

Google Trends kao metoda za rano detektiranje pojave novih slučajeva COVID-a 19

Tado Jurić

Hrvatsko katoličko sveučilište, Zagreb

SAŽETAK Znajući da je *guglanje* aktivnost koju pojedinci najprije poduzimaju pri traženju odgovora na brojna pitanja, pa tako i pitanje zdravlja, Google je uspio stvoriti iznimno korisnu bazu podataka o praćenju i širenju pandemije COVID-a 19 u SAD-u. Donosimo rezultate testiranja metode na primjeru Hrvatske koja omogućuje otkrivanje i predviđanje pojave novih slučajeva COVID-a 19 u ranoj fazi koristeći se analitičkim alatom *Google Trends* (GT). Osnovna korist predloženog pristupa ogleda se u detektiranju novih lokacija i žarišta širenja pandemije, čime se mogu pravovremeno preusmjeriti resursi javnog zdravstva te djelovati preventivno u daljnjem širenju zaraze, primjerice uvođenjem dodatnih mjera na određenom području

KLJUČNE RIJEČI COVID-19; *Google*; *Google Trends*; Hrvatska; predviđanje širenja pandemije

Primjenom umjetne inteligencije u obradi podataka iz baza nadzora javnog zdravlja u SAD-u prvi je put točno predviđeno širenje COVID-a 19.¹ Uslijed takvih uspjeha, jedna od posljedica pandemije nedvojbeno će biti jačanje uloge umjetne inteligencije i usvajanje digitalnih pristupa u raznim poljima znanstvenih istraživanja.²

KAKO ANALIZIRANJE GOOGLE PRETRAŽIVANJA MOŽE PODRŽATI ISTRAŽIVANJE COVID-a 19

Google Trends se borbi protiv pandemije priključio omogućavajući istraživačima dostupnost skupova trendova pretraživanja s ciljem proučavanja povezanosti između pretraživanja pojmova povezanih sa simptomima i širenja COVID-a 19.³ Znajući da je *guglanje* aktivnost koju pojedinci najprije poduzimaju pri traženju odgovora na brojna pitanja, pa tako i pitanje zdravlja, *Google* je uspio stvoriti iznimno korisnu bazu podataka o praćenju i širenju pandemije COVID-19 u SAD-u. Postavlja se pitanje mogu li podaci pretraživanja biti korisni i u hrvatskom slučaju.

U prošlosti su istraživači koristili podatke *Google* pretraživanja kako bi procijenili utjecaj vrućine na zdravlje, poboljšali modele predviđanja izbijanja gripe i pratili učestalost Lajmske bolesti.⁴ Ferguson i suradnici (2005.) pokazali su da pravovremeno otkrivanje pojave gripe može bitno smanjiti njene posljedice u populaciji.⁵ Ginsberg i suradnici (2009.) pokazali su kako poboljšati rano otkrivanje širenja gripe koristeći podatke tražilice *Google* u SAD-u.⁶ Tim su ispitivanjem predstavili i metodu analiziranja velikih skupova podataka dobivenih

putem pretraga na tražilici *Google*. „Budući da je relativna učestalost određenih upita u visokoj korelaciji s postotkom posjeta liječniku u kojem pacijenti pokazuju simptome gripe, može se točno procijeniti razina tjedne aktivnosti gripe u svakoj regiji Sjedinjenih Država.“⁶

U Njemačkoj su rađena ispitivanja korelacije između liječenja simptoma alergije, pojave peludi u zraku i ponašanja pretraživanja na *Google* tražilici.⁷ Aktivnosti pretraživanja koje koriste izraze kao što su „alergija“, „alergije“ i „pelud“ pokazale su korelaciju s pojavom peludi i u SAD-u.⁸ Tim ispitivanjima je uočeno da GT odražava stvarnu pojavu alergijskog rinitisa te je uvršten kao jedna od metoda za praćenje tih pojava kao i učinkovit pristup u predviđanju broja posjeta hitnim službama.¹⁰

Trendovi pretraživanja također su signalizirali koliko ljudi slušaju savjete javnih zdravstvenih službi, što se može koristiti vezano uz nošenje maski i sl.³

Metoda je tim više uspješna što više korisnika koristi internet, odnosno *Google* tražilicu. Do sredine 2020. godine 58% svjetske populacije koristilo je internet, u usporedbi s gotovo 90% u Europskoj uniji.

