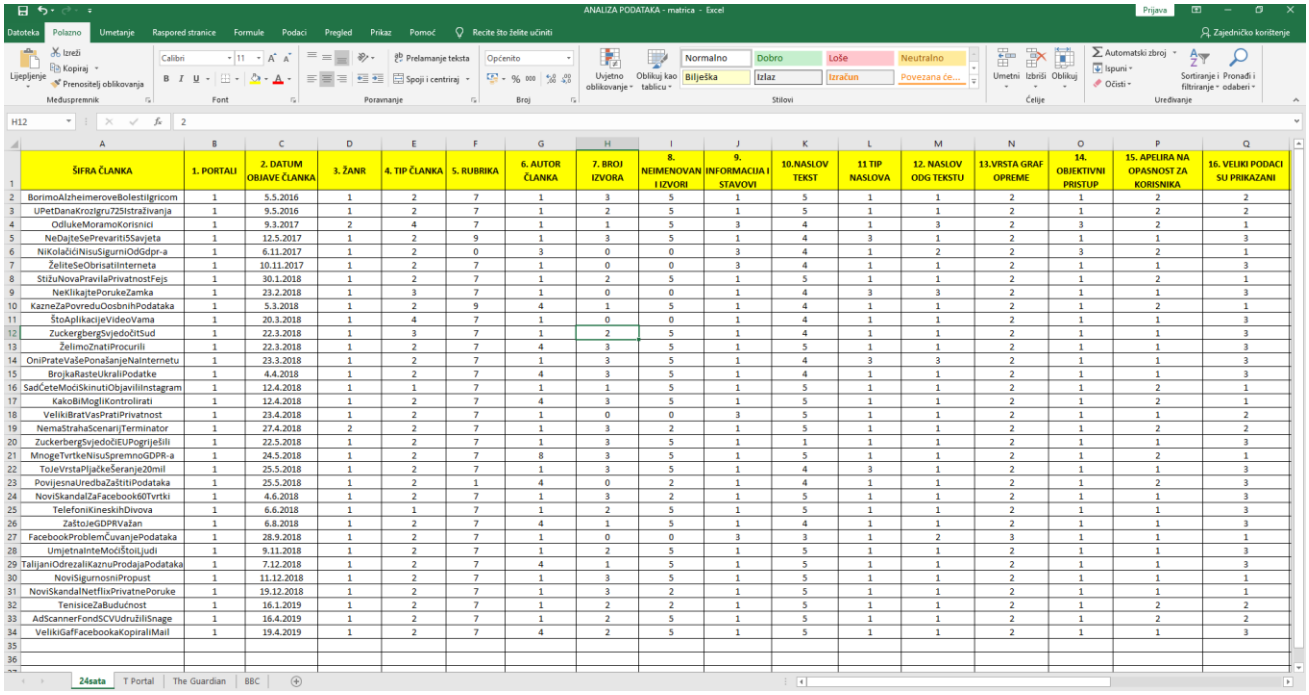
1. Površno
2. Dublja obrada
3. Ne spominje se uopće

--24. Spominje li se u tekstu zaštita podataka?

