

Wikidata

Domić, Tea

Undergraduate thesis / Završni rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Humanities and Social Sciences / Sveučilište u Zagrebu, Filozofski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:131:656027>

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

Download date / Datum preuzimanja: **2024-08-31**



Sveučilište u Zagrebu
Filozofski fakultet
University of Zagreb
Faculty of Humanities
and Social Sciences

Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences](#)



SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI
Ak. god. 2022./2023.

Tea Domić

Wikidata

Završni rad

Mentor: izv.prof.dr.sc. Tomislav Ivanjko

Zagreb, lipanj 2023.

Izjava o akademskoj čestitosti

Izjavljujem i svojim potpisom potvrđujem da je ovaj rad rezultat mog vlastitog rada koji se temelji na istraživanjima te objavljenom i citiranoj literaturi. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Također izjavljujem da nijedan dio rada nije korišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

(potpis)

SADRŽAJ

1. UVOD	1
2. WIKIDATA	2
2.1. Nastanak.....	2
2.1.1. Wikidata u brojkama.....	5
2.1.2. 10 godina Wikidate	8
2.2. Funkcija.....	9
2.3. Obilježja Wikidata	11
2.4. Kako radi Wikidata	12
2.5. Unos podataka na Wikidatu	15
2.6. Sukobi mišljenja na Wikidatai	16
2.7. Prednosti i mane Wikidatae	17
2.8. Wikidata projekti	18
2.8.1. Wikiprojekt COVID19.....	18
2.8.2. WikiProject Music	20
3. ODNOS WIKIPEDIE I WIKIDATE.....	22
3.1. Ukratko o Wikipediji	22
3.2. Utjecaj Wikidate na Wikipediju.....	24
3.3. Glavne sličnosti i razlike.....	25
4. WIKIMEDIA COMMONS I WIKIDATA.....	27
4.1. Općenito o Wikimedia Commonsu.....	27
4.2. Povezanost sa Wikidatom	28
5. ZAKLJUČAK.....	30
6. LITERATURA	31
POPIS SLIKA.....	33
POPIS GRAFOVA	33
SAŽETAK	34
SUMMARY	35

1. UVOD

Wikidata, besplatna i otvorena baza znanja koju mogu čitati i uređivati kako ljudi tako i strojevi zahvaljujući strukturiranim podacima koje sadrži, izvor je podataka za mnoge projekte Wikimedia Zaklade, ali pruža podršku i mnogim drugim stranicama i uslugama izvan samih Wikimedijinih projekata. Od svog osnutka u listopadu 2012. sve je više rasla u smislu zajednice i sadržaja te danas predstavlja jednu od najznačajnijih web stranica za izvor informacija.

Ovaj rad ima za cilj pružiti opći pregled projekta Wikidata, odnosno njen razvoj, glavne karakteristike, ali i prednosti i mane koje obuhvaća, kako bi se otkrio potencijal u raznim disciplinama, ali i opća važnost ovog projekta za cijelu zajednicu. U radu se također istražuje i odnos Wikidate sa sestrinskim projektima, Wikipedijom, besplatnom enciklopedijom i Wikimedia Commonsom, repozitorijem medijskog sadržaja, kako bi se uvidjele njihove sličnosti i razlike te mogućnosti koje pružaju za cjelokupnu dobrobit i napredak društva.

Motivacija za pisanje ovog rada stoji u činjenici da Internet predstavlja nepresušan izvor informacija i uvelike olakšava život sadašnjice, no isto tako i potencijalni rizik oko istinitosti i kvalitete pruženih podataka. Iako je Wikipedija poznata većem brojem ljudi, njezin sestrinski projekt, Wikidata predstavlja bolju polaznu točku za istraživanje konkretnih informacija i pregled podataka na jednom mjestu.

2. WIKIDATA

Wikidata je besplatna, suradnička, višejezična, sekundarna baza podataka koja prikuplja strukturirane podatke za pružanje podrške za Wikipediju, WikimediaCommons i druge „wikije“ Wikimedijinog pokreta i bilo koga u svijetu. Radi se o besplatnoj i otvorenoj bazi znanja koju mogu čitati i uređivati kako ljudi tako i strojevi. Wikidata također pruža podršku mnogim drugim stranicama i uslugama izvan samih Wikimedijinih projekata. Sadržaj Wikidata dostupan je pod besplatnom licencom, izvezen korištenjem standardnih formata i može se međusobno povezati s drugim skupovima otvorenih podataka na povezanom podatkovnom webu. ("Wikidata", 2023.)

