

Analiza komentara na objave videa na platformi YouTube vezanim uz koronavirus

Severović, Ivan

Master's thesis / Diplomski rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka / Sveučilište u Rijeci**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:195:684076>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-04-25**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Informatics and Digital Technologies - INFORI Repository](#)



Sveučilište u Rijeci - Odjel za informatiku
Diplomski studij Informacijsko komunikacijski sustavi

Ivan Severović

Analiza komentara na objave videa na platformi
YouTube vezanim uz koronavirus

Diplomski rad

Mentor: dr.sc. Ana Meštrović

Rijeka, prosinac 2020.

Rijeka, 3.10.2020.

Zadatak za diplomski rad

Pristupnik: Ivan Severović

Naziv diplomskog rada: Analiza komentara na objave videa na platformi YouTube vezanim uz koronavirus

Naziv diplomskog rada na eng. jeziku: Analysis of YouTube video comments related to coronavirus

Sadržaj zadatka:

Zadatak diplomskog rada je analizirati objave i komentare s društvene mreže YouTube koji su vezani uz koronavirus. Prvo je potrebno prikupiti tekstualne podatke na hrvatsome jeziku s platforme YouTube. Nakon toga treba dati pregled prikupljenih podataka. U drugom dijelu rada potrebno je konstruirati mrežu na temelju komentara i komunikacije na mreži YouTube. Potom treba napraviti analizu mreže na globalnoj, središnjoj i lokalnoj razini.

Mentor:

Izv. prof. dr. sc. Ana Meštrović



Voditelj za diplomske radove:

Izv. prof. dr. sc. Ana Meštrović



Zadatak preuzet: 3.10.2020.



(potpis pristupnika)

Sažetak:

Pandemija koronavirusa SARS-Cov-2, odnosno bolest Covid-19, zahvatila je cijeli svijet, a tako i Republiku Hrvatsku, stoga je cilj ovog diplomskog rada prikupiti i analizirati podatke o objavama video zapisa sa YouTube platforme u tom kontekstu s ključnim riječima poput *korona virus*, *korona mjere*, *korona simptomi* itd. Ključni podaci koji se prikupljaju i analiziraju su naziv videozapisa, autor, datum objave, najrelevantniji komentari, te njihovi autori i datumi objave. Ideja za nastanak ovog rada nastala je zbog interesa o ponašanju pojedinaca na društvenim mrežama u vrijeme vladanja pandemije.

Ključne riječi: SARS-Cov-2, Covid-19, YouTube, YouTube Data API, Python

Sadržaj

1. Uvod	5
2. Prikupljanje podataka	7
2.1. YouTube Data API.....	7
2.2. Podešavanje lokalnog projekta	10
2.3. Pretraga videa po ključnoj riječi	12
2.4. Dohvaćanje komentara videozapisa	15
3. Analiza i vizualizacija podataka.....	16
3.1. Statistika prikupljenih podataka.....	16
3.2. Analiza objava videa i komentara po danima.....	18
3.3. Videozapisi s najviše pregleda, komentara i lajkova	22
4. Analiza mreže	25
5. Zaključak	35
Popis literature i izvora.....	36
Popis priloga	38
Popis slika	39
Popis tablica	40

1. Uvod

COVID-19 (korona) je bolest dišnih puteva koja se prvi put pojavila krajem prosinca 2019. godine u kineskom gradu Wuhanu. U siječnju 2020. se u Kini pretvorila u epidemiju i proširila diljem svijeta. Potaknuo ju je do tada nepoznati koronski virus SARS-CoV-2, a prvi slučaj u Hrvatskoj zabilježen 25. veljače 2020. godine. [1]

Komunikacija putem društvenih mreža danas je zastupljenija više nego ikad prije. Na to su uvelike utjecale mjere poduzete u svrhu usporavanja širenja koronavirusa. Kako bi se ograničio fizički kontakt zatvoreni su noćni klubovi, restorani, kafići, teretane, plesne škole i mnoge druge aktivnosti koje su natjerale ljude da to vrijeme zamijene korištenjem društvenih mreža. Ljudska potreba je iskazati svoje mišljenje, a s obzirom da to više ne mogu razgovorom s bliskom osobom u kafiću uz kavu, to će učiniti putem poruke u Messengeru, u obliku komentara na Facebook objavu nekog portala ili YouTube videa, čija je analiza ujedno i svrha ovog diplomskog rada.

YouTube [2] je najveća svjetska web-platforma za dijeljenje video sadržaja, na koju se svake minute učita više od 500 sati video sadržaja i koja ima preko 1.9 bilijuna mjesečno aktivnih korisnika. Osnovani su je 14. veljače 2005. godine S. Chen, C. Hurley i J. Karim Koristi, no brzo nakon njenog nastanka, u studenom 2006. godine, YouTube je kupljen od strane Google LLC-a, multinacionalne korporacije specijalizirane za mrežne usluge i proizvode. U današnje vrijeme YouTube je jedna od najpoznatijih podružnica Googlea. Koristi se za dijeljenje informacija, podučavanje, zabavu, reklamiranje i još mnogo drugih stvari. Njeno sjedište nalazi se u Kaliforniji, preciznije u gradu San Bruno, no koristi se diljem svijeta, a tako i u Republici Hrvatskoj. YouTube se u hrvatskoj javnosti prvi put u većoj mjeri spominjao krajem 2006. godine, kada je na njega jedan korisnik postavio isječke govora tadašnjeg ministra unutarnjih poslova kako bi ga ismijao, danas YouTube u Hrvatskoj koristi većina internet korisnika [3]. S obzirom na veliki broj korisnika, YouTube sadrži mnoštvo podataka koji se mogu iskoristiti za istraživanje i različite analize. U ovom radu, ekstrakcija komentara vezana uz videozapise o koronavirusu koristiti će se za analizu sentimenta i ostale zadatke računalne analize prirodnog jezika, dok će YouTube API omogućiti podešavanje kriterija pretraživanja (vremenski raspon, ključne riječi i lokaciju).

U prvom dijelu rada opisana je ekstrakcija podataka o videozapisima s profila registriranih u Republici Hrvatskoj, te njihovim komentarima vezanih uz virus SARS-CoV-2 [4]. To uključuje prikupljanje podataka od početka 2020. godine i njihovu pohranu u datoteku .csv formata pomoću programskog jezika Python [5]. Analiza takve interakcije (broj pregleda, lajkova) korisnika s videozapisima i komunikacije (komentari) s kreatorima je iznimno važna za razumijevanje komunikacijske krize uzrokovane koronavirusom. Rezultat analize dat će nam uvide kako je većina hrvatske populacije reagirala na događaje i situacije vezane uz epidemiju, kao što s naprimjer mjere i odluke stožera.

Kako bi provođenje analize bilo moguće, potrebno je postaviti projekt na Google konzoli [6], omogućiti YouTube Data API [7], te kreirati skriptu koja komunicira s YouTube Data API-jem. Nakon uspješne uspostave komunikacije skripte i API-ja, definirat ćemo listu ključnih riječi na temelju kojih će nam API vratiti željene rezultate. Njih ćemo proslijediti API-jevoj funkciji search() te u nju dodati vremenski parametar za datum od kojeg želimo da nam skuplja podatke videozapisa i lokacijski za regiju iz koje ih dohvaća. Limit slanja zahtjeva API-ju je postavljen na 100 po danu, tako da bi prije pokretanja pretrage trebalo dobro pročitati dokumentaciju i podesiti sve parametre.

Gore navedeni parametri se odnose na broj videozapisa koji se dohvaća po ključnoj riječi i broj komentara po videozapisu. Ključna riječ se odnosi na upit, a pretraga daje preciznije rezultate ako se API-ju prosljedi upit 'korona cjepivo' umjesto samo jedna riječ 'cjepivo'. Ukoliko imamo 5 ključnih riječi, možemo za svaku pretragu podesiti da prođe kroz 4 stranice (svaka stranica sadrži podatke o 5 videozapisa) i tako dohvatiti podatke za 20 videozapisa po upitu. Nakon toga ćemo za svaki od tih videozapisa pomoću metode `commentThread()` dohvatiti podatke za 10 najbolje rangiranih komentara po broju lajkova i spremiti ih kao datoteku u .csv formatu.

Drugi dio rada uključuje analizu i vizualizaciju dobivenih podataka. Gore navedene datoteke u .csv formatu koje sadrže sve informacije o komentarima (autor, broj lajkova, datum objave) pojedinog videozapisa, te informacije o broju pregleda, lajkova, autoru videa, naslovu i datumu objave biti će spojene u jedinstveni skup podataka. Potom će se iz tog skupa ukloniti duplikati (postoji mogućnost da je za dvije ključne riječi dohvaćen isti videozapis) i komentari koji nisu na hrvatskom jeziku. Iz pročišćenog skupa podataka bit će stvorena tablica najučestalijih riječi u skupu riječi svih komentara i tablice s informacijama o 10 videozapisa s najviše pregleda, komentara i lajkova. Osim toga, stvorit ćemo grafove koji prikazuju broj objava videozapisa i komentara po danima, te na temelju toga izvlačiti zaključke kako objave medija, uvođenje novih mjera ili rast broja zaraženih utječu na korisnike YouTube mreže tokom cijele godine i drugog vala pandemije.

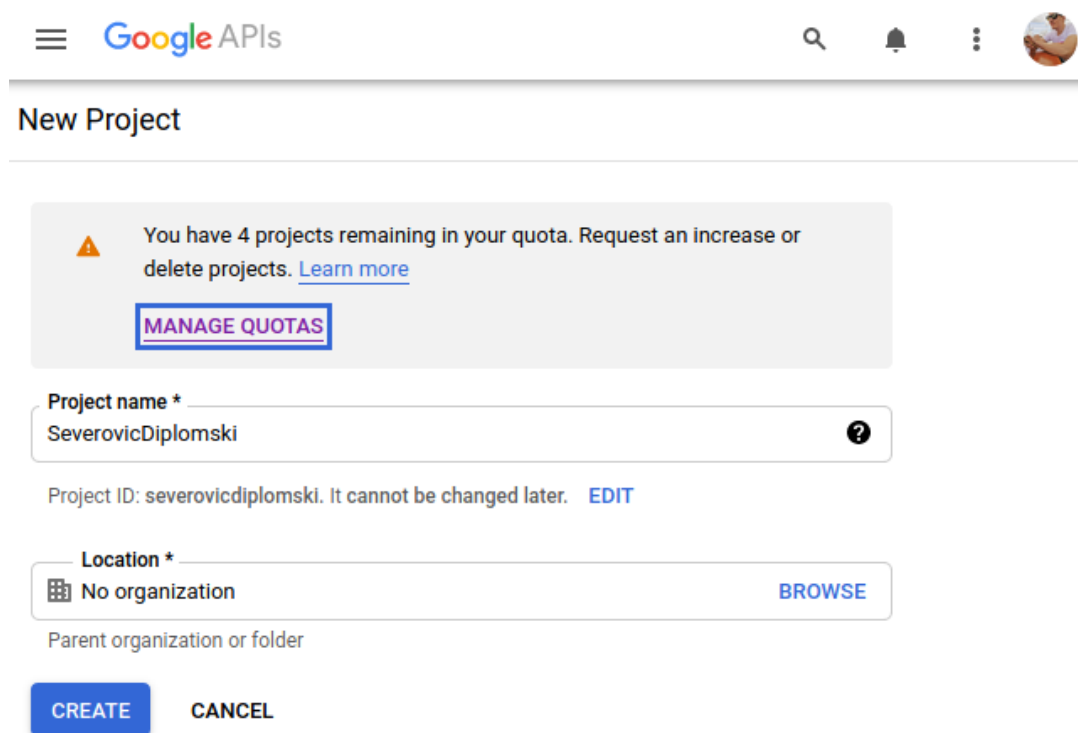
Zadnji dio analize obuhvaćat će pripremu podataka (čvorovi, bridovi i težine) na temelju kojih ćemo kreirati mrežu YTNet1 koja će predstavljati interakciju korisnika tokom cijele pandemije i mrežu YTNet2 koja će je pratiti samo tokom drugog vala od 2.10.2020. Ispitivanjem mjera poput snage čvora dolazit ćemo do informacija o autorima čija su videa najkomentiranija, korisnicima koji ostavljaju najviše komentara, te mnogim drugim globalnim i lokalnim mjerama. Na kraju rada ćemo mreže vizualizirati u Gephiu što će nam dati uvid u izgled i veličinu zajednica na YouTube društvenoj mreži, te saznanja koji YouTube kanali su tokom epidemije uzrokovane koronavirusom bili najpopularniji kod većine hrvatskih korisnika mreže YouTube.

2. Prikupljanje podataka

U ovom poglavlju opisani su postupci prikupljanja podataka o videozapisima i komentarima YouTube videa za riječ 'koronavirus' i slične. Prvi korak je kreiranje projekta na Google konzoli i povezivanje s YouTube Data API servisom preko Python skripte pomoću vjerodajnica (Slika 6).

Nakon toga će nam se otvoriti sve mogućnosti YouTube Data API servisa od pretrage videa po ključnoj riječi i podešavanju raznih parametara kao što su jezik, regija, datum objave i slični. Nakon pretrage videa dohvaćat ćemo podatke kao što su broj pregleda, lajkova, komentara, ime kanala koji je objavio videozapis, datum objave i slični za svaki pronađeni videozapis. Osim videozapisa dohvaćat ćemo i podatke o komentarima koji će se na kraju spremati kao datoteke u .csv formatu, a u drugom dijelu spajati u početni skup podataka za analizu.

2.1. YouTube Data API

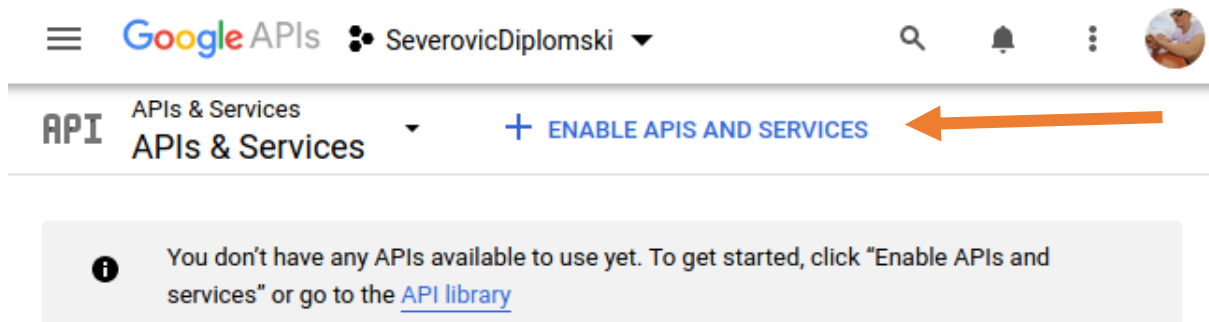


The screenshot shows the 'New Project' page on the Google Cloud console. At the top, there's a navigation bar with the Google APIs logo and search, notification, and menu icons. Below the header, the title 'New Project' is displayed. A warning box indicates that the user has 4 projects remaining in their quota and provides a 'MANAGE QUOTAS' button. The 'Project name' field is filled with 'SeverovicDiplomski'. Below it, the 'Project ID' is shown as 'severovicdiplomski'. The 'Location' is set to 'No organization'. At the bottom, there are 'CREATE' and 'CANCEL' buttons.