1. Da
2. Ne

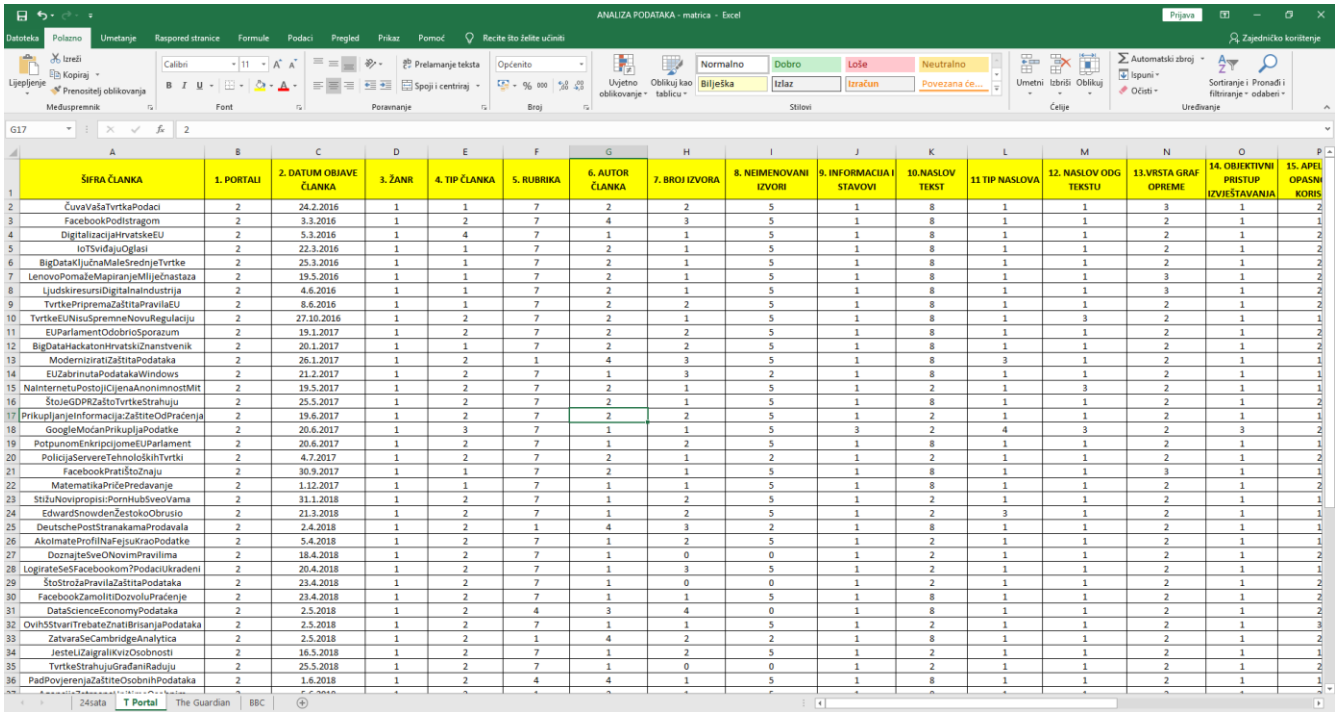
Kodne liste

24sata



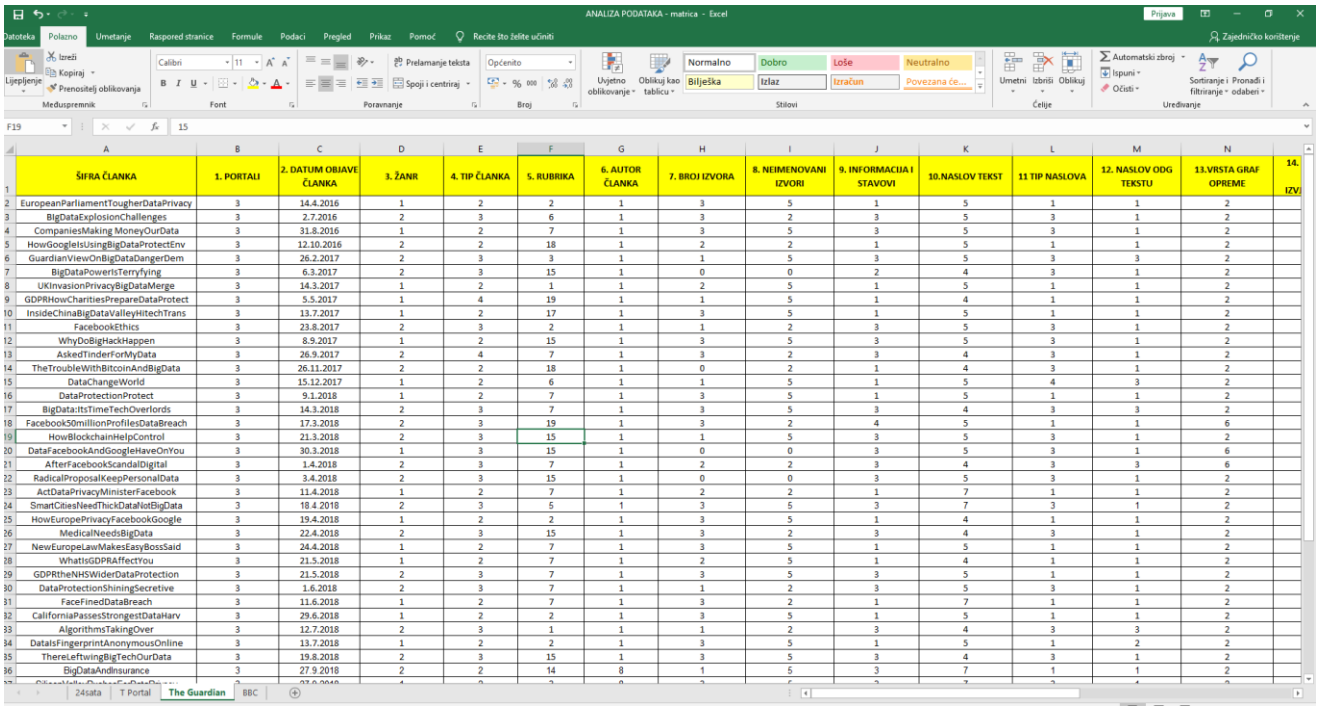
ŠIFRA ČLANKA	1. PORTALI	2. DATUM OBJAVE ČLANKA	3. ŽANR	4. TIP ČLANKA	5. RUBRIKA	6. AUTOR ČLANKA	7. BROJ IZVORA	8. NEIMENOVANI IZVORI	9. INFORMACIJA I STAVOVI	10.NASLOV TEKST	11 TIP NASLOVA	12. NASLOV ODG TEKSTU	13.VRSTA GRAF OPREME	14. OBJEKTIVNI PRISTUP	15. APELIRA NA OPASNOST ZA KORISNIKA	16. VELIKI PODACI SU PRIKAZANI
BorimoAlzheimeroveBolestiIgricom	1	5.5.2016	1	2	7	1	3	5	1	5	1	1	2	1	2	2
UPetDanaKrozIgru725IstraživanjaUPetDanaKrozIgru725IstraživanjaUPetDanaKrozIgru725IstraživanjaUPetDanaKrozIgru725IstraživanjaUPetDanaKrozIgru725IstraživanjaUPetDanaKrozIgru725Istraživanja	1	9.5.2016	1	2	7	1	2	5	1	5	1	1	2	1	2	2
OdlukeMoramoKorisnici	1	9.3.2017	2	4	7	1	1	5	3	4	1	3	2	3	2	1
NeDajteSePrevariti5Savjeta	1	12.5.2017	1	2	9	1	3	5	1	4	3	1	2	1	1	3
NiKolačićiNisuSigurniOdGdpr-a	1	6.11.2017	1	2	0	3	0	0	3	4	1	2	2	3	2	1
ŽeliteSeObrisatiInterneta	1	10.11.2017	1	2	7	1	0	0	3	4	1	1	2	1	1	3
StižuNovaPravilaPrivatnostFejs	1	30.1.2018	1	2	7	1	2	5	1	5	1	1	2	1	2	1
NeKlikajtePorukeZamka	1	23.2.2018	1	3	7	1	0	0	1	4	3	3	2	1	1	3
KazneZaPovreduOosbnihPodataka	1	5.3.2018	1	2	9	4	1	5	1	4	1	1	2	1	2	1
ŠtoAplikacijeVideoVama	1	20.3.2018	1	4	7	1	0	0	1	4	1	1	2	1	1	3
ZuckergbergSvjedočitSud	1	22.3.2018	1	3	7	1	2	5	1	4	1	1	2	1	1	3
ŽelimoŽnatiProcurili	1	22.3.2018	1	2	7	4	3	5	1	5	1	1	2	1	1	3
OniPrateVašePonašanjeNaInternetu	1	23.3.2018	1	2	7	1	3	5	1	4	3	3	2	1	1	3
BrojkaRasteUkraliPodatke	1	4.4.2018	1	2	7	4	3	5	1	4	1	1	2	1	1	3