Unutar EU-a, isto ispitivanje pokazalo je da se uporaba kretala od gotovo 98% u Danskoj do manje od 70% u Bugarskoj, dok je u Hrvatskoj taj omjer 79%. Što se tiče najpopularnije tražilice, u Hrvatskoj je to bez sumnji *Google* s 91,8% korisnika.¹¹

METODOLOŠKO OBJAŠNJENJE

Osnovna hipoteza rada je da je analitički alat *Google Trends* koristan izvor podataka za utvrđivanje, procjenu i predviđanje povećanja broja novih aktivnih slučajeva

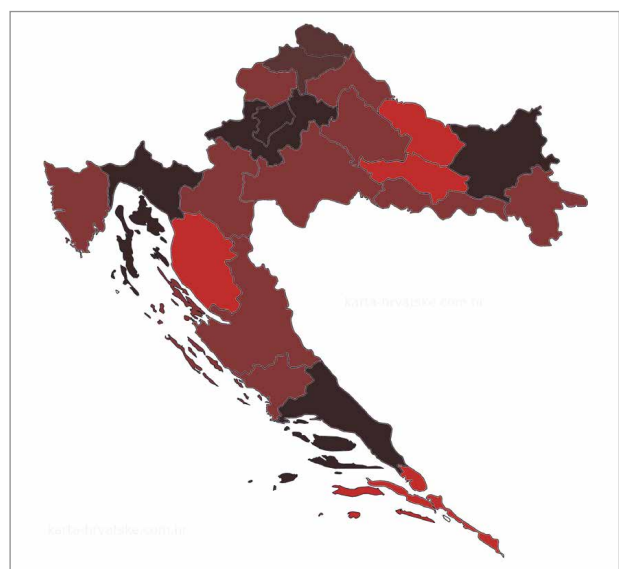
kako na razini cijele države tako i po regijama u Hrvatskoj. Osnovni metodološki koncept praćenje je digitalnih tragova pretraživanja po ključnim pojmovima

pomoću spomenutog analitičkog alata.¹² Najvažnije je ispravno identificirati pojmove čije pretraživanje daje najviše indicija. U radu su korišteni samo anonimni, zbirni podaci. Svi su podaci prikupljeni slijedeći GDPR i etička načela te se nijedan upit u bazi podataka opisanog projekta ne može povezati s određenom osobom. Baza podataka ne sadrži informacije o identitetu, IP adresi ili geolokaciji bilo kojeg korisnika.

U svrhu standardiziranja podataka, preuzeti su skupovi podataka za razdoblje od 20. siječnja 2020. do 20. veljače 2021. Zatim je podijeljena frekvencija ključnih riječi za odabrane pojmove, čime je dobiven indeks učestalosti pretraživanja koja su zatim uspoređena sa službenim statistikama kako bi se pokazao značaj rezultata (vidi daljnja objašnjenja kod Wilde i sur., 2020.¹³). U svrhu odabira najčešćih pojmova, korišteno je izvješće SZO-a (2020.)¹⁴ o simptomima zaraze te pojmovi koji su pokazali da su prediktivni, specifični i dovoljno uobičajeni u uporabi u hrvatskom jeziku. Nakon analize odabrane su sljedeće ključne riječi i teme, prikazane u tablici 1.

Iako su prethodna istraživanja na tom području pokazala izvedivost korištenja navedenih digitalnih podataka, istodobno je važno naglasiti probleme povezane s procjenama i zaključcima. Naime, neupitno je da još uvijek postoje značajna otvorena metodološka pitanja vezana uz korištenje izvora velikih skupova podataka u istraživanjima. Kada se koristi taj pristup, treba imati na umu da je svako od tih pretraživanja provedeno iz svog razloga i ne odgovara na izravna pitanja istraživača. Tako, primjerice, *guglanje* pojma „testiranje na koronavirus“ nije nužno implikacija da je netko bolestan ili ima simptome. Unatoč ograničenjima, postoje dokazi o visokoj korelaciji između takvog tipa istraživanja i službenih izvješća o slučajevima gripe.⁶ K tome, statistika nudi niz dostupnih alata za rješavanje problema nesavršenih podataka.¹⁶

SLIKA 3. Prijavljeni slučajevi u Hrvatskoj po regijama, ožujak 2020. - veljača 2021.