Ona djeluje kao središnja pohrana za strukturirane podatke svojih Wikimedijinih sestrinskih projekata. Dakle, ona je dio Zaklade Wikimedia, odnosno neprofitne organizacije koja upravlja nizom web stranica, a najpoznatija je Wikipedia. Cilj Wikimedije je "*svijet u kojem svako pojedino ljudsko biće može slobodno dijeliti zbroj svih znanja*". (Vrandečić, 2013., str. 90.)

U nastavku poglavlja prikazat će se nastanak Wikidate, njezine glavne karakteristike kao i prednosti i mane ta na koji način radi ovaj Googleov projekt.

2.1. Nastanak

Iako je izvorno zamišljena kao primarno tekstualni izvor, Wikipedia je sve više prikupljala strukturirane podatke, dakle brojeve, datume, koordinate i odnose mnogih vrsta. Ti su podaci postali izvor velike vrijednosti s potencijalnom primjenom u raznim područjima znanosti, tehnologije i kulture. Ovaj razvoj nije iznenađujući s obzirom na to da Wikipediju pokreće sveobuhvatna vizija svijeta u kojem svatko može slobodno dijeliti svo znanje. Danas nema sumnje da se ti podaci mogu pretraživati, analizirati i ponovno koristiti. Stoga je možda iznenađujuće da Wikipedia ne omogućuje izravan pristup većini podataka, bilo putem usluga upita ili putem izvoza podataka koji se mogu preuzeti. Stvarna upotreba podataka je rijetka i često ograničena na vrlo specifične informacije, kao što je geoznačavanje Wikipedijinih članaka koji se koriste u Google kartama. Razlog za

ovaj veliki jaz između vizije i stvarnosti je taj što su Wikipedijini podaci skriveni u 30 milijuna Wikipedijinih članaka na 287 jezika, iz kojih ih je teško izvući. Ova situacija nije baš povoljna za bilo koga tko želi koristiti podatke, ali također predstavlja rastuću prijetnju glavnom cilju Wikipedije, a to je da pruži ažurno i točno enciklopedijsko znanje. Iste informacije često se pojavljuju u člancima na više jezika, a često i u više članaka na jednom jeziku. Na primjer, podatke o stanovništvu za grad Rim mogu se pronaći u engleskim i talijanskim člancima o Rimu, ali brojke su različite. Tako je nastala potreba za osnivanjem Wikidate, gdje će svi podatci biti obuhvaćeni na jednom mjestu. (Vrandečić, Krötzs, 2014. str. 78. – 81.)

Wikidata je osnovana 29. listopada 2012. godine. Osnovao ju je Zdenko, poznatiji kao Denny Vrandečić, hrvatski informatički znanstvenik, prikazan na slici 1., koji je radio u Wikimedia Deutschland, njemačkom ogranku Zaklade Wikimedia. Prethodno je osnovao Semantic Media Wiki, koji koriste mnoge organizacije kao što su NASA, američke obavještajne agencije, Metropolitan Museum of Art i drugi. Osnivač je Wikipedije Hrvatska i izabrani je član Upravnog odbora Zaklade Wikimedia. Živio je u Hrvatskoj, Stuttgartu, Rimu, Los Angelesu, Berlinu i sada u San Francisco Bay Area ("The Knowledge Graph Conference", 25.04.2023.). U Googleu je od listopada 2013. godine. Glavni pravci njegova istraživanja su evaluacija strukture znanja, masivna suradnička izgradnja baze znanja te primjena tehnologije semantičkog weba i slično. (Vrandečić, 2013., str. 92.)



Slika 1. Zdenko Denny Vrandečić – osnivač Wikidate

Izvor: The Knowledge Graph Conference:

<https://www.knowledgegraph.tech/speakers/denny-vrandecic/> (25.04.2023.)