SadCeteMociSkinutiObjaviliInstagram	1	12.4.2018	1	1	7	1	1	5	1	5	1	1	2	1	2	1
KakoBiMogliKontrolirati	1	12.4.2018	1	2	7	4	3	5	1	5	1	1	2	1	2	1
VelikiBratVasPratiPrivatnost	1	23.4.2018	1	2	7	1	0	0	3	5	1	1	2	1	1	2
NemaStrahaScenarijTerminator	1	27.4.2018	2	2	7	1	3	2	1	5	1	1	2	1	2	2
ZuckerbergSvjedočiEUPogriješili	1	22.5.2018	1	2	7	1	3	5	1	1	1	1	2	1	1	3
MnogeTvrtkeNisuSpremnoGDPR-a	1	24.5.2018	1	2	7	8	3	5	1	5	1	1	2	1	2	1
ToJeVrstaPljačkeSeranje20mil	1	25.5.2018	1	2	7	1	3	5	1	4	3	1	2	1	1	3
PovijesnaUredbaZaštitiPodataka	1	25.5.2018	1	2	1	4	0	2	1	4	1	1	2	1	2	3
NoviSkandalZaFacebook60Tvrtki	1	4.6.2018	1	2	7	1	3	2	1	5	1	1	2	1	1	3
TelefoniKineskihDivova	1	6.6.2018	1	1	7	1	2	5	1	5	1	1	2	1	1	3
ZaštoJeGDPRVažan	1	6.8.2018	1	2	7	4	1	5	1	4	1	1	2	1	1	3
FacebookProblemČuvanjePodataka	1	28.9.2018	1	2	7	1	0	0	3	3	1	2	3	1	1	1
UmjetnaInteMoćiŠtoiLjudi	1	9.11.2018	1	2	7	1	2	5	1	5	1	1	2	1	1	3
TalijaniOdrezaliKaznuProdajaPodataka	1	7.12.2018	1	2	7	4	1	5	1	5	1	1	2	1	1	3
NoviSigurnosniPropust	1	11.12.2018	1	2	7	1	3	5	1	5	1	1	2	1	1	1
NoviSkandalNetflixPrivatnePoruke	1	19.12.2018	1	2	7	1	3	2	1	5	1	1	2	1	1	1
TenisiceZaBudućnost	1	16.1.2019	1	2	7	1	2	2	1	5	1	1	2	1	2	2
AdScannerFondSCVUdružiliSnage	1	16.4.2019	1	2	7	1	2	5	1	5	1	1	2	1	2	2
VelikiGafFacebookaKopiraliMail	1	19.4.2019	1	2	7	4	2	5	1	5	1	1	2	1	1	3

Tportal



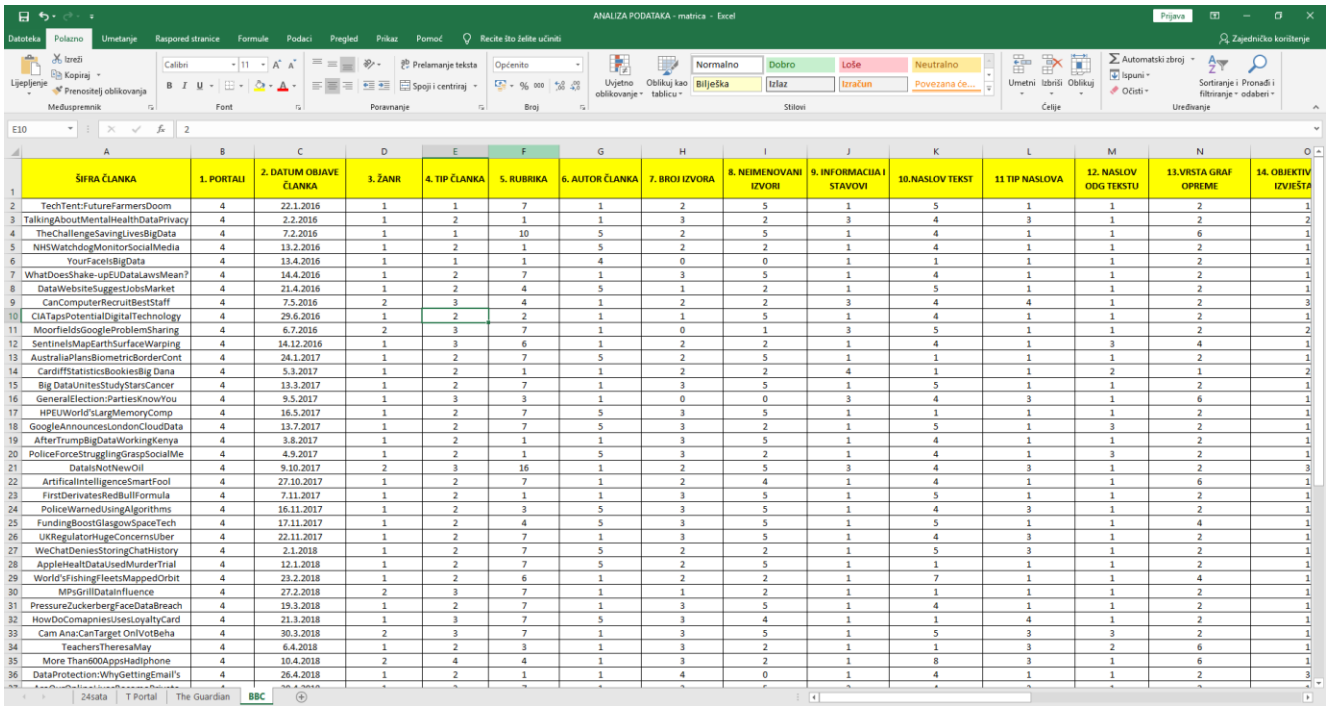