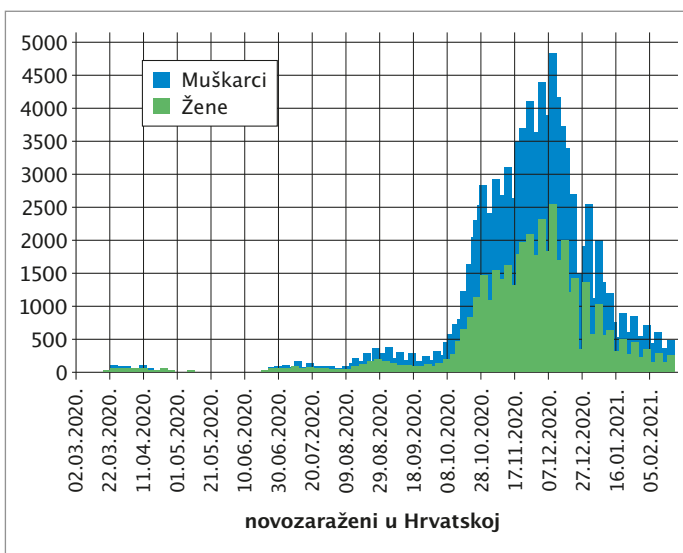


Legenda: Tamnije nijanse ukazuju na veći broj slučajeva. Izvor: koronavirus.hr, 2020.

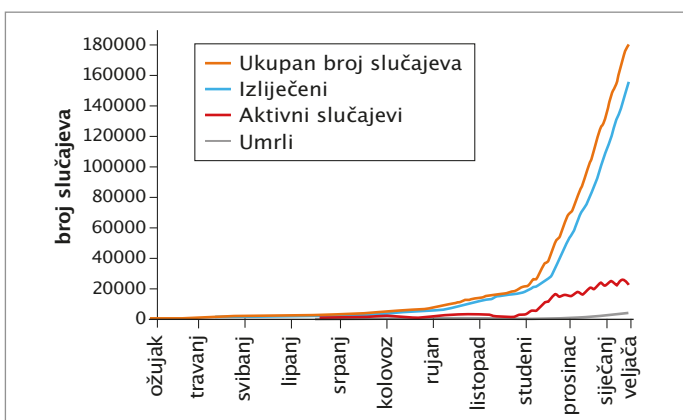
TABLICA 1. Kriteriji za odabir ključne riječi i teme		
Simptomi	Pojmovi	Aktivnosti
<ul style="list-style-type: none"> • Upala pluća • Prehlada • Kašalj • Depresija • Umor • Groznica • Glavobolja • Mučnina • Kratkoća daha • Upala pluća + korona • Suhi kašalj + korona • Curenje iz nosa + koronavirus • Bol u mišićima + korona 	<ul style="list-style-type: none"> • Gripa • Lijek protiv gripe/ koronavirusa/ COVID-a 19 • Opći simptomi gripe/COVID-a 19 • Simptomi gripe • Simptomi COVID-a 19 • Antibiotik • Lijekovi za koronu • Antivirusni lijekovi • Sumamed • Neofen 	<ul style="list-style-type: none"> • Testiranje na koronavirus • PCR test • Zahtjev za minimalnu plaću

Prilagođeno prema: Jurić, T. (2021) Monitoring families and COVID-19 with Google Trends: A Croatian case study (u objavi)

SLIKA 1. Ukupan broj aktivnih slučajeva u Hrvatskoj, ožujak 2020. - veljača 2021. (HZJZ, 2021.)