Slika 1. Kreiranje projekta na Google konzoli

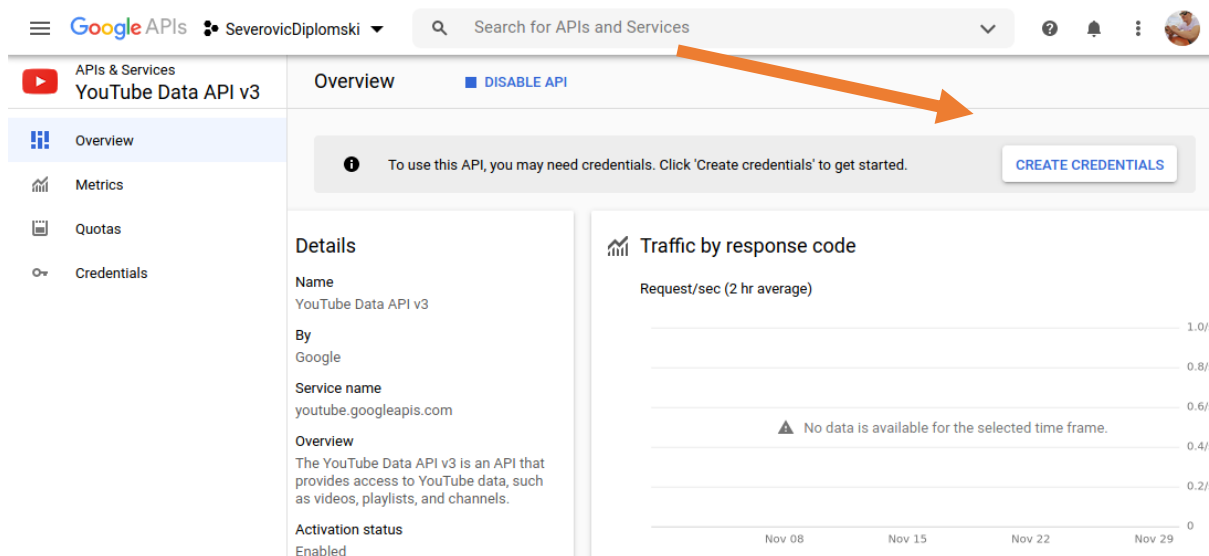
Kako bi pristupili YouTube Data API-ju, potrebno je stvoriti projekt na [Google konzoli](#) [6], kako bi dobili podatke za autorizaciju što nam dozvoljava upućivanje API poziva unutar projekta (Slika 1).

Nakon kreiranja projekta potrebno je omogućiti YouTube Data API, na način da omogućimo API-je i servise (Slika 2).



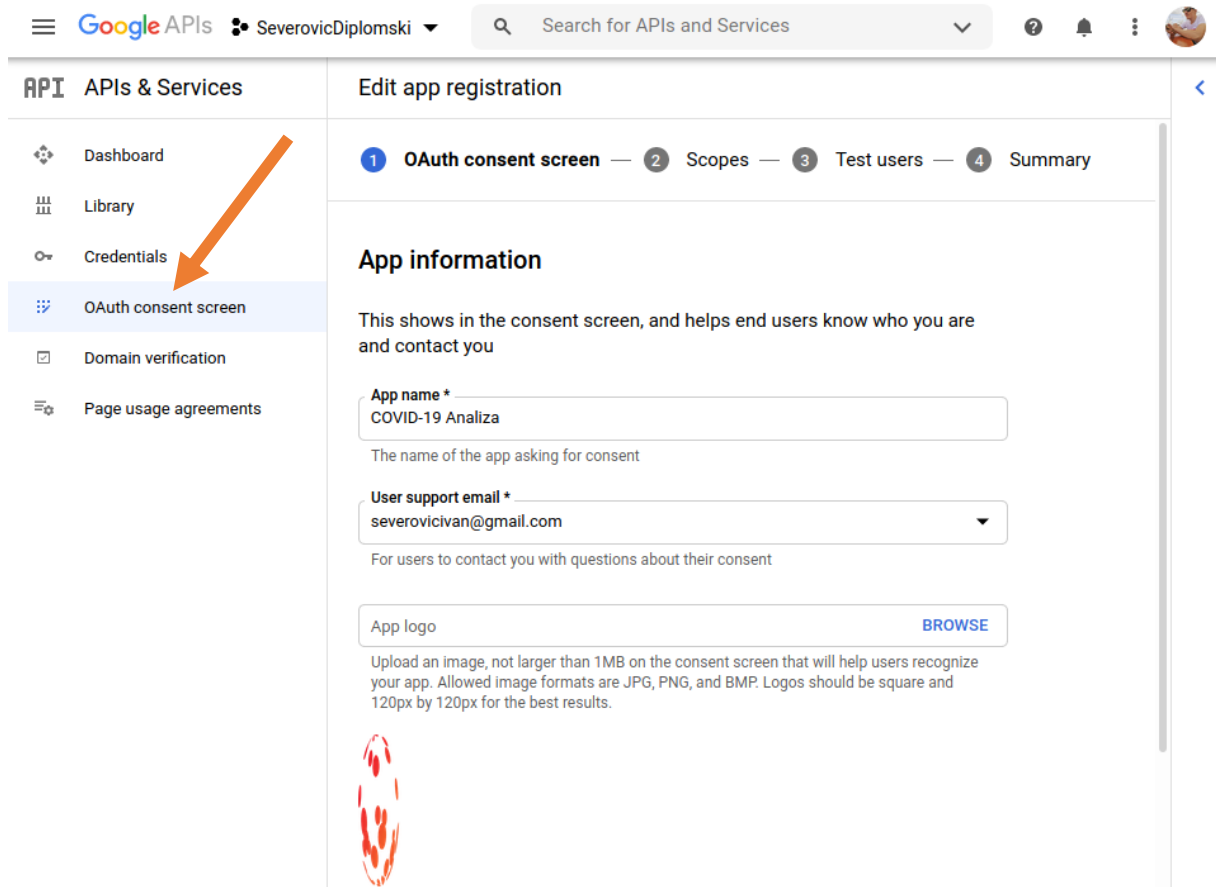
Slika 2 - Odabir API-ja i servisa

Odabirom navedene opcije otvara se pretraživač u koji u unosimo „youtube”, te iz liste rezultata odaberemo **YouTube Data API 3** i omogućimo ga u projektu. Nakon što smo omogućili API, odabirom gumba **Create credentials** ćemo kreirati vjerodajnice kako bi bili u mogućnosti koristiti ga (Slika 3). Vjerodajnica je datoteka koja sadrži ovlaštenje koje je pojedincu izdala treća strana, u ovom slučaju YouTube za korištenje njihovog Data API-ja.



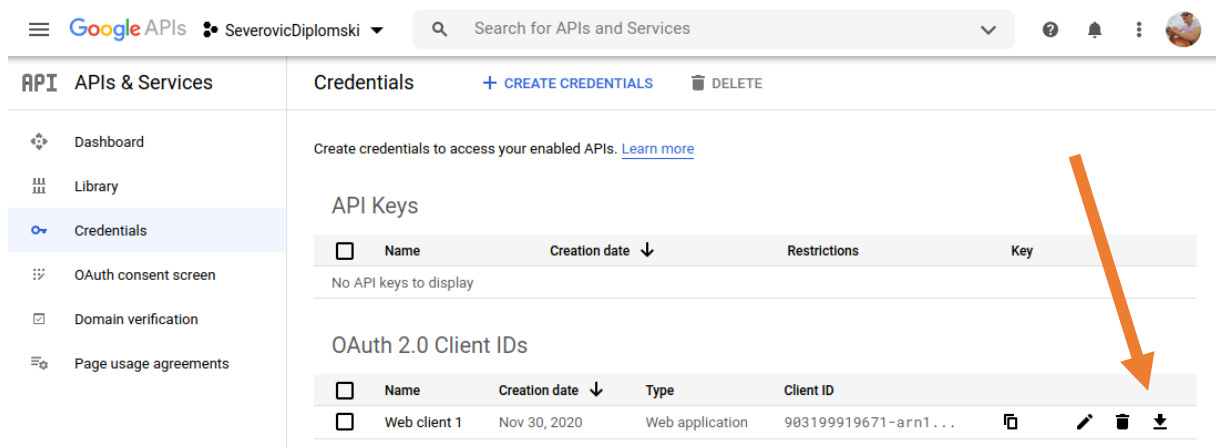
Slika 3. Kreiranje vjerodajnica

Nakon klika na gumb 'Create credentials' otvorit će se nova stranica s proširenim izbornikom s lijeve strane. Tu se treba odabrati opciju **OAuth consent screen** (Slika 4), kliknuti **Create**, popuniti obrazac s odgovarajućim podacima kao što je prikazano na slici te odabirati **Save and continue** sve dok ne dođemo do kraja. Nakon što smo to napravili, u izborniku s lijeve strane odabrat ćemo **Credentials**, potom **+Create credentials** i u padajućem izborniku **OAuth client ID**. Otvorit će nam se obrazac u kojem kao tip aplikacije odaberemo **Web application**, te potom **Create**.



Slika 4. Popunjavanje informacija o aplikaciji

Ovim putem kreirana je JSON datoteka koju ćemo preuzeti (Slika 5) i integrirati ju u naš lokalni projekt čije je podešavanje pojašnjeno u nastavku.



Slika 5. Preuzimanje konfiguracije za povezivanje s API servisom

2.2. Podešavanje lokalnog projekta

Nakon kreiranja glavne skripte, **covid_ekstrakcija.py**, u kojoj ćemo definirati pravila za ekstrakciju komentara, potrebno je instalirati i uključiti **Python** ekstenziju za **Visual Studio Code** kako bi tu skriptu mogli uspješno pokrenuti.

Sljedeći korak je instalacija Google API klijentske biblioteke za Python pomoću naredbe: **pip3 install google-api-python-client** i dodatne biblioteke koje će se baviti provjerom autentičnosti naredbom: **pip3 install google-auth google-auth-oauthlib google-auth-httplib2**. Vjerodajnicu (eng. Credential) koju smo već ranije postavili u repozitorij skupa sa **covid_ekstrakcija.py** skriptom ćemo preimenovati u **client_secret.json** i pohraniti ju u konstantu **CLIENT_SECRETS_FILE** (Slika 6, linija15).

```
covid_ekstrakcija.py > get_videos
5  import google.oauth2.credentials
6
7  from googleapiclient.discovery import build
8  from googleapiclient.errors import HttpError
9  from google_auth_oauthlib.flow import InstalledAppFlow
10 from google.auth.transport.requests import Request
11
12 # The CLIENT_SECRETS_FILE variable specifies the name of a file that contains
13 # the OAuth 2.0 information for this application, including its client_id and
14 # client_secret.
15 CLIENT_SECRETS_FILE = "client_secret_1.json"
16 # This OAuth 2.0 access scope allows for full read/write access to the
17 # authenticated user's account and requires requests to use an SSL connection.
18 SCOPES = ['https://www.googleapis.com/auth/youtube.force-ssl']
19 API_SERVICE_NAME = 'youtube'
20 API_VERSION = 'v3'
21
22 def get_authenticated_service():
23     credentials = None
24     if os.path.exists('token.pickle'):
25         with open('token.pickle', 'rb') as token:
26             credentials = pickle.load(token)
27     # Check if the credentials are invalid or do not exist
28     if not credentials or not credentials.valid:
29         # Check if the credentials have expired
30         if credentials and credentials.expired and credentials.refresh_token:
31             credentials.refresh(Request())
32         else:
33             flow = InstalledAppFlow.from_client_secrets_file(
34                 CLIENT_SECRETS_FILE, SCOPES)
35             credentials = flow.run_console()
36     # Save the credentials for the next run
37     with open('token.pickle', 'wb') as token:
38         pickle.dump(credentials, token)
39     return build(API_SERVICE_NAME, API_VERSION, credentials = credentials)
```

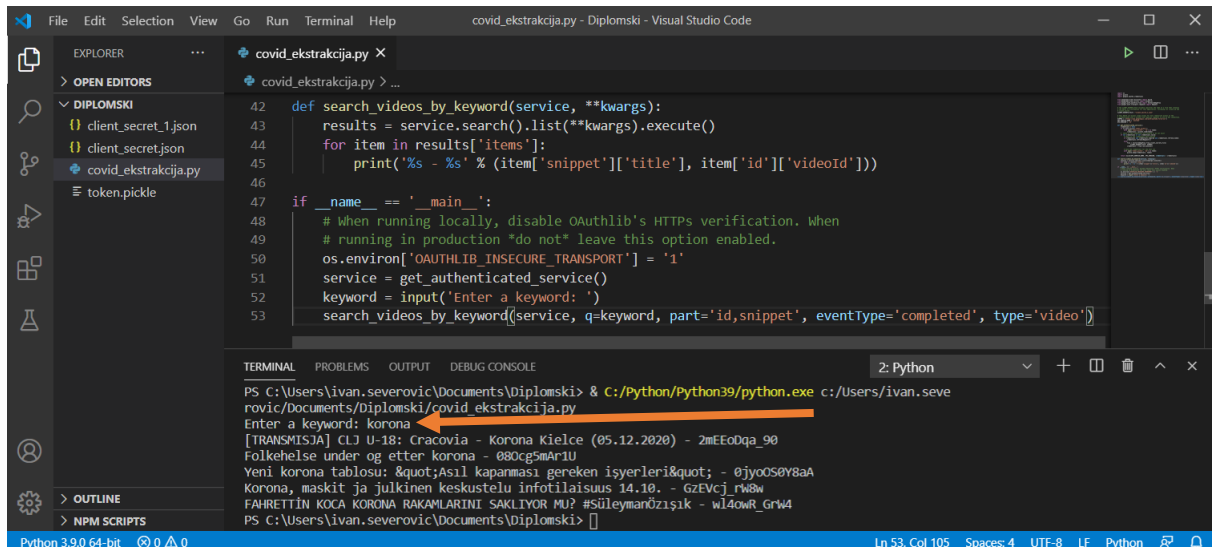
Slika 6. Stvaranje poveznice za autorizaciju

S obzirom da se Google API klijent koristi za pristup svim Google API-ima, opseg ćemo ograničiti samo na YouTube (Slika 6, linija 18). Nakon što smo definirali opseg unutar konstante **SCOPES** možemo izgraditi servis koji će biti odgovoran za interakciju s API-jem. To ćemo napraviti preko funkcije **get_authenticated_service** (Slika 6, linija 22) nakon što joj proslijedimo gore definirane konstante. Nakon pokretanja skripte u terminalu će nam se prikazati autorizacijski URL kojeg ćemo otvoriti u internet pregledniku, te odabrati naš račun za prijavu. Potom je potrebno našoj Python skripti dati dozvole za pregled, uređivanje i brisanje videozapisa, ocjena, komentara i titlova na YouTubeu. Nakon potvrde svih dozvola pojavit će se kod koji je potrebno kopirati i zalijepiti na predviđeno mjesto unutar terminala (**Enter the authorization code**).

Nakon što smo zalijepili kod, skripta se izvela do kraja bez izbacivanja pogreške, što znači da smo pravilno postavili klijenta, te povezali našu skriptu s API-jem. Jedini problem je taj što ćemo cijeli postupak autorizacije morati ponavljati svaki put kada pokrećemo skriptu. Kako bi to izbjegli, izmjenom postojeće funkcije **get_authenticated_service()** predmemorirat ćemo vjerodajnice i pohraniti ih u datoteku **token.pickle** koristeći Pythonov pickle format. Ponovni proces autorizacije će se pokrenuti samo ako datoteka s vjerodajnicama ne postoji ili su vjerodajnice pohranjene u njoj istekle.

2.3. Pretraga videa po ključnoj riječi

Sada napokon možemo početi s našim zadatkom, odnosno dohvaćanjem podataka o YouTube video zapisima koji sadrže ključne riječi poput virus, korona, maske, mjere, covid, itd. Za početak ćemo napraviti testnu funkciju `search_videos_by_keyword()` koja vraća naslove videa ovisno o unesenoj ključnoj riječi. Kao prvi argument imat ćemo funkcijsku varijablu `service` koja nas spaja na API. Ostale argumente dodavati ćemo po želji i potrebi, a bitno je da njihovi nazivi i vrijednosti budu u skladu s [YouTube Data API dokumentacijom](#). Nama najbitniji parametar je `q` koji pohranjuje unesenu ključnu riječ (Slika 7).