ŠIFRA ČLANKA	1. PORTALI	2. DATUM OBJAVE ČLANKA	3. ŽANR	4. TIP ČLANKA	5. RUBRIKA	6. AUTOR ČLANKA	7. BROJ IZVORA	8. NEIMENOVANI IZVORI	9. INFORMACIJA I STAVOVI	10.NASLOV TEKST	11 TIP NASLOVA	12. NASLOV ODG TEKSTU	13.VRSTA GRAF OPREME	14. OBJEKTIVNI PRISTUP IZVJEŠTAVANJA	15. APEL OPASN KORIS
ČuvaVašaTvrtkaPodaci	2	24.2.2016	1	1	7	2	2	5	1	8	1	1	3	1	2
FacebookPodIstragom	2	3.3.2016	1	2	7	4	3	5	1	8	1	1	2	1	1
DigitalizacijaHrvatskeEU	2	5.3.2016	1	4	7	1	1	5	1	8	1	1	2	1	2
IoTSvidjaUoglasi	2	22.3.2016	1	1	7	2	1	5	1	8	1	1	2	1	2
BigDataKljučnaMaleSredinjeTvrtke	2	25.3.2016	1	1	7	2	1	5	1	8	1	1	2	1	2
LenovoPomažeMapiranjeMiliječnastaza	2	19.5.2016	1	1	7	2	1	5	1	8	1	1	3	1	2
LjudskiresursiDigitalnaIndustrija	2	4.6.2016	1	1	7	2	1	5	1	8	1	1	3	1	2
TvrtkePripremaZaštitaPravilaEU	2	8.6.2016	1	1	7	2	2	5	1	8	1	1	2	1	2
TvrtkeEUnisuSpremneNovuRegulaciju	2	27.10.2016	1	2	7	2	1	5	1	8	1	3	2	1	1
EUParlamentOdobrioSporazum	2	19.1.2017	1	2	7	2	2	5	1	8	1	1	2	1	2
BigDataHackatonHrvatskiZnanstvenik	2	20.1.2017	1	1	7	2	2	5	1	8	1	1	2	1	2
ModernizirajuZaštitaPodataka	2	26.1.2017	1	2	1	4	3	5	1	8	3	1	2	1	1
EUZabrinutaPodatakaWindows	2	21.2.2017	1	2	7	1	3	2	1	8	1	1	2	1	1
NaInternetuPostojiCijenaAnonimnostMit	2	19.5.2017	1	2	7	2	1	5	1	2	1	3	2	1	1
ŠtoJeGDPRZaštoTvrtkeStrahuju	2	25.5.2017	1	2	7	2	1	5	1	8	1	1	2	1	2
PrikupljanjeInformacija:ZaštiteOdPraćenja	2	19.6.2017	1	2	7	2	2	5	1	2	1	1	2	1	1
GoogleMočanPrikupljaPodatke	2	20.6.2017	1	3	7	1	1	5	3	2	4	3	2	3	2
PotpunomEnkripcijomeEUParlament	2	20.6.2017	1	2	7	1	2	5	1	8	1	1	2	1	1
PolicijaServereTehnološkihTvrtki	2	4.7.2017	1	2	7	2	1	2	1	2	1	1	2	1	2
FacebookPratiŠtoZnaju	2	30.9.2017	1	1	7	2	1	5	1	8	1	1	3	1	1