SLIKA 2. Broj izliječenih i aktivnih slučajeva u Hrvatskoj, ožujak 2020. - veljača 2021.



Izvor: koronavirus.hr, 2020.

POJAVA I ŠIRENJE KORONAVIRUSA U HRVATSKOJ

Prvi slučaj zaraze virusom SARS-CoV-2 u Hrvatskoj potvrđen je 25. veljače 2020.¹⁷ Sredinom veljače 2021. u Hrvatskoj je bilo 237.725 zaraženih. U ožujku 2020. godine donesena je Odluka o proglašenju epidemije bolesti COVID-19 na području Republike Hrvatske (11. ožujka 2020.).¹⁸

Početak rujna 2020. zabilježen je rekordan broj od 369 novih, a sredinom listopada 793 novih slučajeva. Tijekom listopada broj se udvostručuje, a do kraja listopada utrostručuje.¹⁷

Do kraja rujna 2020., prema procjenama Hrvatskog zavoda za javno zdravstvo, otprilike 10% stanovništva, tj. oko 400.000 osoba zaraženo je koronavirusom i razvilo je antitijela.¹⁹ Takozvani imunitet krda, tj. situacija u kojoj se infekcija više ne može snažno proširiti u populaciji zbog dovoljnog broja jedinki s razvijenim imunitetom, očekuje se početkom ljeta 2021. godine.

U pandemijskoj godini 2020. Hrvatska je imala najveći broj smrtnih slučajeva (osim u vrijeme II. svjetskog rata) otkako se bilježi evidencija umrlih – posebno tijekom studenog i prosinca 2020.²⁰ Prema preliminarnim procjenama, 2020. godine umrlo je 56.677 osoba,²¹ što je porast od 5000 u usporedbi s prosjekom posljednjih pet godina.

REZULTATI

Izbijanjem pandemije u Hrvatskoj i „zaključavanjem“ u ožujku 2020. godine, hrvatski građani počinju intenzivno *guglati* pojmove povezane s pandemijom bolesti COVID-19. Najčešće se pretražuju upiti: „što je suhi kašalj“, „što se smatra groznicom“, „koronavirus“, „razlika između gripe i koronavirusa“ i sl. Najčešći pretraživani pojmovi su uspoređeni sa službenim pokazateljima aktivnih slučajeva kako bi se provjerila povezanost između tih pojava.

Uočava se da porast pretraživanja „kašalj + korona“, „upala pluća + korona“, „suhi kašalj + korona“, „curenje iz nosa + koronavirus“, „bol u mišićima + korona“ korelira s porastom broja novih slučajeva. S druge strane, smanjenje opsega pretraživanja navedenih pojmova korelira sa smanjenjem broja novih slučajeva.²²

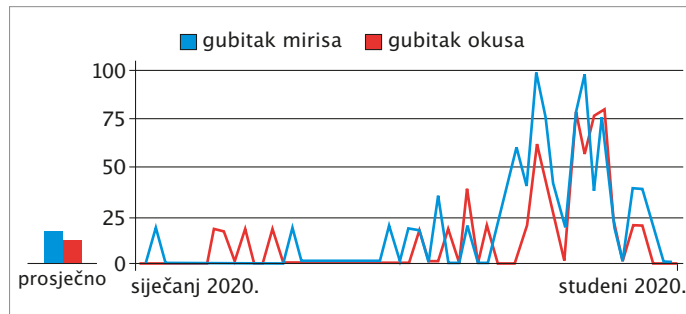
Kao i u prethodnom slučaju, porast pretraživanja „gubitak mirisa“ i „gubitak okusa“ korelira s porastom novih slučajeva oboljenja.

Istraženi su i upiti „Neofen“ i „Sumamed“, koji su pokazali da se potražnja za tim lijekovima povećala posebno u vrijeme izbijanja pandemije i u jesen 2020., kada je zabilježen najveći broj novih slučajeva.

Idući korisni uvidi vezani su uz geolokacije, čime se može pratiti stvaranje novih žarišta te sukladno tome preventivno djelovati uvođenjem dodatnih mjera u određenom prostoru poput zatvaranja škola i sl. Tim pristupom se može predvidjeti i porast pritiska na zdravstveni sustav u određenoj regiji.