The screenshot shows the Visual Studio Code editor with a file named `covid_ekstrakcija.py`. The code defines a function `search_videos_by_keyword` that takes a `service` object and keyword arguments. It uses the `service.search()` method to find videos and prints their titles and IDs. A `__main__` block prompts the user for a keyword and calls the function. The terminal shows the command being executed and the output of the script, which lists several videos related to 'korona'.

```
def search_videos_by_keyword(service, **kwargs):
    results = service.search().list(**kwargs).execute()
    for item in results['items']:
        print('%s - %s' % (item['snippet']['title'], item['id']['videoId']))

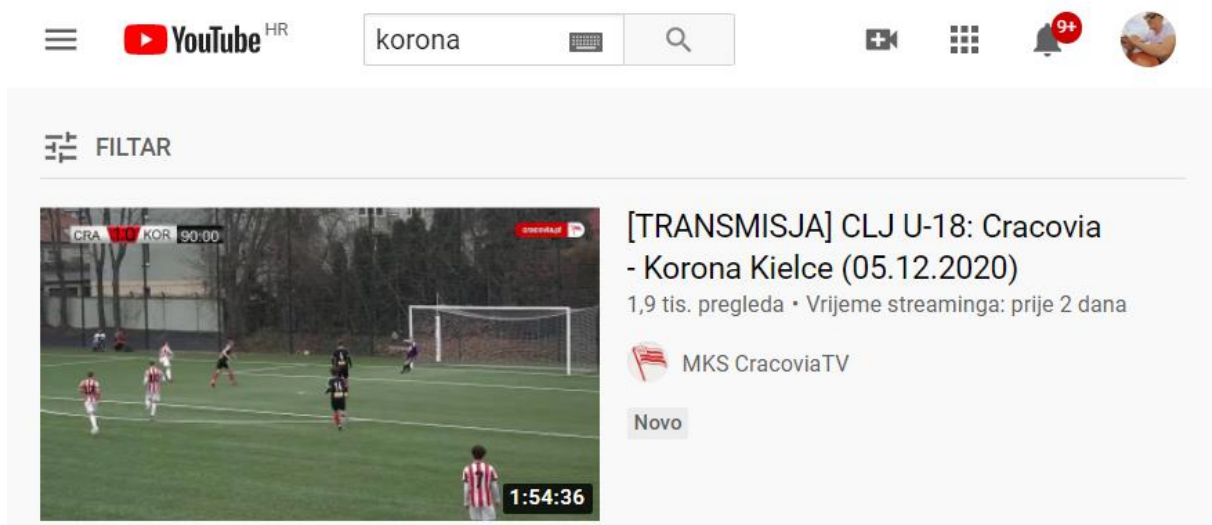
if __name__ == '__main__':
    # When running locally, disable OAuthLib's HTTPS verification. When
    # running in production "do not" leave this option enabled.
    os.environ['OAUTHLIB_INSECURE_TRANSPORT'] = '1'
    service = get_authenticated_service()
    keyword = input('Enter a keyword: ')
    search_videos_by_keyword(service, q=keyword, part='id,snippet', eventType='completed', type='video')
```

Terminal output:

```
PS C:\Users\ivan.severovic\Documents\Diplomski> & C:/Python/Python39/python.exe c:/Users/ivan.seve
rovic/Documents/Diplomski/covid_ekstrakcija.py
Enter a keyword: korona
[TRANSMISJA] CLJ U-18: Cracovia - Korona Kielce (05.12.2020) - 2mEEoDqa_90
Folkehelse under og etter korona - 080cg5mArIU
Yeni korona tablosu: "Asil kapanmasi gereken isyerleri" - 0jyoOS0Y8aA
Korona, maskit ja julkinen keskustelu infotilaisuus 14.10. - GzEvcj_rW8w
FAHRETTİN KOCA KORONA RAKAMLARINI SAKLIYOR MU? #SüleymanÖzişik - w14owR_GrW4
PS C:\Users\ivan.severovic\Documents\Diplomski>
```

Slika 7. Funkcija za pretragu po ključnoj riječi

Kao što vidimo po prvom rezultatu pretrage (Slika 8) nakon što smo upisali ključnu riječ **korona** unutar VS Code terminala (Slika 7) i YouTube web aplikacije, dobili smo isti rezultat, odnosno naslov videa na kojemu je snimka utakmice poljske lige između Cracovie i Korone Kielce.



Slika 8. Rezultat pretrage bez filtra

S obzirom da nas zanimaju samo rezultati vezani uz hrvatski jezik nastali od 1.1.2020. kao argumente kod poziva funkcije dodat ćemo još parametre **relevanceLanguage** i **publishedAfter**. Nakon pokretanja skripte i postavljanja podešavanja filtra u YouTube web aplikaciji u skladu s gore navedenim parametrima vidimo da smo dobili željene rezultate (Slika 9).

The image shows a YouTube search results page for the query 'korona'. The search bar at the top contains the word 'korona'. Below the search bar, there are three video results from Al Jazeera Balkans. The first video is titled 'Korona virus u Hrvatskoj: Vlada donijela nove odluke' (1,5 tis. pregleda • prije 1 dan). The second video is titled 'Vakcina protiv korona virusa na Balkanu ni za godinu' (10 tis. pregleda • prije 4 mjeseca). The third video is titled 'Simptomi blaže prehlade kod mladića zaraženog korona...' (25 tis. pregleda • prije 9 mjeseci). An orange arrow points from the third video title to a terminal window overlay. The terminal window shows the command prompt 'PS C:\Users\ivan.severovic\Documents\Diplomski> & C:/Python/Python39/pytho n.exe c:/Users/ivan.severovic/Documents/Diplomski/covid_ekstrakcija.py' and the output of the script, which lists the same three video titles and their IDs: 'Korona virus u Hrvatskoj: Vlada donijela nove odluke - 06bu1zLJr_o', 'Vakcina protiv korona virusa na Balkanu ni za godinu - 34RYELeDhYE', and 'Simptomi blaže prehlade kod mladića zaraženog korona virusom - UinQVwq1vh8'.

Slika 9. Rezultati pretrage nakon filtriranja

Kao što vidimo na gornjoj slici, rezultati dobiveni pokretanjem skripte su identični onima dobivenim putem korištenjem pretraživača i filtra na YouTube web aplikaciji te stoga

možemo nastaviti s uređivanjem naše skripte za dohvaćanje podataka o videima proširivanjem broja dohvaćenih videa koji je za sad 5 (dohvati samo videa s prve stranice).

U novu funkciju **get_videos()** sam dodao while petlju koja se zaustavlja nakon 10 requestova (varijabla **max_pages**), odnosno, nakon što se preuzmu JSON podaci za 10 video zapisa za datu ključnu riječ. Parametar sadržan u JSON odgovoru svakog dohvaćenog videa, **nextPageToken**, ukazuje nam na to postoji li još stranica s videima ili smo došli do kraja pretrage.

Nakon što smo dobili videa koja se podudaraju s unesenom ključnom riječi, funkciju **search_videos_by_keyword()** nadogradit ćemo tako da osim naslova dohvaća ID videa, naziv kanala i opis, a umjesto ručnog unosa jedne ključne riječi proslijedit ćemo joj listu ključnih riječi.

U međuvremenu je s korištenjem YouTube Data API-ja nastao problem limita koji ograničava broj zahtjeva po danu na 100, tako da sam parametar **max_pages** postavio na 2, što znači da će se dohvatiti 10 videa za jednu ključnu riječ, a u listu **keywords** umetnuo 5 ključnih *riječi korona, lockdown, cjepivo, izolacija i pandemija* što će također dati 50 videa kao rezultat pretrage.

Primijetio sam da parametri **relevanceLanguage** i **regionCode** baš i ne djeluju s obzirom da su za ključnu riječ **lockdown** davali rezultate s lokacija iz cijelog svijeta koji nisu samo na hrvatskom jeziku, odlučio sam funkciji pretrage dodati parametre **location** i **locationRadius**. Nakon pokretanja skripte s novim parametrima dobili smo dobre lokacijske rezultate, ali oni su i dalje bili na engleskom jeziku, te sam iz tog razloga bio primoran ukloniti riječ **lockdown** iz liste ključnih riječi.

U sljedećem poglavlju opisana je pohrana komentara dohvaćenih video zapisa, kako bi oni bili spremni za daljnju analizu.

2.4. Dohvaćanje komentara videozapisa

Za dohvaćanje komentara videa koristiti ćemo `commentThread` parametar. Zbog ograničenja API iz svakog videa dohvaćati će se informacije za 10 komentara s najviše oznaka „sviđa mi se“ (eng. Like), takozvanim lajkovima, a osim broja lajkova još ćemo dohvatiti podatak o autoru komentara i broju odgovora na komentar.

Svaki komentar bit će potrebno dodatno očistiti i od dijakritičkih znakova prije nego što ga pohranimo u okvir podataka (eng. dataframe) koji ćemo spremiti kao datoteku u .csv formatu za svaki videozapis skupa s ostalim informacijama o videu i komentaru. Osim toga konstruirat ćemo i datoteku koja sadrži informacije o svim dohvaćenim video zapisima koji imaju barem jedan komentar (Slika 10). U toj novokreiranoj datoteci koja sadrži sve informacije o preuzetim videima, primjećujemo da komentari često sadrže osjećajnike (eng. emoticon), a neki i nisu na hrvatskom jeziku. Kako bi dobili što čišće tekstualne podatke, osjećajnike ćemo ukloniti uz pomoć biblioteke **demoji**, a informaciju o jeziku komentara ćemo pokušati izvući pomoću biblioteke **langdetect** kako bi u kasnijoj analizi one koji nisu na hrvatskom jeziku lako isključili iz skupa podataka.

Automatsko spremanje

Slika 10. Spojeni skup svih prikupljenih podataka

Nakon što smo očistili komentare od osjećajnika i detektirali jezik komentara te na temelju toga uklonili neželjene redove iz skupa podataka (prazne komentare koji su prethodno sadržavali samo osjećajnike i komentare na stranom jeziku) možemo krenuti sa završnim dijelom rada, odnosno analizom i vizualizacijom podataka.

3. Analiza i vizualizacija podataka

3.1. Statistika prikupljenih podataka

Naziv mjere	Cijela pandemija	Drugi val
Broj videozapisa	83	73
Broj komentara	359	343
Broj riječi	4488	6008
Broj različitih riječi	2420	3349

Tablica 1. Podaci o korpusu

S obzirom da je u ovom trenutku pisanja diplomskog rada broj zaraženih u drugom valu počeo opadati, kako bih napravio što bolju analizu, odlučio sam kreirati još jedan skup podataka i u njega pohraniti podatke o videozapisima kreiranim nakon 2.10.2020. što je bio početak drugog vala. Taj skup i sve stvari vezane za njega sam pohranio u mapu **DrugiVal**,

Tablica 1 prikazuje podatke o broju videozapisa, komentara, broja svih i jedinstvenih riječi (nakon uklanjanja stop riječi) za cijelu pandemiju i drugi val. Prva dva podatka iz tablica će nam biti važna u sljedećem poglavlju kako bi provjerili jesmo li dohvatili sve videozapise i komentare. Na desnoj tablici je približan broj videozapisa i komentara kao na lijevoj jer je dnevni limit API-ja ograničen na dohvaćanje 100 videozapisa, a nakon čišćenja podataka taj broj se još smanji. Kada ne bi bili ograničeni limitom, lijeva tablica u kojoj su pohranjeni podaci od početka pandemije bi sadržavala dosta veće brojeve u odnosu na desnu koja sadrži one od početka drugog vala.

Da bi se poboljšala čitljivost koda, stvorena je skripta **covid_analiza.py** u kojoj ćemo za početak definirati funkciju **frekvencije_rijeci()**. Unutar nje ćemo očistiti datoteku **HrvatskiKomentari.csv** pretvoriti u dataframe i pohraniti ga u varijablu **podaci**. Od svih informacija potrebni su nam komentari tako da ćemo iz cijelog dataframea odabrati samo taj stupac i pretvoriti ga u listu komentara.

Kreirat ćemo praznu listu **lista_rijeci** koju ćemo pomoću **for** petlje popuniti sa svim riječima svih komentara. Riječi unutar komentara smo odvojili funkcijom **split()**. Za svaku riječ potrebno je odrediti broj njenog pojavljivanja unutar liste riječi. To ćemo napraviti pomoću funkcije **count()** i pohraniti rezultate u listu **frekvencije**.

Od liste riječi i liste frekvencija kreirat ćemo listu uređenih dvojki pomoću funkcije **zip()**. S obzirom da ona sadrži i duplikate, njih ćemo eliminirati funkcijom **set()** tako da će nam u listi ostati samo različite riječi kao **ključ** i njen broj pojavljivanja u skupu svih komentara kao **vrijednost**.

Nakon sortiranja liste pronađeno je **2215** jedinstvenih riječi, a kao one s najviše ponavljanja dobili smo veznike, priloge i prijedloge što je bilo i očekivano. Njih smo eliminirali pomoću biblioteke **text_hr** i njene funkcije **get_all_std_words()** tako da smo prije dodavanja riječi komentara u varijablu **lista_rijeci** prvo provjerili nalazi li se riječ u već ranije definiranoj listi stop riječi (eng. stop words). Nakon ponovnog sortiranja i pohrane, broj jedinstvenih riječi je s **2215** pao na **1440**.

S obzirom da nam je prikupljeni skup podataka dosta mali, a u listi stop riječi biblioteke **text_hr** nema zamjenica, rezultati nam nisu savršeni. Hrvatski jezik nema toliko razvijene alate za analizu sentimenta kao engleski tako da skripta trenutno nije automatizirana za izvlačenje mišljenja (eng. opinion mining) o određenom videu na temelju komentara. Nakon ručnog uklanjanja najfrekventnijih zamjenica došli smo do konačnog skupa od **1420** riječi. Na Tablici 2 su prikazane najfrekventnije riječi.

Rijec	Frekvencija
bravo	28
ivane	21
ljudi	20
cast	20
pernar	18
hvala	18
treba	14
bog	12
pozdrav	11
narod	11
jedan	10
moze	9
hdz	9
vrijeme	9
dalje	9
pravu	8
virus	8
puno	8
imas	7

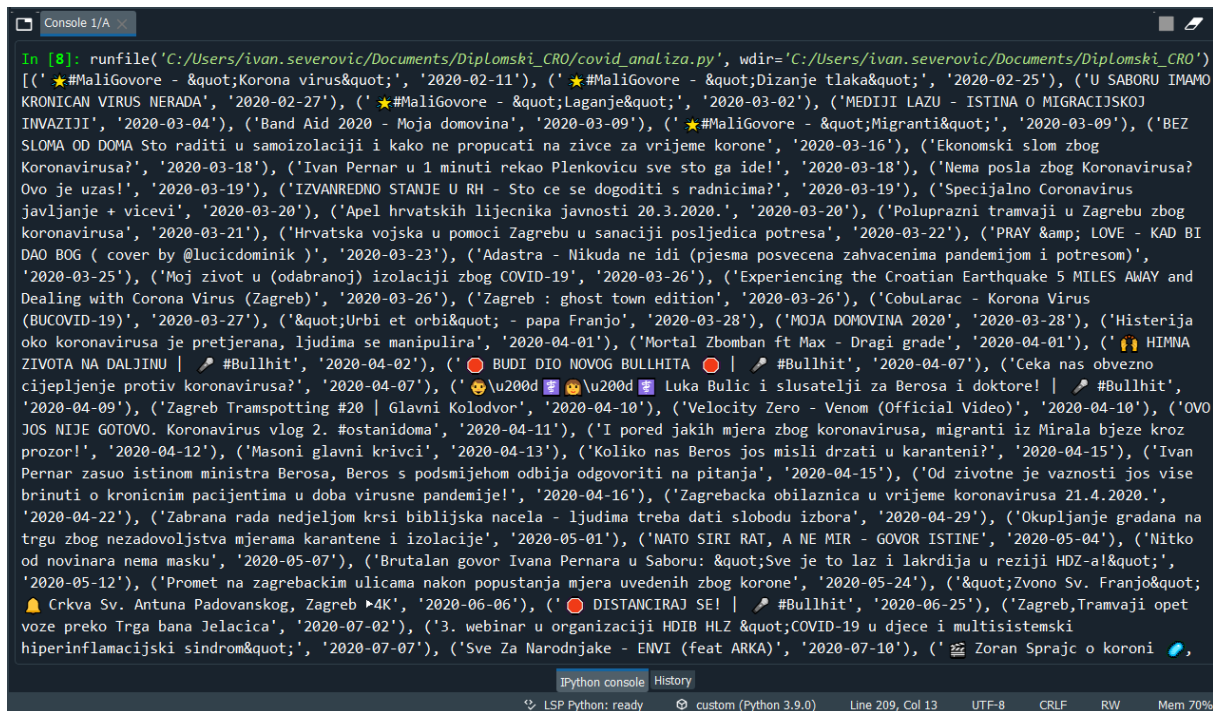
Rijec	Frekvencija
ljudi	28
bravo	22
cast	18
treba	16
hvala	15
moze	15
puno	12
jako	12
narod	12
draga	11
hdz	11
bog	11
dobro	10
dalje	10
mogu	9
plenkovic	9
sad	9
pozdrav	9
nisu	9

Tablica 2. Top 20 najfrekventnijih riječi (cijela pandemija i drugi val)

Analizirajući sentiment videa odokativno, uzimajući u obzir komentare korisnika te veliki broja pregleda i lajkova u kratkom vremenu, uvidio sam da je uvođenjem strožijih mjera počela rasti popularnost kanala koji ukazuju na besmisao mjera glazbenim video spotovima poput [Antene Zagreb](#), onih koji su pristalice ležernijih mjera kakve su u Švedskoj i ukazuju na manipulaciju medija kao što je [MatrixTV](#) ili čak onih koji vjeruju da je ova pandemija teorija zavjere poput [IvanPernarTV](#).