MatematikaPričePredavanje	2	1.12.2017	1	1	7	1	1	5	1	8	1	1	2	1	2
StižuNovipropisi:PornHubSveoVama	2	31.1.2018	1	2	7	1	2	5	1	2	1	1	2	1	1
EdwardSnowdenŽestokoObrusio	2	21.3.2018	1	2	7	1	2	5	1	2	3	1	2	1	1
DeutschePostStranakamaProdavala	2	2.4.2018	1	2	1	4	3	2	1	8	1	1	2	1	1
AkoImateProfilNaFejsuKraoPodatke	2	5.4.2018	1	2	7	1	2	5	1	2	1	1	2	1	1
DoznajteSveONovimPravilima	2	18.4.2018	1	2	7	1	0	0	1	2	1	1	2	1	2
LogirateSeSFacebookom?PodaciUkradeni	2	20.4.2018	1	2	7	1	3	5	1	2	1	1	2	1	1
ŠtoStrožaPravilaZaštitaPodataka	2	23.4.2018	1	2	7	1	0	0	1	2	1	1	2	1	2
FacebookZamoliliDozvoluPraćenje	2	23.4.2018	1	2	7	1	1	5	1	8	1	1	2	1	2
DataScienceEconomyPodataka	2	2.5.2018	1	2	4	3	4	0	1	8	1	1	2	1	2
Ovih5StvariTrebateZnatiBrisanjaPodataka	2	2.5.2018	1	2	7	1	1	5	1	2	1	1	2	1	3
ZatvaraSeCambridgeAnalytica	2	2.5.2018	1	2	1	4	2	2	1	8	1	1	2	1	2
JesteLiZaigraliKvizOsobnosti	2	16.5.2018	1	2	7	1	2	5	1	2	1	1	2	1	1
TvrtkeStrahujuGrađaniRaduju	2	25.5.2018	1	2	7	1	0	0	1	2	1	1	2	1	2
PadPovjerenjaZaštiteOsobnihPodataka	2	1.6.2018	1	2	4	4	1	5	1	8	1	1	2	1	1

The Guardian



ŠIFRA ČLANKA	1. PORTALI	2. DATUM OBJAVE ČLANKA	3. ŽANR	4. TIP ČLANKA	5. RUBRIKA	6. AUTOR ČLANKA	7. BROJ IZVORA	8. NEIMENOVANI IZVORI	9. INFORMACIJA I STAVOVI	10.NASLOV TEKST	11 TIP NASLOVA	12. NASLOV ODG TEKSTU	13.VRSTA GRAF OPREME	14. IZV
EuropeanParliamentTougherDataPrivacy	3	14.4.2016	1	2	2	1	3	5	1	5	1	1	2	
BigDataExplosionChallenges	3	2.7.2016	2	3	6	1	3	2	3	5	3	1	2	
CompaniesMaking MoneyOurData	3	31.8.2016	1	2	7	1	3	5	3	5	3	1	2	