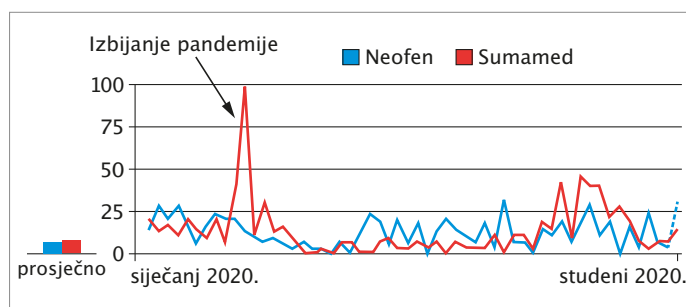
Potruga za lijekovima korelira s naglim porastom broja novih slučajeva u studenom 2020. u sjevernim

SLIKA 4. Google pretraživanje „gubitak mirisa“ i „gubitak okusa“ od siječnja 2020. do veljače 2021. u Hrvatskoj



Izvor: Jurić, T. (2021) Monitoring families and COVID-19 with Google Trends: A Croatian case study (u objavi)

SLIKA 5. Google pretraživanje „Neofen“ i „Sumamed“ od siječnja 2020. do veljače 2021. u Hrvatskoj



Izvor: Jurić, T. (2021) Monitoring families and COVID-19 with Google Trends: A Croatian case study (u objavi)

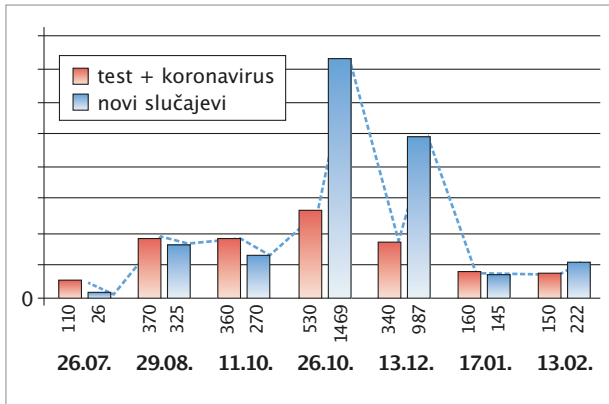
SLIKA 6. Korelacija po regijama



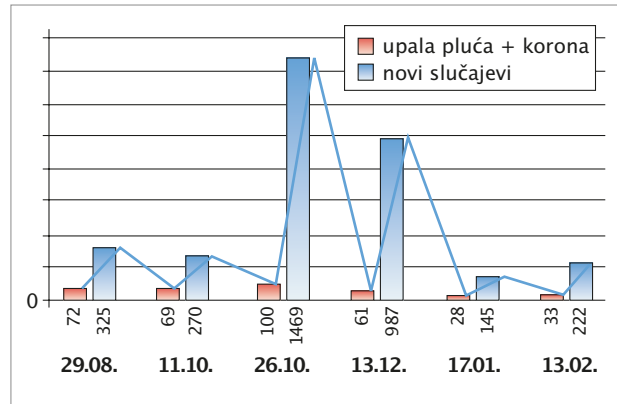
hrvatskim regijama,²³ kao i činjenicom da je Virovitičko-podravska županija imala najmanji broj zaraženih u tom razdoblju.

U daljnjem postupku prikazuje se standardizacija podataka, pri čemu se indeks učestalosti pretraživanja na Google tražilici uspoređuje sa službenim statistikama u svrhu pokazivanja značaja rezultata. Poseban fokus je na tzv. drugi val širenja infekcije, kada su slučajevi počeli eksponencijalno rasti.¹⁹

SLIKA 7. Korelacija između Googleova indeksa pretraživanja za upit „test + koronavirus“ i službenog broja prijavljenih novih slučajeva pacijenata oboljelih od COVID-a 19 u Hrvatskoj prema odabranim datumima 2020. i 2021. godine



SLIKA 8. Korelacija između Googleova indeksa pretraživanja za upit „upala pluća + korona“ i službenog broja prijavljenih novih slučajeva pacijenata COVID-19 u Hrvatskoj po odabranim datumima 2020. i 2021. godine



Prilagođeno prema: Jurić, T. (2021) Monitoring families and COVID-19 with Google Trends: A Croatian case study (u objavi)

Aktivnosti pretraživanja pojmovima „PCR + COVID“; „PCR + test“; „koronavirus + test“ su u visokoj korelaciji sa službenim izvješćima.