3.2. Analiza objava videa i komentara po danima

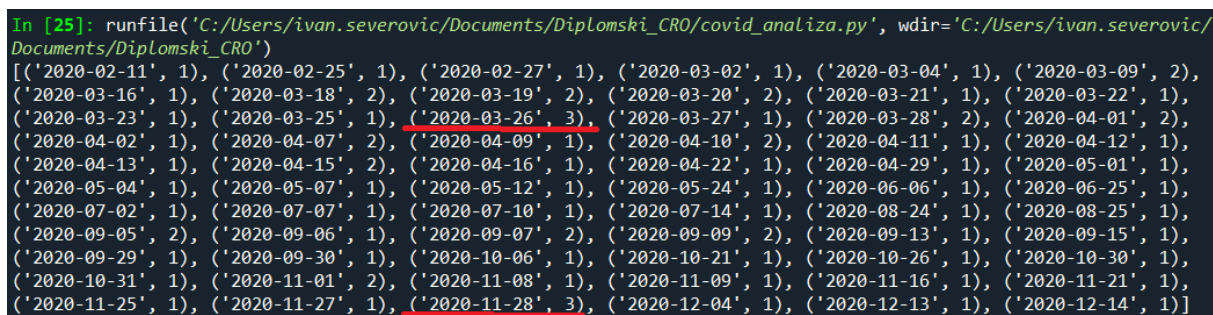
Za ovo poglavlje kao početne skupove podataka koristio sam dokumente **HrvatskiKomentari.csv** iz mape Pandemija i DrugiVal. Kod prikaza objave videa po danu, ključna su nam 2 stupca iz skupa podataka, a to su **VideoTitle** i **VideoPublished**. Nakon što smo iz ta dva stupca stvorili uređene parove i zadržali samo one jedinstvene, njihovim prebrojavanjem došli smo do brojke od **83** videa za cijelu pandemiju i **73** za drugi val. Na isti smo način sa stupcima **Comment** i **CommentPublished** došli do potvrde da je broj komentara jednak broju redova iz dva skupa podataka, odnosno **359** i **343**. Dobivena lista parova (video, datum objave) za cijelu pandemiju prikazana je na slici 11.



```
In [8]: runfile('C:/Users/ivan.severovic/Documents/Diplomski_CRO/covid_analiza.py', wdir='C:/Users/ivan.severovic/Documents/Diplomski_CRO')
[('★#MaliGovore - "Korona virus"', '2020-02-11'), ('★#MaliGovore - "Dizanje tlaka"', '2020-02-25'), ('U SABORU IMAMO KRONICAN VIRUS NERADA', '2020-02-27'), ('★#MaliGovore - "Laganje"', '2020-03-02'), ('MEDIJI LAZU - ISTINA O MIGRACIJSKOJ INVAZIJI', '2020-03-04'), ('Band Aid 2020 - Moja domovina', '2020-03-09'), ('★#MaliGovore - "Migranti"', '2020-03-09'), ('BEZ SLOMA OD DOMA Sto raditi u samoizolaciji i kako ne propucati na zivce za vrijeme korone', '2020-03-16'), ('Ekonomski slom zbog Koronavirusa?', '2020-03-18'), ('Ivan Pernar u 1 minuti rekao Plenkovicu sve sto ga ide!', '2020-03-18'), ('Nema posla zbog Koronavirusa? Ovo je uzas!', '2020-03-19'), ('IZVANREDNO STANJE U RH - Sto ce se dogoditi s radnicima?', '2020-03-19'), ('Specijalno Coronavirus javljanje + vicevi', '2020-03-20'), ('Apel hrvatskih lijebnika javnosti 20.3.2020.', '2020-03-20'), ('Poluprazni tramvaji u Zagrebu zbog koronavirusa', '2020-03-21'), ('Hrvatska vojska u pomoci Zagrebu u sanaciji posljedica potresa', '2020-03-22'), ('PRAY & LOVE - KAD BI DAO BOG ( cover by @lucicdominik )', '2020-03-23'), ('Adastra - Nikuda ne idi (pjesma posvecena zahvacenima pandemijom i potresom)', '2020-03-25'), ('Moj zivot u (odabranoj) izolaciji zbog COVID-19', '2020-03-26'), ('Experiencing the Croatian Earthquake 5 MILES AWAY and Dealing with Corona Virus (Zagreb)', '2020-03-26'), ('Zagreb : ghost town edition', '2020-03-26'), ('CobuLarac - Korona Virus (BUCOVID-19)', '2020-03-27'), ('"Urb et orbi" - papa Franjo', '2020-03-28'), ('MOJA DOMOVINA 2020', '2020-03-28'), ('Histerija oko koronavirusa je pretjerana, ljudima se manipulira', '2020-04-01'), ('Mortal Zbombar ft Max - Dragi grade', '2020-04-01'), ('HIMNA ZIVOTA NA DALJINU | #Bullhit', '2020-04-02'), ('BUDI DIO NOVOG BULLHITA | #Bullhit', '2020-04-07'), ('Čeka nas obvezno cijepljenje protiv koronavirusa?', '2020-04-07'), ('u200d 🇺🇦 Luka Bulic i slusatelji za Berosa i doktore! | #Bullhit', '2020-04-09'), ('Zagreb Tramspotting #20 | Glavni Kolodvor', '2020-04-10'), ('Velocity Zero - Venom (Official Video)', '2020-04-10'), ('OVO JOS NIJE GOTOVO. Koronavirus vlog 2. #ostanidoma', '2020-04-11'), ('I pored jakih mjera zbog koronavirusa, migranti iz Mirala bjeze kroz prozor!', '2020-04-12'), ('Masoni glavni krivci', '2020-04-13'), ('Koliko nas Beros jos misli drzati u karanteni?', '2020-04-15'), ('Ivan Pernar zasuo istinom ministra Berosa, Beros s podsmijehom odbija odgovoriti na pitanja', '2020-04-15'), ('Od zivotne je vaznosti jos vise brinuti o kroninim pacijentima u doba virusne pandemije!', '2020-04-16'), ('Zagrebacka obilaznica u vrijeme koronavirusa 21.4.2020.', '2020-04-22'), ('Zabrana rada nedjeljom krši biblijska nacela - ljudima treba dati slobodu izbora', '2020-04-29'), ('Okupljanje gradana na trgu zbog nezadovoljstva mjerama karantene i izolacije', '2020-05-01'), ('NATO SIRI RAT, A NE MIR - GOVOR ISTINE', '2020-05-04'), ('Nitko od novinara nema masku', '2020-05-07'), ('Brutalan govor Ivana Pernara u Saboru: "Sve je to laz i lakrdija u reziji HDZ-a"', '2020-05-12'), ('Promet na zagrebackim ulicama nakon popustanja mjera uvedenih zbog korone', '2020-05-24'), ('"Zvono Sv. Franjo" 🛎 Crkva Sv. Antuna Padovanskog, Zagreb ▶4K', '2020-06-06'), ('DISTANCIRAJ SE! | #Bullhit', '2020-06-25'), ('Zagreb, Tramvaji opet voze preko Trga bana Jelacica', '2020-07-02'), ('3. webinar u organizaciji HDIB HLZ "COVID-19 u djece i multisistemske hiperinflamacijski sindrom"', '2020-07-07'), ('Sve Za Narodnjake - ENVI (feat ARKA)', '2020-07-10'), ('Zoran Sprajc o koroni 🇺🇦
```

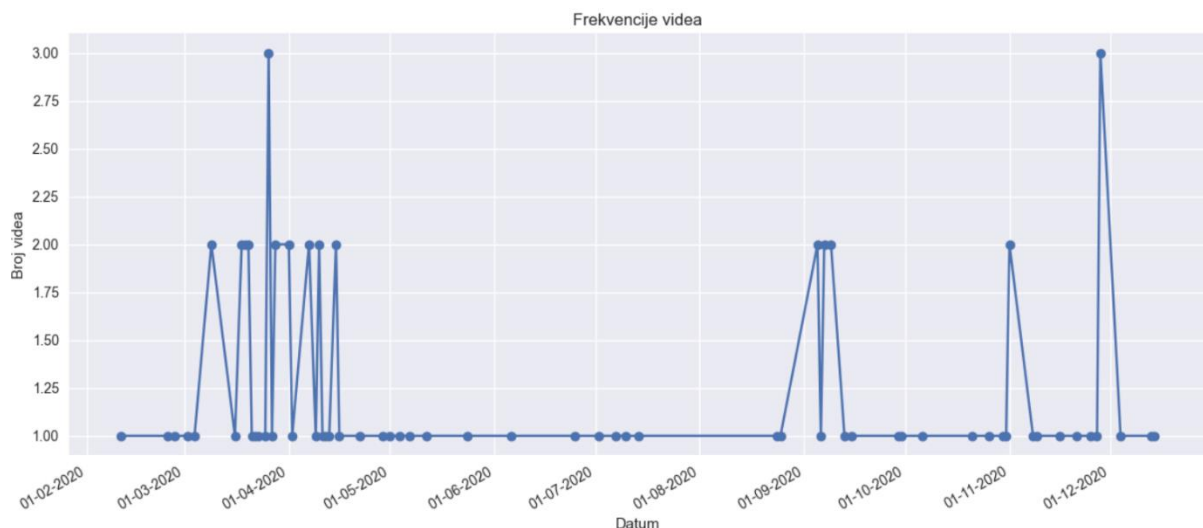
Slika 11. Lista uređenih parova (video, datum objave)

Nakon kreiranja liste parova datuma i naslova videa, kreirana je lista **frekvencije** koja se sastoji od parova (datum, broj objavljenih videa na taj datum). Sa slike 12 i slike 13 vidimo da je 26.3.2020. i 28.11.2020. objavljeno najviše videozapisa. Prvi datum možemo povezati s reakcijama na potres i uvođenje mjere o zabrani napuštanja mjesta prebivališta i stalnog boravka u Republici Hrvatskoj, a drugi s ponovnim uvođenjem mjera, odnosno zatvaranja kafića, restorana i teretana.



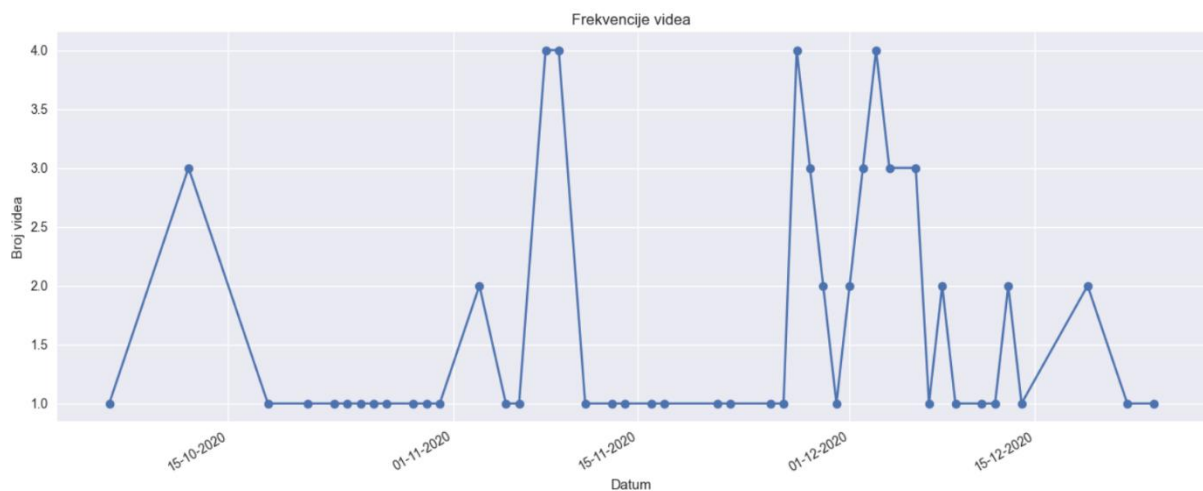
```
In [25]: runfile('C:/Users/ivan.severovic/Documents/Diplomski_CRO/covid_analiza.py', wdir='C:/Users/ivan.severovic/Documents/Diplomski_CRO')
[('2020-02-11', 1), ('2020-02-25', 1), ('2020-02-27', 1), ('2020-03-02', 1), ('2020-03-04', 1), ('2020-03-09', 2), ('2020-03-16', 1), ('2020-03-18', 2), ('2020-03-19', 2), ('2020-03-20', 2), ('2020-03-21', 1), ('2020-03-22', 1), ('2020-03-23', 1), ('2020-03-25', 1), ('2020-03-26', 3), ('2020-03-27', 1), ('2020-03-28', 2), ('2020-04-02', 1), ('2020-04-07', 2), ('2020-04-09', 1), ('2020-04-10', 2), ('2020-04-11', 1), ('2020-04-12', 1), ('2020-04-13', 1), ('2020-04-15', 2), ('2020-04-16', 1), ('2020-04-22', 1), ('2020-04-29', 1), ('2020-05-01', 1), ('2020-05-04', 1), ('2020-05-07', 1), ('2020-05-12', 1), ('2020-05-24', 1), ('2020-06-06', 1), ('2020-06-25', 1), ('2020-07-02', 1), ('2020-07-07', 1), ('2020-07-10', 1), ('2020-07-14', 1), ('2020-08-24', 1), ('2020-08-25', 1), ('2020-09-05', 2), ('2020-09-06', 1), ('2020-09-07', 2), ('2020-09-09', 2), ('2020-09-13', 1), ('2020-09-15', 1), ('2020-09-29', 1), ('2020-09-30', 1), ('2020-10-06', 1), ('2020-10-21', 1), ('2020-10-26', 1), ('2020-10-30', 1), ('2020-10-31', 1), ('2020-11-01', 2), ('2020-11-08', 1), ('2020-11-09', 1), ('2020-11-16', 1), ('2020-11-21', 1), ('2020-11-25', 1), ('2020-11-27', 1), ('2020-11-28', 3), ('2020-12-04', 1), ('2020-12-13', 1), ('2020-12-14', 1)]
```

Slika 12. Lista učestalosti objava videozapisa po danima



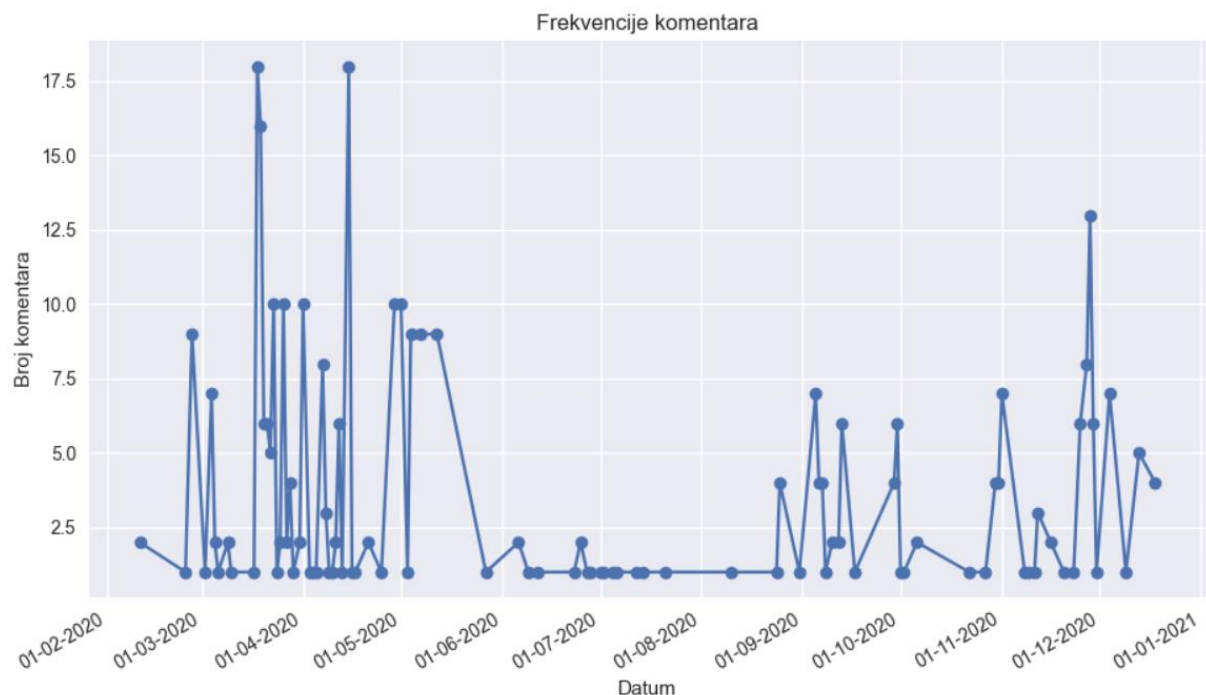
Slika 13. Broj objavljenih videa po danima tokom cijele pandemije

Analizom skupa podataka drugog vala, dolazimo do otkrivanja još jednog događaja koji je bio društveno popraćen, a iz prvog skupa nam nije bio vidljiv. Radi se o prosvjedu protiv mjera stožera u Zagrebu 8.11.2020 (Slika 14). Drugi skok je vidljiv krajem studenog kao i na prethodnom grafu, a onaj nedugo nakon toga odnosi se na rasprave u saboru vezano za pooštavanje mjera.