HowGoogleIsUsingBigDataProtectEnv	3	12.10.2016	2	2	18	1	2	2	1	5	1	1	2	
GuardianViewOnBigDataDangerDem	3	26.2.2017	2	3	3	1	1	5	3	5	3	3	2	
BigDataPowerIsTerryfying	3	6.3.2017	2	3	15	1	0	0	2	4	3	1	2	
UKInvasionPrivacyBigDataMerge	3	14.3.2017	1	2	1	1	2	5	1	5	1	1	2	
GDPRHowCharitiesPrepareDataProtect	3	5.5.2017	1	4	19	1	1	5	1	4	1	1	2	
InsideChinaBigDataValleyHitechTrans	3	13.7.2017	1	2	17	1	3	5	1	5	1	1	2	
FacebookEthics	3	23.8.2017	2	3	2	1	1	2	3	5	3	1	2	
WhyDoBigHackHappen	3	8.9.2017	1	2	15	1	3	5	3	5	3	1	2	
AskedTinderForMyData	3	26.9.2017	2	4	7	1	3	2	3	4	3	1	2	
TheTroubleWithBitcoinAndBigData	3	26.11.2017	2	2	18	1	0	2	1	4	3	1	2	
DataChangeWorld	3	15.12.2017	1	2	6	1	1	5	1	5	4	3	2	
DataProtectionProtect	3	9.1.2018	1	2	7	1	3	5	1	5	1	1	2	
BigData:ItsTimeTechOverlords	3	14.3.2018	2	3	7	1	3	5	3	4	3	3	2	
Facebook50millionProfilesDataBreach	3	17.3.2018	2	3	19	1	3	2	4	5	1	1	6	
HowBlockchainHelpControl	3	21.3.2018	2	3	15	1	1	5	3	5	3	1	2	
DataFacebookAndGoogleHaveOnYou	3	30.3.2018	1	3	15	1	0	0	3	5	3	1	6	
AfterFacebookScandalDigital	3	1.4.2018	2	3	7	1	2	2	3	4	3	3	6	
RadicalProposalKeepPersonalData	3	3.4.2018	2	3	15	1	0	0	3	5	3	1	2	
ActDataPrivacyMinisterFacebook	3	11.4.2018	1	2	7	1	2	2	1	7	1	1	2	
SmartCitiesNeedThickDataNotBigData	3	18.4.2018	2	3	5	1	3	5	3	7	3	1	2	
HowEuropePrivacyFacebookGoogle	3	19.4.2018	1	2	2	1	3	5	1	4	1	1	2	
MedicalNeedsBigData	3	22.4.2018	2	3	15	1	3	2	3	4	3	1	2	
NewEuropeLawMakesEasyBossSaid	3	24.4.2018	1	2	7	1	3	5	1	5	1	1	2	
WhatIsGDPRAffectYou	3	21.5.2018	1	2	7	1	2	5	1	4	1	1	2	