U nastavku je metoda testirana prema najčešćim simptomima: kašalj, upala pluća i bolovi u mišićima.

Aktivnosti pretraživanja vezanih uz simptome „kašalj + korona“, „upala pluća + korona“, „bolovi u mišićima + korona“ u snažnoj su korelaciji sa službenim podacima zabilježenih novih slučajeva.²⁴

ZAKLJUČAK

Google Trends analitički je alat koji pruža jedan od najnovativnijih širokopoljanskih sustava praćenja COVID-a 19 koji su dostupni. Aktivnosti pretraživanja na Googleu pri kojima se koriste upiti „PCR + Covid“, „PCR + test“ ili pak simptomi poput „kašalj + korona“, „upala pluća + korona“, „bolovi u mišićima + korona“ u velikoj su korelaciji sa službeno prijavljenim slučajevima bolesti,

dok je pad tih pretraživanja u korelaciji s padom broja novih slučajeva.

Testiranje metode pokazuje da se pristup može koristiti i vezano uz geolokacije, čime se može pratiti stvaranje novih žarišta te sukladno tome preventivno djelovati uvođenjem dodatnih mjera na određenoj lokaciji, poput zatvaranja škola i sl., kao i predvidjeti porast pritiska na zdravstveni sustav u određenoj regiji. Glavna prednost takvog pristupa u odnosu na službene pokazatelje je što se pravovremeno uočava izbijanje novih žarišta te može poslužiti kao rani alarm. Tom metodom javno zdravstvo dobiva novo oružje u borbi s pandemijom.

Međutim, bitno je naglasiti potreban oprez zbog niza metodoloških ograničenja kao i da se pristup može koristiti samo uz analizu cjelokupnoga društvenog konteksta. Primjerice, neuobičajeni događaji kao što su novonastale nedoumice vezane uz cjepivo AstraZeneca mogu u određenom razdoblju bitno utjecati na Google pretraživanja.

Google Trends as a method to predict new COVID-19 cases

Tado Jurić

Catholic University of Croatia, Zagreb, Croatia

SUMMARY Knowing that "googling" is an activity that individuals take first to find answers to many questions, including health, Google has managed to create an extremely useful database for monitoring the COVID-19 pandemic in the United States. In this paper, we present the results of testing a method on the example of the Republic of Croatia that enables the detection and prediction of new cases of COVID-19 at an early stage, using the analytical tool Google Trends (GT). The main benefit of the proposed approach is reflected in the timely detection of new locations and hotspots of pandemic spread. Such detection can redirect public health resources promptly and enable act preventively in the further spread of the infection, for example by introducing additional measures in a certain area.

KEY WORDS COVID-19; Croatia; Google; Google Trends; pandemic spread forecasting

LITERATURA

1. Niiler E. An AI Epidemiologist Sent the First Warnings of the Wuhan Virus. Jan 2020. Available from: <https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/>. Accessed on Jan 29th 2020.
2. Hantrais L, Allin P, Kritikos M, et al. Covid-19

- and the digital revolution. Contemporary Social Science: J Acad Soc Sci. Oct 2020.
3. Google search: Covid 19. (<https://www.google.com/search?client=firefox-b-d&q=covid+19>)
4. Gabrilovich E. Using symptoms search trends to inform COVID-19 research. Google Health. Sep 2020.

- Available from: <https://blog.google/technology/health/using-symptoms-search-trends-inform-covid-19-research>. Accessed Jan 4th 2021.
5. Ferguson NM, Cummings DA, Cauchemez S, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature. 2005 Sep;

437(7056):209–14.

6. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature*. 2009 Feb 19;457(7232):1012–4.
7. König V, Mösges R. A model for the determination of pollen count using google search queries for patients suffering from allergic rhinitis. *J Allergy (Cairo)*. 2014;2014:381983.
8. Willson TJ, Lospinoso J, Weitzel E, et al. Correlating regional aeroallergen effects on internet search activity. *Otolaryngol Head Neck Surg*. 2015 Feb;152(2):228–32.
9. Kang MG, Song WJ, Choi S, et al. Google unveils a glimpse of allergic rhinitis in the real world. *Allergy* 2015;70:124–8.
10. Ram S, Zhang W, Williams M, Pengetnze Y. Predicting asthma-related emergency department visits using big data. *IEEE J Biomed Health Inform*. 2015 Jul;19(4):1216–23.; Bousquet J, O'Hehir RE, Anto JM, et al. Assessment of thunderstorm-induced asthma using Google Trends. *J Allergy Clin Immunol*. 2017 Sep;140(3):891–893.e7.
11. Internet World Stats. 2020. <https://www.internetworldstats.com/stats.htm>.
12. <https://trends.google.com/trends/?geo=HR>
13. Wilde J, Chen W, Lohmann S. COVID-19 and the future of US fertility: what can we learn from Google?. *IZA Discussion Papers*. Oct 2020. Available from: <https://www.econstor.eu/bitstream/10419/227303/1/dp13776.pdf>. Accessed Jan 10th 2021.
14. WHO. Report of the WHO–China Joint Mission on Coronavirus Disease 2019 (COVID–19). <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>. Accessed Jan 10th 2021.
15. Zhang B, Zaman A, Silenzio V, et al. The Relationships of Deteriorating Depression and Anxiety With Longitudinal Behavioral Changes in Google and YouTube Use During COVID–19: Observational Study. *JMIR Ment Health*. 2020 Nov 23;7(11):e24012.
16. <https://www.r-project.org/>
17. Vlada RH. Službena stranica Vlade za pravodobne i točne informacije o koronavirusu. 2020. <https://www.koronavirus.hr/>.
18. Ministarstvo zdravstva RH. <https://zdravlje.gov.hr/>; Civilna zaštita RH. <https://civilna-zastita.gov.hr/>. Accessed Jan 4th 2021.
19. HZJZ. <https://www.hzjz.hr/priopcenja-mediji/koronavirus-najnoviji-podatci/>. Accessed Jan 20th 2021.
20. Mustač D. Demografska obilježja epidemije novog koronavirusa u Hrvatskoj. May 2021. Available from: <https://clps.hr/demografska-obiljezja-epidemije-novog-koronavirusa-u-hrvatskoj>. Accessed Mar 12th 2021.
21. DZS. Natural change in population. 2021. Available from: [https://www.dzs.hr/Hrv/publication/FirstRelease/results.asp?pString=Prirodno%20kreta&pSearchString=%20Prirodno%20kreta%](https://www.dzs.hr/Hrv/publication/FirstRelease/results.asp?pString=Prirodno%20kreta&pSearchString=%20Prirodno%20kreta%20). Accessed Mar 1st 2021.
22. Jurić, T. Monitoring families and COVID-19 with Google Trends: A Croatian case study. 2021 (u objavi).
23. Novi list. U Varaždinskoj županiji od subote strože protuepidemijske mjere. Nov 2020. Available from: <https://www.novolist.hr/novosti/hrvatska/u-varazdinskoj-zupaniji-od-subote-stroze-protuepidemij-ske-mjere/>. Accessed Dec 20th 2020.
24. Jurić T. Google Trends as a method to predict new COVID-19 cases. *MedRxiv*. Mar 2021. Available from: <https://www.medrxiv.org/content/10.1101/2021.03.12.21253452v1>

ADRESA ZA DOPISIVANJE

Doc. dr. sc. Tado Jurić, prof.
Hrvatsko katoličko sveučilište, Odjel za povijest
Ilica 242, 10000 Zagreb
E-mail: tado.juric@unicath.hr
Telefon: + 385 1 3706 628