Slika 14. Broj objavljenih videa po danima u drugom valu

Slika 15 koja prikazuje broj objavljenih komentara po danima u 2020. godini govori nam kako na broj komentara najviše utječu javni mediji i pooštavanje mjera, a ne broj zaraženih (Slika 16). Za drugi val to i ne možemo sa tolikom sigurnošću potvrditi, tako da ćemo za taj skup podataka pearsonovom metodom [8] provjeriti postoji li korelacija između broja komentara i broja zaraženih. Na mrežnim stranicama koronavirusa [4] mogu se pronaći podaci o dnevnom broju zaraženih pohranjene u JSON formatu. Nama su od interesa oni s datumom od 2.10.2020 do 25.12.2020. Njih ćemo ubaciti u Python skriptu, te pohraniti u obliku dataframe-a sa stupcima Broj zaraženih i Datum. Već ranije je stvoren dataframe s datumima i brojem komentara iz drugog vala. Njih ćemo spojiti po datumu, a nakon obacivanja redova za datume bez komentara nastaje skup podataka (Slika 17).



Slika 15. Broj objavljenih komentara po danima



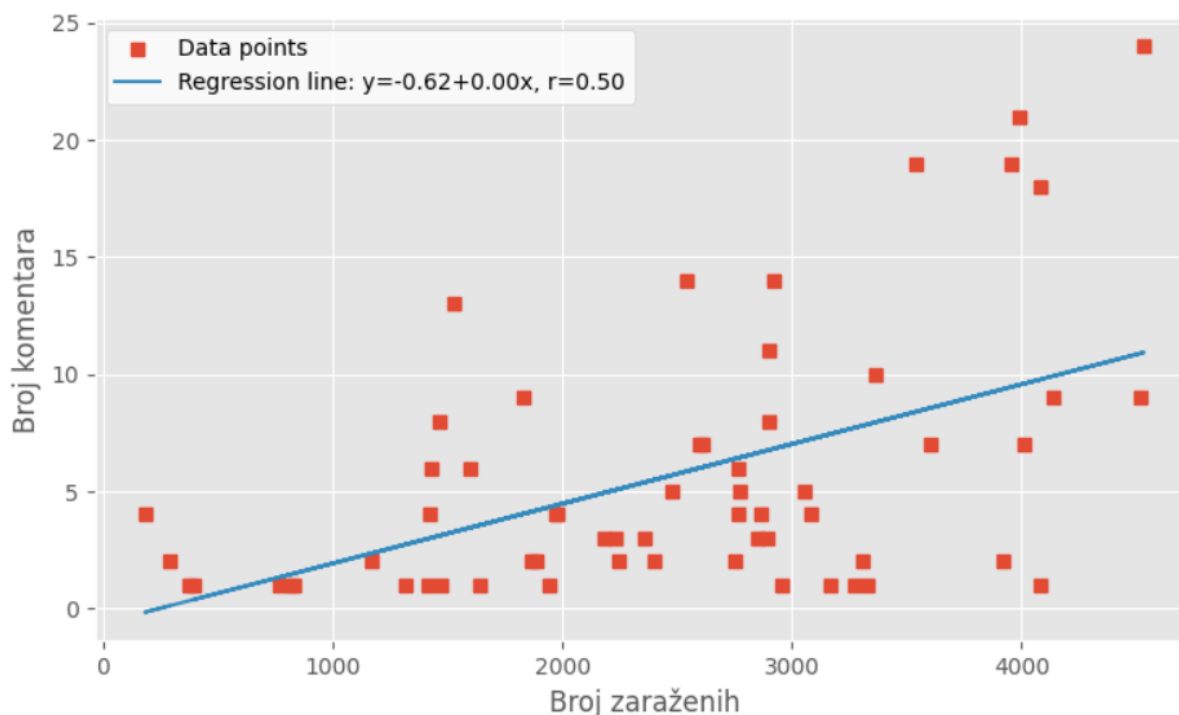
Slika 16. Broj zaraženih koronavirusom po danu

Kako bi dobili pearsonovu korelaciju potrebno je u Python skriptu uključiti biblioteku **scipy.stats** i njenu funkciju **pearsonr**. Ona kao argumente prima dvije varijable, a kao rezultat vraća koeficijent povezanosti. U našem slučaju varijable su stupci Slučajevi i Broj komentara sa slike 17 pretvoreni u liste, a izračunom smo došli do saznanja da je koeficijent korelacije **0.504**. Kako bi si to i predočili, prikazat ćemo taj odnos pomoću grafa. Na osi x ćemo postaviti vrijednosti broja zaraženih, a na y broj komentara (Slika 18).

Datum	Slucajevi	Broj komentara	Komentar
2020-12-04	3955	19	['E moj Grmoja nema te na svijetu.Da te nema Sabor bi bio prazan.Ima divnih ljudi ,a te iskre...
2020-12-03	4534	24	['Steta sto ga nema u Prime Time-u, odlican je Zvonko.', 'Tko Reinera snabdijeva zastitnim ma...
2020-12-02	3539	19	['Ustaj Matija Gubec, ustaj vise iz tog groba.', 'Dace rijec ili nece dati rijec. Moj rajner...
2020-12-01	2900	11	['svi kao nesto protiv a kad je glasanje onda su svi za, isto ko sto je bilo za davanje ovlas...
2020-11-30	1830	9	['to je manje nego je i jedan hdz ovac ukraolovo je zalosno i tuzno', 'Sramota.... Sramota', ...
2020-11-29	2919	14	['Lijepi moj Zagreb, koliko god ne volim onu guzvu u ovo vrijeme na ulicama Zagreba a zapravo...
2020-11-28	3987	21	['Zagreb je zimu prelep. Pozdrav iz Novog Sada, i dodi kad budes mogao da snimis nove autobus...
2020-11-27	4080	18	['Sotona se ponasa lijepo prema covjeku tako da ga zavede. Treba biti mudar i neda ti se zave...
2020-11-26	4009	7	['Bravo Gospodine Bulj zacepите usta toj partiznskoj unuki koja nam stavlja sol narane svojim ...
2020-11-25	3603	7	['Plenkovicu, ako si ti krscanin, onda sam ja niskopodni tramvaj.', 'Zombi si ti Plenkovicu! ...
2020-11-23	1973	4	['264 na tjedan? Znam da se mora prilagoditi za rast populacije, ali apsolutno je to manje ne...
2020-11-22	3308	2	['Tako je mi smo kraljevi kceri i sinovi vrjedni smo', 'Draga Loredana,sve je istina sto pro...
2020-11-20	2958	1	['Vama je u cilju da stavite ljude u kavez kao majmune zbog obicne gripe']
2020-11-19	3164	1	['Slazem se jer njima je u cilju staviti ljude u kavez']
2020-11-17	1945	1	['Moze titl na hrvatski']

Slika 17. Broj komentara i zaraženih po danima

Usporedbom skupa podataka i grafa iz drugog vala vidimo da u njemu ipak postoji veza između broja zaraženih i aktivnosti korisnika na društvenoj mreži YouTube. Što je više zaraženih korisnici su aktivniji. To se najviše vidi za podatak od 3.12.2020, dok bih skok u komentarima od 28.11.2020 ipak povezao s aktivnošću vezanom za uvođenje mjera zabrane rada ugostiteljskim objektima, sportskim objektima itd.



Slika 18. Utjecaj broja zaraženih na aktivnost YouTube korisnika

3.3. Videozapisi s najviše pregleda, komentara i lajkova

Među 10 videozapisa s najviše pregleda, lajkova i komentara uvjerljivo prednjače kanali IvanPernarTV i Antena Zagreb (Slika 19,20,21).

popular - DataFrame

Index	Channel	Video Title	Video Published	Video Views
279	Antena Zagreb	👤👤 Luka Bulic i slusatelji za Berosa i doktore! 🗣️ #Bullhit	2020-04-09 18:49:13	120747
59	Antena Zagreb	🔴 DISTANCIRAJ SE! 🗣️ #Bullhit	2020-06-25 18:00:12	97486
9	IvanPernarTV	MEDIJI LAZU - ISTINA O MIGRACIJSKOJ INVAZIJI	2020-03-04 11:25:06	96236
146	IvanPernarTV	Histerija oko koronavirusa je pretjerana, ljudima se manipulira	2020-04-01 09:17:02	87972
31	IvanPernarTV	Ekonomski slom zbog Koronavirusa?	2020-03-18 11:01:00	84824
271	Antena Zagreb	🗣️ HIMNA ZIVOTA NA DALJINU 🗣️ #Bullhit	2020-04-02 18:19:36	67507
320	IvanPernarTV	Ivan Pernar zasuo istinom ministra Berosa, Beros s podsmijehom odbija odgovoriti na pitanja	2020-04-15 10:08:15	58067
221	IvanPernarTV	Ivan Pernar u 1 minuti rekao Plenkovicu sve sto ga ide!	2020-03-18 12:42:11	49020
15	IvanPernarTV	Ceka nas obvezno cijepljenje protiv koronavirusa?	2020-04-07 16:35:37	47667
334	Most	GRMOJA ISTINOM UZDRMAO ANDREJA PLENKOVIKA!!!!	2020-11-25 18:49:02	45853

Format Resize ☐ Background color ☐ Column min/max Save and Close Close

Slika 19. Videozapisi s najviše pregleda

Pretraga videozapisa je uzimala u obzir sve video objave od početka 2020. godine tako da oni noviji nisu pri vrhu jer su imali manje vremena za skupiti veći broj pregleda, lajkova i komentara. Bez obzira na to, u listu videozapisa s najviše pregleda (Slika 19) je uspio ući i kanal [Most](#). Nakon pregleda videozapisa shvatio sam da je to zbog izbjegavanja odgovora predsjednika vlade na pitanje hoće li ministar gospodarstva i održivog razvoja biti odgovoran za štetne posljedice mjera kao što su zatvaranje restorana, kafića, plesnih škola, teretana itd. Na listi videozapisa s najviše komentara je samo kanal IvanPernarTV koji na najkomentiranijem videozapisu ima čak 784 komentara (Slika 21).

likes - DataFrame

Index	Channel	Video Title	Video Published	Video Likes
275	Antena Zagreb	👤👤 Luka Bulic i slusatelji za Berosa i doktore! ...	2020-04-09 18:49:13	2752
0	IvanPernarTV	MEDIJI LAZU - ISTINA O MIGRACIJSKOJ INVAZIJI	2020-03-04 11:25:06	2673
142	IvanPernarTV	Histerija oko koronavirusa je pretjerana, ljudima se manipulira	2020-04-01 09:17:02	2347
30	IvanPernarTV	Ekonomski slom zbog Koronavirusa?	2020-03-18 11:01:00	2159
56	Antena Zagreb	🔴 DISTANCIRAJ SE! 🗣️ #Bullhit	2020-06-25 18:00:12	1779
287	IvanPernarTV	Brutalan govor Ivana Pernara u Saboru: "Sve je to laz i lakrdija u reziji HDZ-a!"	2020-05-12 10:09:11	1707
248	IvanPernarTV	Okupljanje gradana na trgu zbog nezadovoljstva mjerama karantene i izolacije	2020-05-01 14:00:12	1493
271	Antena Zagreb	🗣️ HIMNA ZIVOTA NA DALJINU 🗣️ #Bullhit	2020-04-02 18:19:36	1372
15	IvanPernarTV	Ceka nas obvezno cijepljenje protiv koronavirusa?	2020-04-07 16:35:37	1279
221	IvanPernarTV	Ivan Pernar u 1 minuti rekao Plenkovicu sve sto ga ide!	2020-03-18 12:42:11	1271

Format Resize ☐ Background color ☐ Column min/max Save and Close Close

Slika 20. Videozapisi s najviše lajkova

comments - DataFrame

Index	Channel	Video Title	Video Published	Video Comments
258	IvanPernarTV	Zabrana rada nedjeljom krši biblijska načela - ljudima treba dati slobodu izbora	2020-04-29 09:39:48	784
248	IvanPernarTV	Okupljanje građana na trgu zbog nezadovoljstva mjerama karantene i izolacije	2020-05-01 14:00:12	684
142	IvanPernarTV	Histerija oko koronavirusa je pretjerana, ljudima se manipulira	2020-04-01 09:17:02	648
15	IvanPernarTV	Ceka nas obvezno cijepljenje protiv koronavirusa?	2020-04-07 16:35:37	606
30	IvanPernarTV	Ekonomski slom zbog Koronavirusa?	2020-03-18 11:01:00	557
0	IvanPernarTV	MEDIJI LAZU - ISTINA O MIGRACIJSKOJ INVAZIJI	2020-03-04 11:25:06	506
287	IvanPernarTV	Brutalan govor Ivana Pernara u Saboru: "Sve je to laz i lakrdija u reziji HDZ-a!"	2020-05-12 10:09:11	503
320	IvanPernarTV	Ivan Pernar zasuo istinom ministra Berosa, Beros s podsmijehom odbija odgovoriti na pitanja	2020-04-15 10:08:15	408
48	IvanPernarTV	Nema posla zbog Koronavirusa? Ovo je užas!	2020-03-19 15:28:11	400
86	IvanPernarTV	IZVANREDNO STANJE U RH - Sto će se dogoditi s radnicima?	2020-03-19 08:00:10	341

Format Resize Background color Column min/max Save and Close Close

Slika 21. Videozapisi s najviše komentara

Pri analizi skupa podataka prikupljenih za vrijeme trajanja drugog vala pandemije uvidio sam da korisnici mreže YouTube zapravo dosta prate politiku i Hrvatski Sabor. Kako je nakon izbora za zastupnike koji će ući u Hrvatski sabor 5.7.2020. Ivan Pernar ispao, popularnost mu je pala i zamijenio ga je Most koji zauzima šest mjesta u listi deset najgledanijih videa drugog vala (Slika 22). Na listi su još Antena Zagreb, NewsBar, Matrix TV i MC Seki za kojeg sam prvo mislio da je greškom ušao u listu, ali prolazeći kroz stihove pjesme shvatio sam da **search()** funkcija dohvaća videa čak i ako nađe ključnu riječ u opisu videa, a ne samo naslovu. U opisu video spota je tekst pjesme, a stih „Otkad došo nam je virus cijene otišle u plafon“ sadrži ključnu riječ **virus**.