GDPRtheNHSWiderDataProtection	3	21.5.2018	2	3	7	1	3	5	3	5	1	1	2	
DataProtectionShiningSecretive	3	1.6.2018	2	3	7	1	1	2	3	5	3	1	2	
FaceFinedDataBreach	3	11.6.2018	1	2	7	1	3	2	1	7	1	1	2	
CaliforniaPassesStrongestDataHarv	3	29.6.2018	1	2	2	1	3	5	1	5	1	1	2	
AlgorithmsTakingOver	3	12.7.2018	2	3	1	1	1	2	3	4	3	3	2	
DataIsFingerprintAnonymousOnline	3	13.7.2018	1	2	2	1	3	5	1	5	1	2	2	
ThereLeftwingBigTechOurData	3	19.8.2018	2	3	15	1	3	5	3	4	3	1	2	
BigDataAndInsurance	3	27.9.2018	2	2	14	8	1	5	3	7	1	1	2	

BBC



ŠIFRA ČLANKA	1. PORTALI	2. DATUM OBJAVE ČLANKA	3. ŽANR	4. TIP ČLANKA	5. RUBRIKA	6. AUTOR ČLANKA	7. BROJ IZVORA	8. NEIMENOVANI IZVORI	9. INFORMACIJA I STAVOVI	10.NASLOV TEKST	11 TIP NASLOVA	12. NASLOV ODG TEKSTU	13.VRSTA GRAF OPREME	14. OBJEKTIV IZVJEŠTA
TechTent:FutureFarmersDoom	4	22.1.2016	1	1	7	1	2	5	1	5	1	1	2	1
TalkingAboutMentalHealthDataPrivacy	4	2.2.2016	1	2	1	1	3	2	3	4	3	1	2	2
TheChallengeSavingLivesBigData	4	7.2.2016	1	1	10	5	2	5	1	4	1	1	6	1
NHSWatchdogMonitorSocialMedia	4	13.2.2016	1	2	1	5	2	2	1	4	1	1	2	1
YourFaceIsBigData	4	13.4.2016	1	1	1	4	0	0	1	1	1	1	2	1
WhatDoesShake-upEUDataLawsMean?	4	14.4.2016	1	2	7	1	3	5	1	4	1	1	2	1
DataWebsiteSuggestJobsMarket	4	21.4.2016	1	2	4	5	1	2	1	5	1	1	2	1
CanComputerRecruitBestStaff	4	7.5.2016	2	3	4	1	2	2	3	4	4	1	2	3
CIATapsPotentialDigitalTechnology	4	29.6.2016	1	2	2	1	1	5	1	4	1	1	2	1
MoorfieldsGoogleProblemSharing	4	6.7.2016	2	3	7	1	0	1	3	5	1	1	2	2
SentinelsMapEarthSurfaceWarping	4	14.12.2016	1	3	6	1	2	2	1	4	1	3	4	1
AustraliaPlansBiometricBorderCont	4	24.1.2017	1	2	7	5	2	5	1	1	1	1	2	1