views - DataFrame

Index	Channel	Video Title	Video Published	Video Views
228	Most	GRMOJA ISTINOM UZDRMAO ANDREJA PLENKOVIČA!!!!	2020-11-25 18:49:02	45959
150	Most	BULJ OSTRO PO VLADI: PLENKOVIČU, SRAMI SE - POGLEDAJ!!	2020-12-06 17:18:00	44572
27	Most	GRMOJA ISTINOM RAZOTKRIO BEROSA!!	2020-12-04 06:45:11	17540
201	Most	BULJ ISTINOM UTISAO KATARINU PEOVIČ!!	2020-11-26 15:44:36	16978
14	Antena Zagreb	NOVI TAJNI ZVUK VRIJEDI VISE! #TajniZvuk2020	2020-10-23 10:57:19	15621
297	MC Seki	MC SEKI & ZIZI - AFTER NA KROVU 2 ft. LERDI CREW	2020-12-24 23:00:12	14818
62	Matrix TV	Zagreb - PROSVJED - Prvi GOVORNIK / 08.11.2020	2020-11-08 15:06:15	14400
0	NewsBar	Kupovina božićnih poklona u vrijeme korone Skec @NewsBar	2020-12-03 08:26:23	13371
135	Most	KAKVA PORUKA MARIJE SELAK RASPUDIĆ GORDANU JANDROKOVIČU!!	2020-12-02 21:53:39	11027
175	Most	IZ VLADE ANDREJA PLENKOVIČA NISU OČEKIVALI OVU PORUKU MOSTOVACA!!!	2020-11-27 17:54:48	10014

Format Resize Background color Column min/max Save and Close Close

Slika 22. Videozapisi s najviše pregleda tokom drugog vala pandemije

Gore spomenuti videozapis je prvi na listi s najviše lajkova (Slika 23), dok je na listi s najviše komentara (Slika 24) videozapis Matrix TV-a na kojemu govornik s prosvjeda protiv vladinih mjera naglašava kako su mjerama oduzeta prava i slobode zajamčene ustavom. Mnogo korisnika ga je podržalo putem komentara pa se iz toga da zaključiti da većina ljudi nije zadovoljna uvođenjem mjera.

likes - DataFrame

Index	Channel	Video Title	Video Published	Video Likes
297	MC Seki	MC SEKI & ZIZI - AFTER NA KROVU 2 ft. LERDI CREW	2020-12-24 23:00:12	3052
62	Matrix TV	Zagreb - PROSVJED - Prvi GOVORNIK / 08.11.2020	2020-11-08 15:06:15	915
27	Most	GRMOJA ISTINOM RAZOTKRIO BEROŠA!!	2020-12-04 06:45:11	553
0	NewsBar	Kupovina božićnih poklona u vrijeme korone Skec @NewsBar	2020-12-03 08:26:23	468
228	Most	GRMOJA ISTINOM UZDRMAO ANDREJA PLENKOVICA!!!!	2020-11-25 18:49:02	433
201	Most	BULJ ISTINOM UTISAO KATARINU PEOVIC!!	2020-11-26 15:44:36	413
135	Most	KAKVA PORUKA MARIJE SELAK RASPUDIC GORDANU JANDROKOVICU!!	2020-12-02 21:53:39	404
150	Most	BULJ OSTRO PO VLADI: PLENKOVICU, SRAMI SE - POGLEDAJ!!	2020-12-06 17:18:00	387
158	Most	PETROV PORUCIO PLENKOVICU - NECE ICI!	2020-11-27 14:40:17	342
248	Matrix TV	Pripreme za Advent u Zagrebu	2020-11-27 14:31:10	335

Format Resize Background color Column min/max Save and Close Close

Slika 23. Videozapisi s najviše lajkova tokom drugog vala pandemije

comments - DataFrame

Index	Channel	Video Title	Video Published	Video Comments
62	Matrix TV	Zagreb - PROSVJED - Prvi GOVORNIK / 08.11.2020	2020-11-08 15:06:15	208
297	MC Seki	MC SEKI & ZIZI - AFTER NA KROVU 2 ft. LERDI CREW	2020-12-24 23:00:12	193
201	Most	BULJ ISTINOM UTISAO KATARINU PEOVIC!!	2020-11-26 15:44:36	165
228	Most	GRMOJA ISTINOM UZDRMAO ANDREJA PLENKOVICA!!!!	2020-11-25 18:49:02	150
183	Most	SRAMOTA NAD SRAMOTAMA U REZIJI VLADE ANDREJA PLENKOVICA!!	2020-11-29 14:21:25	149
248	Matrix TV	Pripreme za Advent u Zagrebu	2020-11-27 14:31:10	136
265	Matrix TV	Subota jutro - U Zagrebu nema ljudi	2020-12-12 11:26:22	114
135	Most	KAKVA PORUKA MARIJE SELAK RASPUDIC GORDANU JANDROKOVICU!!	2020-12-02 21:53:39	104
27	Most	GRMOJA ISTINOM RAZOTKRIO BEROŠA!!	2020-12-04 06:45:11	80
192	Most	VLADA OPET NA SILU SRLJA S MJERAMA BEZ KRITERIJA ILI ANALIZE!	2020-12-04 13:51:47	62

Format Resize Background color Column min/max Save and Close Close

Slika 24. Videozapisi s najviše komentara tokom drugog vala pandemije

4. Analiza mreže

U ovom poglavlju kreirali smo graf društvene mreže YouTube i obavili analizu na globalnoj, središnjoj i lokalnoj razini. Općenito, mreža ili graf definirana je kao skup vrhova i skup bridova koji povezuju vrhove [9]. Naša mreža će se sastojati od čvorova koji će predstavljati YouTube kanal koji je kreirao videozapis, a s druge strane registriranog YouTube korisnika, koji ujedno može biti i kanal, koji je komentirao videozapis. Veza između dva YouTube kanala će biti uspostavljena ukoliko je ostvaren komentar.

Za kreiranje grafa ćemo koristiti **networkx** biblioteku [9]. Stupce 'Channel' i 'Author' iz datoteke **HrvatskiKomentari.csv**, čij je postupak generiranja prikazan na slici 20, ćemo izdvojiti u listu uređenih parova **cvorovi** i ukloniti petlje (Slika 23, linija 241). U listu **tezine** ćemo pohraniti podatke o broju ponavljanja određenog uređenog para prebrojavanjem liste 'cvorovi' pomoću funkcije **count**.

```
229 def lista_bridova():
230     podaci = pd.read_csv('HrvatskiKomentari.csv')
231     cvorovi = [tuple(x) for x in podaci[['Channel', 'Author']].to_numpy()]
232     graph_df = pd.DataFrame()
233     cvorovi1 = []
234     cvorovi2 = []
235     tezine = []
236     for brid in cvorovi:
237         obrnuti = "(" + brid[1] + ", " + brid[0] + ")"
238         if obrnuti in cvorovi:
239             print(obrnuti)
240         if(brid[0] == brid[1]):
241             cvorovi.remove(brid)
242         else:
243             cvorovi1.append(brid[0])
244             cvorovi2.append(brid[1])
245             tezine.append(cvorovi.count(brid))
246     graph_df['Cvor1'] = cvorovi1
247     graph_df['Cvor2'] = cvorovi2
248     graph_df['Tezine'] = tezine
249     graph_df = graph_df.drop_duplicates()
250     graph_df.to_csv("Bridovi.csv", index=False, encoding='utf-8-sig')
```

Slika 25. Kreiranje liste bridova

Nakon zamjene pozicija čvorova, funkcija **print** (Slika 25, linija 239) u konzoli nije ispisala ništa. To znači da autori videozapisa uglavnom ne komentiraju videa drugih kanala već uglavnom objavljuju sadržaj. Od liste čvorova i težina kreiran je dataframe **graph_df** sa stupcima **Cvor1**, **Cvor2** i **Tezina**. Nakon odbacivanja identičnih redova stvorili smo datoteku **Bridovi.csv** koja će nam poslužiti kao početni skup podataka pri analizi mreže i kreiranju grafa. Mreža koja sadrži podatke od početka pandemije referirat će se kao **YNet1**.

Unutar funkcije **analiza_mreze()** stvorit ćemo varijablu **bridovi** te u nju pohraniti datoteku **Bridovi.csv** kao dataframe. Kreirat ćemo prazan graf **YNet1** s naredbom **nx.Graph()**. U varijablu **podaci** pohranit ćemo uređene trojke (čvor 1, čvor 2, težina) iz dataframea. Svaka trojka sadrži podatke o jednom bridu, a pomoću njih ćemo graditi prethodno kreirani graf kroz petlju funkcijom **add_edge()**.

Konstrukcija grafa **YNet1** nam je dala mogućnost da mrežu analiziramo na globalnoj, središnjoj i lokalnoj razini. Što globalne razine tiče, saznali smo da je broj čvorova **364**, od čega kreatora videozapisa **40**, a komentatora **324**. Broj veza je **334**. Za lokalnu razinu smo izračunom stupnja za svakog autora došli do podatka da su videozapisi kanala

IvanPernarTV najkomentiraniji (136), a korisnik HZCAR-Croatia/trains je napisao najviše komentara (4).

Na isti način smo kreirali i mrežu YTNet2 za drugi val pandemije. Analizom smo saznali da ona ima **313** čvorova od kojih su **26** kreatori videozapisa, a **287** je komentatora. Broj veza je **298**. Videozapisi kanala Most dobivaju najviše komentara (123), a korisnik P K je napisao najviše komentara (9). Ostale globalne mjere i način kako doći do njih ću prikazati u nastavku za obje mreže.

Mjera	Vrijednost
Broj čvorova	364
Broj bridova	334
Prosječni stupanj	1,83516E+15
Prosječna snaga	1,93956E+16
Broj povezanih komponenti	33
Broj čvorova najveće povezane komponente	179
Broj bridova najveće povezane komponente	180
Prosječna duljina najkraćih puteva	1,57549E+16
Dijametar mreže	2,09091E+15
Ekscentričnost mreže	1,85077E+14
Globalna učinkovitost	9,91683E+15
Globalni koeficijent grupiranja	3,31418E+15
Prosječni koeficijent grupiranja	8,2036E+15
Asortativnost	-3,22442E+15
Broj nepostojećih bridova	65732
Gustoća mreže	5,05555E+15
Prosječna centralnost blizine	8,61541E+15
Prosječna centralnost međupoloženosti	1,56791E+16

Tablica 3. Globalne mjere za YTNet1 (1.1.2020 – 25.12.2020)

Podaci u tablici 3 se odnose općenito na mrežu. Na primjer prosječni stupanj ili snaga čvora ne znače da je prosječan broj komentara po objavi bio **1.835**. Prosječni broj veza smo dobili dijeljenjem zbroja stupnjeva svih čvorova grafa s brojem čvorova, a prosječnu snagu dijeljenjem zbroja snage svih čvorova s brojem čvorova grafa. Snaga čvora je zbroj težina bridova povezanih s čvorom, a stupanj čvora broj bridova povezanih s njim. Da smo htjeli saznati prosječan broj komentara ili bi napravili bipartitni graf, te izračunali prosječan stupanj particije autora ili podijelili zbroj stupnjeva čvorova autori s brojem tih čvorova.

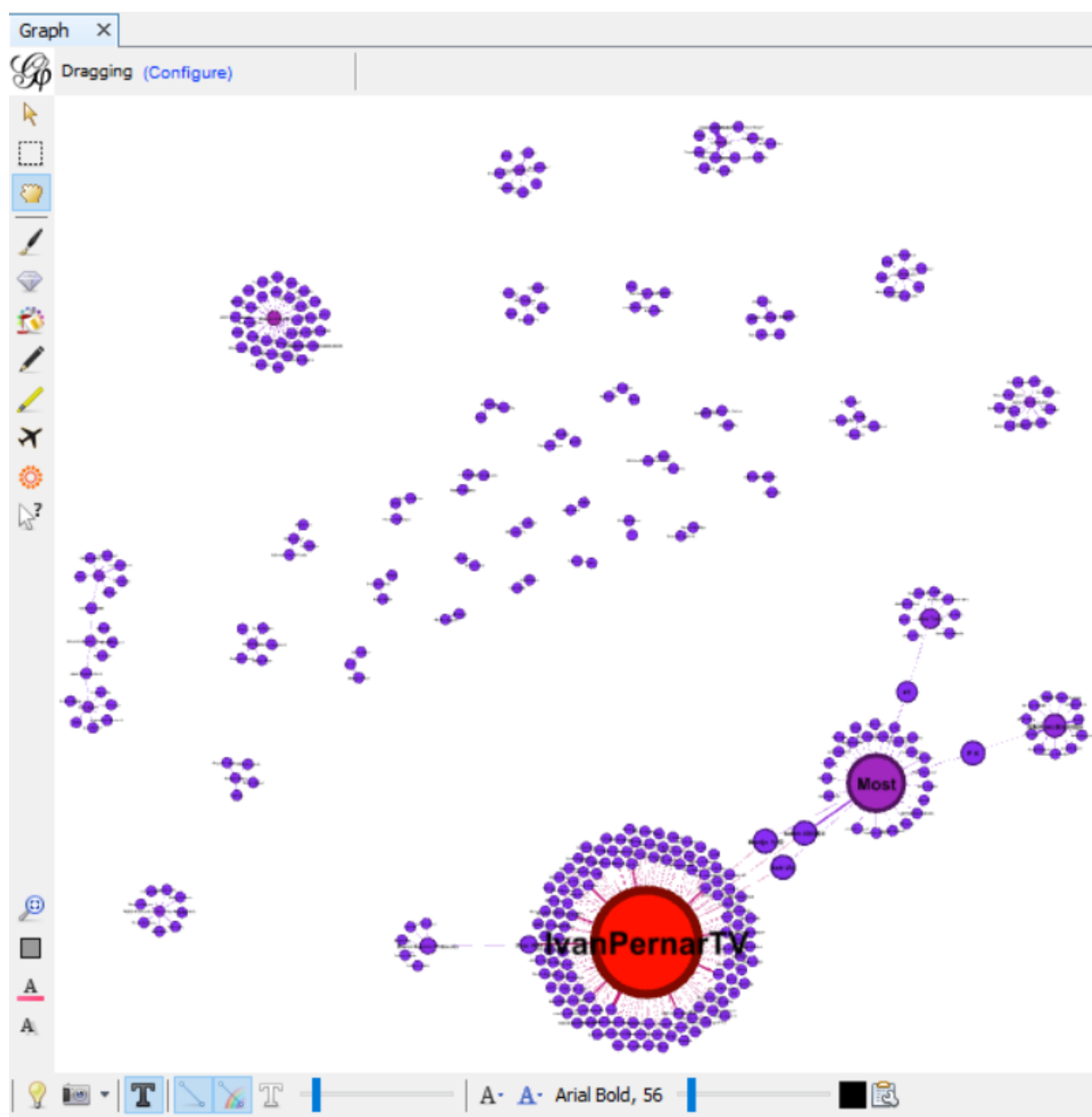
Povezanost je važno svojstvo mreže jer neke mjere mreže ima smisla računati samo na povezanom dijelu mreže. Za mrežu kažemo da je povezana ukoliko između svaka dva čvora postoji put. Vrlo često se događa da nije cijela mreža povezana, već postoje povezane komponente mreže. Povezana komponenta mreže (engl. connected component) je podskup skupa čvorova za koji vrijedi: 1) za svaka dva vrha iz podskupa postoji put između ta dva čvora, 2) podskup nije dio većeg podskupa s istim svojstvima (to je najveći takav podskup).

Za kompleksne mreže uobičajeno je da većina vrhova (čvorova) pripada jedinstvenoj komponenti koja se naziva najveća povezana komponenta (engl. giant connected component, GCC), [10]. Takav slučaj je i kod naše mreže gdje vidimo kako je komponenta koja spaja čvorove IvanPernarTV i Most najveća (Slika 26), a sastoji se od 179 čvorova i 180 bridova.

Pomoću networkx biblioteke i funkcije **number_connected_components(YTNet1)** smo došli do broja 33 što odgovara vizualizaciji sa slike 26.

Nakon učitavanja datoteke Bridovi.csv u alat za vizualizaciju podataka Gephi [11] i pokretanja opcije za izračun duljine prosječnog puta dobili smo broj **3.2039496884918304** što nam ukazuje da postoje korisnici koji su ostavili komentare na videozapise s više različitih YouTube kanala. Po dodatnom podešavanju parametara kao što su boja prema **centralnosti međupoloženosti** i veličina prema **stupnju** čvora, došli smo do boljeg grafičkog prikaza mreže i potvrde pretpostavke iz prethodne rečenice.

Tri korisnika su komentirali videa kanala IvanPernarTV i Most što nam daje naslutiti da bi ta dva kanala mogla objavljivati sličan sadržaj. Pregledavajući potom oba YouTube kanala, vidi se sličnost u objavi videozapisa među kojima je dosta onih koji osuđuju mjere i postupanja vladajuće stranke vezano za upravljanje korona krizom.



Slika 26. YTNet1 mreža autora i komentatora videozapisa

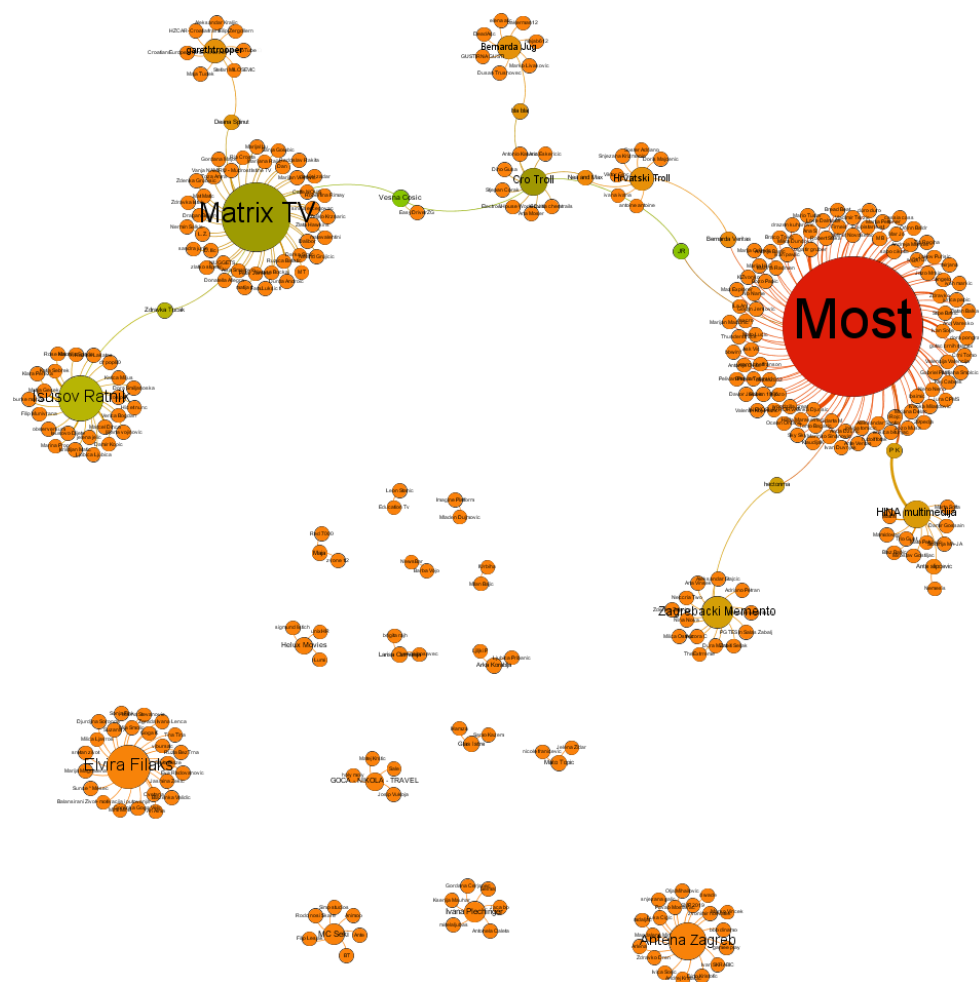
Podatak o broju zajednica izračunat je u alatu Gephi koji koristi Luvain algoritam [12] za analizu zajednica. Mreža ima **37** zajednica, a pet njih je unutar najveće povezane komponente. To su kanali Emanuel Putovanja, IvanPernarTV, Most, Cro Troll i HINA Multimedija. Modularnost je **0,810**. Ostale mjere koje nisam objasnio na prethodnoj tablici ću objasniti na tablici koja sadrži podatke o mreži YTNet2 (Tablica 4).

Mjera	Vrijednost
Broj čvorova	313
Broj bridova	298
Prosječni stupanj	1,90415E+15
Prosječna snaga	2,17891E+16
Broj povezanih komponenti	16
Broj čvorova najveće povezane komponente	224
Broj bridova najveće povezane komponente	224
Prosječna duljina najkraćih puteva	1,68085E+16
Dijametar mreže	23125
Ekscentričnost mreže	2,01834E+16
Globalna učinkovitost	1,43181E+16
Globalni koeficijent grupiranja	0
Prosječni koeficijent grupiranja	0
Asortativnost	-3,648E+15
Broj nepostojećih bridova	485300
Gustoća mreže	6,10306E+15
Prosječna centralnost blizine	1,13015E+16
Prosječna centralnost međupoloženosti	6,52313E+15

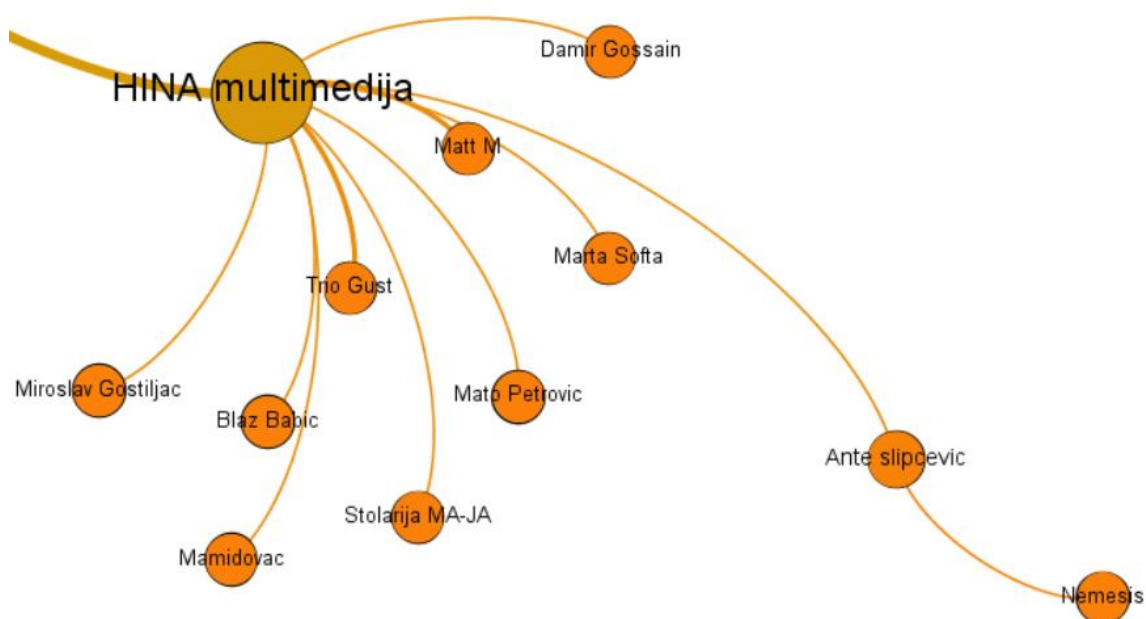
Tablica 4. Globalne mjere za YTNet2 (2.10.2020 – 25.12.2020)

Po broju čvorova i bridova, vidimo da je ova mreža nešto manja mreže YTNet2, ali ima veću prosječnu snagu i stupanj čvora. Postoji 16 povezanih komponenti od kojih najveća ima 224 čvora i 224 brida. Nakon vizualizacije u Gephiju vidimo da se ona sastoji od kanala: Zagrebački Memento, HINA multimedija, Most, Hrvatski Troll, Cro Troll, Bernarda Jug, EasyDriverZG, MatrixTV, garethtrooper i Isusov Ratnik (Slika 27). Algoritam u Gephiju je izračunao kako mreža ima **23** zajednice i modularnost **0,803**. Naknadnom provjerom sam ustvrdio da je broj zajednica nešto manji od broja kanala koji su kreirali videozapise. U jednom i u drugom grafu je za 3 manji zbog osoba koje su ostavile jedan komentar na popularnom kanalu i jedan (koji je ujedno i jedini) na manje popularnom. Na taj način se manje popularan kanal ubrojio u istu zajednicu s onim popularnijim. Primjer takvog slučaja vidimo na slici 28 gdje je kanal Nemesis ubrojen u istu zajednicu s kanalom HINA multimedija zbog Ante Slipčevića koji je ostavio komentar na oba kanala.

Što se tiče mjera udaljenosti, one su važne kod proučavanja mreža iz više razloga. Općenito, one ukazuju na duljine putanja u mreži, te koliko je povezanost u mreži dobra. Udaljenost između dva čvora u mreži (grafu) definira se kao duljina najkraćeg puta između ta dva čvora. Posebno se razmatra prosječna duljina najkraćih puteva (engl. average shortest path length). Računa se kao zbroj najkraćih puteva između svaka dva čvora u mreži podijeljen s ukupnim mogućim brojem puteva [10]. U ovom grafu ona iznosi **1.68**.



Slika 27. YTNet2 mreža autora i komentatora videozapisa



Slika 28. Spajanje 2 kanala u jednu zajednicu

Dijametar mreže (engl. diameter), D , je najveća udaljenost među čvorovima grafa, odnosno najveća vrijednost između svih mogućih najkraćih puteva u mreži. U mreži YTNet2 ona iznosi **2,3125**. Ekscentričnost (engl. eccentricity) čvora je najveća udaljenost čvora od drugih čvorova u mreži. Tu smo morali zbrojiti prosječne ekscentričnosti čvora u svakoj komponenti i podijeliti ih s brojem komponenti. U mreži YTNet2 ona iznosi **2,018**.

Globalna učinkovitost je mjera koja kvantificira koliko se učinkovito informacija može slati kroz mrežu pretpostavljajući da je učinkovitost slanja informacije između dva čvora i i j proporcionalna recipročnoj vrijednosti njihove udaljenosti. U mreži YTNet2 ona iznosi **0,143**.

Tranzitivnost za neusmjerenu mrežu računa se prebrojavanjem povezanih trojki čvorova (engl. triplets) u mreži. Ona određuje tendenciju mreže da se slični čvorovi spajaju sa sličnim čvorovima. Koeficijent grupiranja (engl. clustering coefficient) je mjera koja nam govori koliko su čvorovi u mreži međusobno grupirani. Njime mjerimo koliko je svaki čvor povezan sa svojim susjednim čvorovima. Na razini cijele mreže određuje se koliko je cijeli graf, blizu tome da tvori potpuni graf (engl. complete graph). Iz meni nepoznatih razloga obje mjere u mreži YTNet2 su **0**.

Asortativnost određuje tendenciju mreže da se slični čvorovi spajaju sa sličnim čvorovima. Koeficijent asortativnosti, r , određuje se prema Pearsnovoj korelaciji i može poprimiti vrijednost iz segmenta $[-1,1]$. Pozitivne vrijednosti od r ukazuju na to da je mreža asortativna, a negativne na to da mreža nije asortativna. Za mrežu YTNet2 njena vrijednost je **-0,3648**.

Centralnost međupoloženosti govori koliko je vjerojatno da se čvor nalazi na putu između neka dva čvora. Pokazuje koji čvorovi su više vjerojatni za prijenos komunikacije između dva druga čvora. Koristi se i kao pokazatelj gdje bi se mreža raspala, tj koji bi čvorovi bili otkinuti ako nestane dio čvorova. Jednaka je broju najkraćih puteva koji prolaze kroz čvor podijeljenom sa svim najkraćim putevima u mreži [10]. U mreži YTNet2 ona iznosi **0.0065**.

Nakon izračuna globalnih mjera potrebno je izračunati i lokalne. To su mjere koje se tiču samih čvorova. Za početak ćemo prikazati kanale i korisnike s najviše komentara, a potom odrediti centralne čvorove prema centralnosti stupnja čvora, centralnosti međupoloženosti i centralnosti blizine za obje mreže. Što se tiče kanala s najviše komentara to su IvanPernarTV za mrežu YTNet1 (Tablica 5) i Most za mrežu YTNet2 (Tablica 6).

Autor	Stupanj
IvanPernarTV	136
Antena Zagreb	33
Most	29
HINA multimedija	14
garethtrooper	13
Osnove Elektrotehnike	10
Cro Troll	9
MORH Ministarstvo obrane Republike Hrvatske	8
Linda Spajic	7
Newsify	7

Tablica 5. Kanali s najviše komentara (YTNet1)

Autor	Stupanj
Most	123
Matrix TV	44
Isusov Ratnik	31
Elvira Filaks	24
Antena Zagreb	18
HINA multimedija	16
Zagrebacki Memento	14
Cro Troll	11
Hrvatski Troll	9
garethtrooper	8

Tablica 5. Kanali s najviše komentara (YTNet2)

Kod pretrage najkomentiranih kanala sam dohvaćao naziv i snagu svakog čvora koji se nalazi u listi autora, dok sam za dohvaćanje korisnika s najviše komentara dohvaćao naziv i snagu svakog čvora koji se ne nalazi u listi autora za svaki od grafova (Slika 38). Prilikom izračuna ostalih mjera sam prolazio kroz sve čvorove grafova.

Korisnik	Broj komentara
HZCAR-Croatia/trains	4
Hhgg Hhhh	3
alo alo	3
sabo cikotic	3
TheJelena 1222	2
Darko Simic	2
tomodrx	2
JR	2
P K	2
Deni Domina	2

Tablica 7. Najaktivniji korisnici (YTNet1)

Korisnik	Broj komentara
P K	7
Anka Djurisc	6
sabo cikotic	6
olga tir gruber	4
Matija Gecek	3
Vesna Cosic	3
Ivan Solje	3
Katia Sebrek	3
Zdravko	2
Zdravka Tucak	2

Tablica 8. Najaktivniji korisnici (YTNet2)

Stupanj čvora u analizi društvenih mreža označava koliko čvorova izravno može dohvatiti zadani čvor. Centralnost stupnja čvora određuje se normalizacijom stupnja čvora, dakle $k/(N-1)$ gdje je N ukupni broj čvorova. Čvor koji ima najveći stupanj (može ih biti i više) se zove hub. Većinom veći stupanj čvora označava i veću važnost ili popularnost čvora u mreži [10]. U mreži YTNet1 hub je IvanPernarTV (Tablica 9), a u YTNet2 Most (Tablica 10).

Autor	Centralnost stupnja
IvanPernarTV	3,44353E+16
Antena Zagreb	9,09091E+15
Most	7,7135E+15
HINA multimedija	3,30579E+15
Osnove Elektrotehnike	2,75482E+16
garethtrooper	2,47934E+16
Cro Troll	2,47934E+16
MORH Ministarstvo obrane Republike Hrvatske	2,20386E+15
Linda Spajic	1,92837E+15
Newsify	1,92837E+15

Tablica 9. Centralnost stupnja čvora (YTNet1)

Autor	Centralnost stupnja
Most	3,07692E+15
Matrix TV	1,34615E+15
Isusov Ratnik	7,69231E+15
Elvira Filaks	7,37179E+15
Antena Zagreb	5,76923E+15
Zagrebacki Memento	4,48718E+15
Cro Troll	3,52564E+16
HINA multimedija	3,52564E+16
Bernarda Jug	2,5641E+15
Hrvatski Troll	2,5641E+15

Tablica 10. Centralnost stupnja čvora (YTNet2)

Autor	Centralnost blizine
IvanPernarTV	2,71068E+15
sabo cikotic	2,15515E+15
Ask Vic	2,15515E+15
Marija H-G	2,15515E+15
Most	1,79596E+16
Hhgg Hhhh	1,79227E+16
Ammon Ray	1,74917E+16
tiho6662002	1,74917E+16
Goran Jerkovic	1,74917E+16
HARD MADNESS	1,74917E+16

Tablica 11. Centralnost stupnja čvora (YTNet2)

Centralnost blizine određuje koliko brzo čvor može dosegnuti bilo koji drugi čvor u mreži, tj. koliko u prosjeku iznosi najkraći put do svih ostalih čvorova u mreži. Ova mjera je važna za slučajeve kada se zahtjeva velika brzina prijenosa informacije. Što je manja vrijednost, to je optimalnije za prijenos informacije. Predstavlja važnost čvora u topološkom smislu, jer je čvor sa najvišom centralnošću blizine najbliži ostalim čvorovima. Računa se kao recipročna vrijednost zbroja svih duljina najkraćih puteva od zadanog čvora do ostalih čvorova [10]. U mreži YTNet1 najveću centralnost blizine ima čvor IvanPernarTV (Tablica 11), a u mreži YTNet2 čvor Most (Tablica12).