CardiffStatisticsBookiesBig Dana	4	5.3.2017	1	2	1	1	2	2	4	1	1	2	1	2
Big DataUnitesStudyStarsCancer	4	13.3.2017	1	2	7	1	3	5	1	5	1	1	2	1
GeneralElection:PartiesKnowYou	4	9.5.2017	1	3	3	1	0	0	3	4	3	1	6	1
HPEUWorld'sLargMemoryComp	4	16.5.2017	1	2	7	5	3	5	1	1	1	1	2	1
GoogleAnnouncesLondonCloudData	4	13.7.2017	1	2	7	5	3	2	1	5	1	3	2	1
AfterTrumpBigDataWorkingKenya	4	3.8.2017	1	2	1	1	3	5	1	4	1	1	2	1
PoliceForceStrugglingGraspSocialMe	4	4.9.2017	1	2	1	5	3	2	1	4	1	3	2	1
DataIsNotNewOil	4	9.10.2017	2	3	16	1	2	5	3	4	3	1	2	3
ArtificalIntelligenceSmartFool	4	27.10.2017	1	2	7	1	2	4	1	4	1	1	6	1
FirstDerivatesRedBullFormula	4	7.11.2017	1	2	1	1	3	5	1	5	1	1	2	1
PoliceWarnedUsingAlgorithms	4	16.11.2017	1	2	3	5	3	5	1	4	3	1	2	1
FundingBoostGlasgowSpaceTech	4	17.11.2017	1	2	4	5	3	5	1	5	1	1	4	1
UKRegulatorHugeConcernsUber	4	22.11.2017	1	2	7	1	3	5	1	4	3	1	2	1
WeChatDeniesStoringChatHistory	4	2.1.2018	1	2	7	5	2	2	1	5	3	1	2	1
AppleHealtDataUsedMurderTrial	4	12.1.2018	1	2	7	5	2	5	1	1	1	1	2	1
World'sFishingFleetsMappedOrbit	4	23.2.2018	1	2	6	1	2	2	1	7	1	1	4	1
MPsGrillDataInfluence	4	27.2.2018	2	3	7	1	1	2	1	1	1	1	2	1
PressureZuckerbergFaceDataBreach	4	19.3.2018	1	2	7	1	3	5	1	4	1	1	2	1
HowDoComapniesUsesLoyaltyCard	4	21.3.2018	1	3	7	5	3	4	1	1	4	1	2	1
Cam Ana:CanTarget OnlYotBeha	4	30.3.2018	2	3	7	1	3	5	1	5	3	3	2	1
TeachersTheresaMay	4	6.4.2018	1	2	3	1	3	2	1	1	3	2	6	1
More Than600AppsHadIphone	4	10.4.2018	2	4	4	1	3	2	1	8	3	1	6	1
DataProtection:WhyGettingEmail's	4	26.4.2018	1	2	1	1	4	0	1	4	1	1	2	3