Autor	Centralnost blizine
Most	2,28023E+15
JR	2,17446E+16
Cro Troll	2,1111E+16
Vesna Cosic	1,9273E+16
hectorima	1,78486E+15
P K	1,7769E+16
Matrix TV	1,76509E+16
Bernarda Veritas	1,76119E+16
Rojc	1,7306E+16
Ocean Ondine	1,7306E+16

Tablica 12. Centralnost blizine (YTNet2)

Autor	Centralnost međupoloženosti
IvanPernarTV	2,20835E+16
Most	1,06304E+16
Ask Vic	4,61394E+15
Marija H-G	4,61394E+15
P K	3,03183E+16
HINA multimedija	2,87962E+15
JR	2,31496E+16
Cro Troll	2,11254E+15
Hhgg Hhhh	1,5707E+16
Simun Emanuel Putovanja	1,33175E+15

Tablica 13. Centralnost međupoloženosti (YTNet1)

Autor	Centralnost međupoloženosti
Most	3,99209E+15
Cro Troll	2,76548E+15
Matrix TV	2,72941E+15
JR	2,3415E+16
Vesna Cosic	2,30316E+15
Isusov Ratnik	1,00029E+16
Zdravka Tucak	9,84418E+15
hectorima	6,031E+15
Zagrebacki Memento	5,78778E+16
P K	5,2189E+15

Tablica 14. Centralnost međupoloženosti (YTNet2)

Centralnost međupoloženosti sam već objasnio kod globalnih mjera. Za mrežu YTNet1 čvor s najvećom centralnošću međupoloženosti je IvanPernarTV (Tablica 13), a za mrežu YTNet2 Most (Tablica 14).

Zadnja mjera za izračun je centralnost svojstvenog vektora. Ona određuje koliko dobro je čvor povezan za ostalim čvorovima koji su jako dobro povezani. Određuje važnost čvora u mreži uzimajući u obzir važnosti susjednih čvorova. To je mjera koja se čvoru dodjeljuje u odnosu na rezultate svih čvorova u mreži tako da veza do čvorova veće važnosti pridonosi rezultatu čvora više nego jednaka veza do čvora manje važnosti. Ova vrijednost je proporcionalna sumi svojstvenih centralnosti svih čvorova spojenih direktno s tim čvorom [10]. Za mrežu YTNet1 najveću vrijednost centralnosti svojstvenog vektora ima IvanPernarTV (Tablica 15), a za mrežu YTNet2 Most (Tablica 16).

Autor	Centralnost svojstvenog vektora
IvanPernarTV	7,06696E+15
alo alo	1,67535E+16
Hhgg Hhhh	1,12415E+15
Ana Pales	1,1169E+16
tomodrx	1,1169E+16
fiona h	1,1169E+16
da milano	1,1169E+16
Darko Simic	1,1169E+16
Deni Domina	1,1169E+16
Marijana Beram	1,1169E+16

Tablica 15. Centralnost svojstvenog vektora (YTNet1)

Autor	Centralnost svojstvenog vektora
Most	7,05774E+15
sabo cikotic	2,80588E+16
Anka Djurisc	2,80588E+16
olga tir gruber	1,87059E+15
P K	1,51762E+16
Ivan Solje	1,40294E+16
Jozo Mrsic	9,35294E+15
M B	9,35294E+15
Robert Strika	9,35294E+15
lbsimic	9,35294E+15

Tablica 16. Centralnost svojstvenog vektora (YTNet2)

5. Zaključak

U ovom radu opisan je postupak analize podataka prikupljenih s društvene mreže YouTube sa svrhom boljeg razumijevanja utjecaja pandemije uzrokovane Sars-Cov-2 virusom na građane Republike Hrvatske. Za dohvaćanje podataka korišten je YouTube Data API, sučelje koje dohvaća informacije o videozapisima s YouTube platforme na temelju unesene ključne riječi ili više njih. Kako bi analiza bila što bolja stvorena su dva skupa podataka. Prvi (S1) je sadržavao podatke o videozapisima objavljenima od 1.1.2020 za ključne riječi 'koronavirus', 'korona mjere', 'korona cjepivo' i slične, dok je drugi (S2) sadržavao iste podatke, ali za razdoblje drugog vala (od 2.10.2020). Prikupljeni skupovi podataka sadržavali su informacije o videozapisu (autor, naslov, datumu objave, broj pregleda, lajkova i komentara), te njegovom komentaru ili više njih (autor, sadržaj komentara, datum objave, broj lajkova). Početne skupove podataka smo nakon čišćenja (micanja osjećajnika iz komentara, brisanje komentara koji nisu na hrvatskom jeziku) krenuli analizirati.

Prilikom sortiranja datuma objava za potrebe vizualizacije broja objavljenih videozapisa po danima kod skupa S1 sam primjetio da je prvi videozapis vezan uz koronavirus objavljen 2 tjedna prije nego što se pojavio prvi slučaj u Republici Hrvatskoj. Kasnijom analizom i praćenjem stvorene vizualizacije uočio sam da broj objavljenih videozapisa varira kao reakcija na objave javnih medija, rast broja zaraženih, te uvođenje mjera za sprječavanje širenja pandemije od strane stručnog stožera. Sada, godinu dana nakon izbijanja pandemije, na temelju variranja broja zaraženih u cijelom svijetu znamo da je virus sezonalan. Ona je također utjecala na graf broja objavljenih videa po danima jer tokom ljeta rasta slučajeva nije bilo, nije bilo objava medija, uvođenja mjera od strane stožera, pa tako ni videozapisa na YouTube-u. Najpopularniji kanal za vrijeme prvog vala bio je IvanPernarTV koji je žestoko kritizirao mjere stožera i odluke vlade. Iako ga je YouTube zajednica podržavala na izborima u srpnju je ostao bez mjesta u saboru. Analizom skupa S2 postalo je jasno da su njegovi korisnici počeli pratiti kanal Most i MatrixTV koji nisu bili toliko popularni u skupu S1. S obzirom na naslove videa, njihov sadržaj, te datum objave da se zaključiti da je ta zajednica koja ih prati protiv vladajuće stranke i njihovih mjera. U zadnjem poglavlju rada smo iz skupa S1 konstruirali mrežu YNet1 i iz skupa S2 mrežu YNet2 kako bi dobili uvid u zajednice, te imali vizualnu predodžbu mreža. Zajednice pratitelja prethodno spomenutih kanala su u obje mreže bile uključene u najveću povezanu komponentu što se dalo nagovijestiti prethodnim izračunima lokalnih mjera čvorova grafova.

Izrada ovog diplomskog rada mi je bila iznimno zanimljiva. Tema me privukla jer sam i sam prebolio koronu, te me zanimalo kako ostali ljudi gledaju na cijelu situaciju uzrokovanu pandemijom. Simptome prehlade sam imao tri dana, a od svega mi je najteže pala izolacija koja me dosta oštetila novčano i psihički. To mi je dalo razumijevanje za komentare većine korisnika YouTube mreže koji pokazuju negodovanje prilikom uvođenja represivnih mjera koje ograničavaju život građana, onemogućavaju slobodu kretanja, a nekima i pravo na rad. S obzirom da nisam mogao doći do podataka o starosti YouTube korisnika, pretpostavljam da je to uglavnom populacija mlađe ili srednje životne dobi. To sam zaključio iz razgovora s vršnjacima i starijim članovima obitelji kojima su, za razliku od mlađih, jedini izvor informacija tv program i javni mediji, a oni ipak imaju drugačiji i ozbiljniji pogled na cijelu situaciju uzrokovanu pandemijom koronavirusa.

Popis literature i izvora

- [1] Wikipedija, »Pandemija koronavirusa 2019./20.,« 27 Studeni 2020. [Mrežno]. Dostupno: https://hr.wikipedia.org/wiki/Pandemija_koronavirusa_2019./20.. [Pokušaj pristupa 23 Prosinac 2020].
- [2] »Youtube About,« Youtube LLC, 2020. [Mrežno]. Dostupno: <https://www.youtube.com/intl/hr/about/>. [Pokušaj pristupa 10 12 2020].
- [3] »YouTube,« Wikipedija, 30 11 2020. [Mrežno]. Dostupno: <https://hr.wikipedia.org/wiki/YouTube>. [Pokušaj pristupa 28 12 2020].
- [4] »Koronavirus.hr,« Vlada Republike Hrvatske, 2020. [Mrežno]. Dostupno: <https://www.koronavirus.hr/>. [Pokušaj pristupa 28 12 2020].
- [5] »Python,« Python Software Foundation, 2020. [Mrežno]. Dostupno: <https://www.python.org/>. [Pokušaj pristupa 28 12 2020].
- [6] »Google APIs,« Google LLC, 2020. [Mrežno]. Dostupno: <https://console.developers.google.com/>. [Pokušaj pristupa 28 12 2020].
- [7] »YouTube Data API,« Google LLC, 2020. [Mrežno]. Dostupno: <https://developers.google.com/youtube/>. [Pokušaj pristupa 27 12 2020].
- [8] M. Stojiljković, »NumPy, SciPy, and Pandas: Correlation With Python,« 2019.. [Mrežno]. Dostupno: <https://realpython.com/numpy-scipy-pandas-correlation-python/>. [Pokušaj pristupa 29 Prosinac 2020.].
- [9] »NetworkX,« NetworkX developers, 8 2020. [Mrežno]. Dostupno: <https://networkx.org/>. [Pokušaj pristupa 28 12 2020].
- [10] A. Meštrović, »Merlin,« 11 11 2019. [Mrežno]. Dostupno: https://moodle.srce.hr/2019-2020/pluginfile.php/3248518/mod_resource/content/1/Upravljanje_znanjem_nastavni_materijal_dio2_kompleksne%20mre%C5%BEE_27_11_18_v4.pdf. [Pokušaj pristupa 15 12 2020].
- [11] »Gephi,« Gephi.org, 2017. [Mrežno]. Dostupno: <https://gephi.org/>. [Pokušaj pristupa 15 12 2020].
- [12] V. Blondel, R. L. R. J.L. Guillaume i E. Lefebvre, »Fast unfolding of communities in large networks,« *Journal of Statistical Mechanics: Theory and Experiment*, 2008.
- [13] »YouTube API extracting comments,« GOTRAINED PYTHON TUTORIALS, 2019. [Mrežno]. Dostupno: <https://python.gotrained.com/youtube-api-extracting-comments/>. [Pokušaj pristupa 20 12 2020].
- [14] R. Agrawal, »Extracting & Pre-processing the YouTube Comments,« 23 11 2020. [Mrežno]. Dostupno: <https://www.analyticssteps.com/blogs/extracting-pre-processing-youtube-comments>. [Pokušaj pristupa 15 12 2020].

- [15] A. Maedche, *Ontology Learning for the Semantic Web*, Springer, 2002.
- [16] M. Russell, *Mining the Social Web*, O'Reilly Media, 2011.
- [17] A. Gomez-Perez, M. Fernandez-Lopez i O. Corcho, *Ontological Engineering*, Springer, 2004.
- [18] S. Mader, *WikiPatterns: A Practical Guide to Improving Productivity and Collaboration in our Organization*, Indianapolis: Wiley Publishing, 2008.

Popis priloga

Prilog 1: **covid_ekstrakcija.py**

Prilog 2: **covid_analiza.py**

Prilog 3: **client_secret.json**

Prilog 4: **token.pickle**

Prilog 5: **HrvatskiKomentari.csv**

Prilog 6: **HrvatskiKomentariDrugiVal.csv**

Prilog 7: **FrekvencijeRijeci.csv**

Prilog 8: **FrekvencijeRijeciDrugiVal.csv**

Prilog 9: **ObjaveVidea.csv**

Prilog 10: **ObjaveKomentara.csv**

Prilog 11: **ObjaveVideaDrugiVal.csv**

Prilog 12: **ObjaveKomentaraDrugiVal.csv**

Prilog 13: **Bridovi.csv**

Prilog 14: **BridoviDrugiVal.csv**

Prilog 15: **GlobalneMjere.csv**

Prilog 16: **GlobalneMjereDrugiVal.csv**

Prilog 17: **Mreza.png**

Prilog 18: **MrezaDrugiVal.png**

Popis slika

Slika 1 - Kreiranje projekta na Google konzoli	7
Slika 2 - Odabir API-ja i servisa	8
Slika 3 - Kreiranje vjerodajnica	8
Slika 4 - Popunjavanje informacija o aplikaciji	9
Slika 5 - Preuzimanje konfiguracije za povezivanje s API servisom.....	9
Slika 6 - Stvaranje poveznice za autorizaciju	10
Slika 7 - Funkcija za pretragu po ključnoj riječi	12
Slika 8 - Rezultat pretrage bez filtra.....	12
Slika 9 – Rezultati pretrage nakon filtriranja	13
Slika 10 - Spojeni skup svih prikupljenih podataka	15
Slika 11 - Lista uređenih parova (video, datum objave).....	18
Slika 12 - Lista učestalosti objava videozapisa po danima.....	18
Slika 13 - Broj objavljenih videa po danima	19
Slika 14 - Broj objavljenih videa u drugom valu	19
Slika 15 - Broj objavljenih komentara po danima.....	20
Slika 16 - Broj zaraženih koronavirusom po danu	20
Slika 17 - Broj komentara i zaraženih po danima	21
Slika 18 - Utjecaj broja zaraženih na aktivnost YouTube korisnika.....	21
Slika 19 - Videozapisi s najviše pregleda	22
Slika 20 - Videozapisi s najviše lajkova	22
Slika 21 - Videozapisi s najviše komentara	23
Slika 22 - Videozapisi s najviše pregleda tokom drugog vala pandemije	23
Slika 23 - Videozapisi s najviše lajkova tokom drugog vala pandemije.....	24
Slika 24 - Videozapisi s najviše komentara tokom drugog vala pandemije	24
Slika 25 - Kreiranje liste bridova	25
Slika 26 – YTNet1 mreža autora i komentatora videozapisa	27
Slika 27 – YTNet2 mreža autora i komentatora videozapisa	29
Slika 28 - Spajanje 2 kanala u jednu zajednicu.....	29

Popis tablica

Tablica 1 - Podaci o korpusu (cijela pandemija i drugi val).....	17
Tablica 2 – Top 20 najfrekventnijih riječi (cijela pandemija i drugi val).....	17
Tablica 3 - Globalne mjere za YTNet1 (1.1.2020 – 25.12.2020).....	26
Tablica 4 - Globalne mjere za YTNet2 (2.10.2020 – 25.12.2020).....	28
Tablica 5 - Kanali s najviše komentara (YTNet1).....	30
Tablica 6 - Kanali s najviše komentara (YTNet2).....	31
Tablica 7 - Najaktivniji korisnici (YTNet1)	31
Tablica 8 - Najaktivniji korisnici (YTNet2)	31
Tablica 9 - Centralnost stupnja čvora (YTNet1)	32
Tablica 10 - Centralnost stupnja čvora (YTNet2)	32
Tablica 11 - Centralnost blizine (YTNet1).....	32
Tablica 12 - Centralnost blizine (YTNet2).....	33
Tablica 13 - Centralnost međupoloženosti (YTNet1)	33
Tablica 14 - Centralnost međupoloženosti (YTNet2)	33
Tablica 15 - Centralnost svojstvenog vektora (YTNet1).....	34
Tablica 16 - Centralnost svojstvenog vektora (YTNet2).....	34