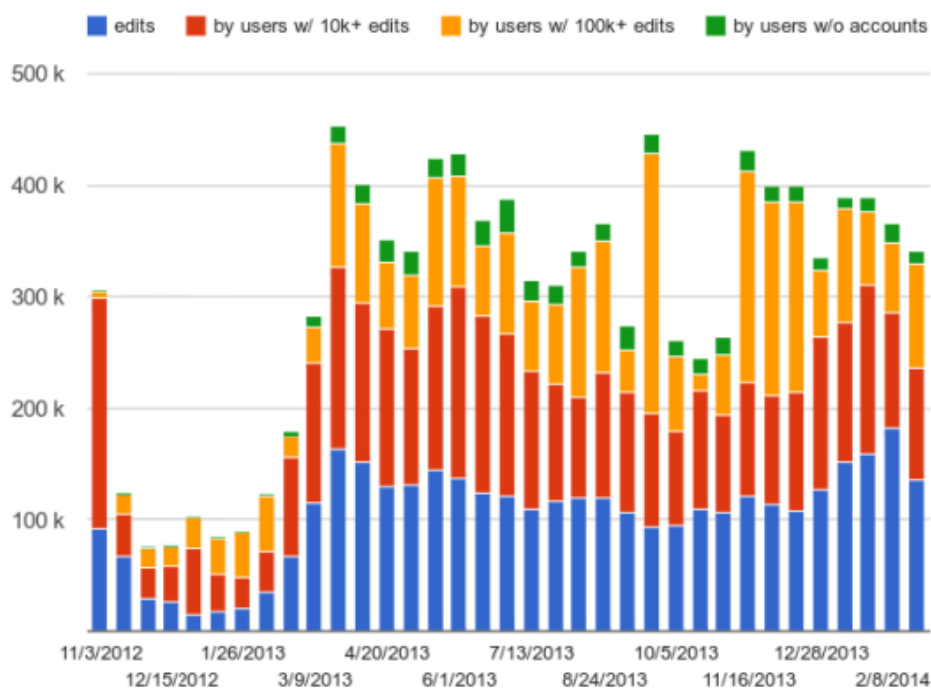
Wikidata ima za cilj prevladati te probleme stvaranjem novog načina na koji Wikipedia globalno upravlja svojim podacima. U početku su urednici mogli samo stvarati entitete i povezivati ih s člancima na Wikipediji. U siječnju 2013. tri Wikipedije, prva na mađarskom, zatim hebrejskom i talijanskom, povezale su se s Wikidatom. U međuvremenu, zajednica je stvorila više od tri milijuna projekata. U veljači je uslijedila engleska Wikipedija, a u ožujku iste godine su se sve Wikipedije povezale s Wikidatom (Vrandečić, Krotzsch, 2014., str. 78. – 83.).

Prvi izazov za Wikidata bio je koordinirati izdanja Wikipedije na 287 jezika. Da bi Wikidata bila doista višejezični projekt, objekt koji predstavlja primjerice grad Rim mora biti isti na svim jezicima. Srećom, Wikipedia već ima blisko povezan mehanizam - jezične veze koje se pojavljuju na lijevoj strani svakog članka, povezujući članke na različitim jezicima. Ove se poveznice sastoje od korisnički uređenih tekstualnih unosa na dnu svakog članka, što rezultira kvadratom broja poveznica: svaki od 207 članaka o gradu Rimu sadrži popis koji sadrži 206 poveznica na sve ostale članke o Rimu. Do kraja 2012. Wikipedia na 66 jezika sadržavala je više teksta jezične veze nego stvarnog sadržaja članka. Naravno, bilo je očito da bi bilo bolje pohranjivati i upravljati jezičnim vezama na jednom mjestu, što je glavni prioritet Wikidata. Za svaki članak na Wikipediji kreirala se stranica na Wikidata koja upravlja poveznicama na povezane članke na Wikipediji na svim jezicima. Takve stranice na Wikipodacima nazivaju se stavke. U početku se za svaku stavku mogla pohraniti samo ograničena količina podataka, kao što je popis jezičnih veza, oznaka, popis aliasa i opis u jednom retku. Zajednica Wikidata stvorila je tako botove za premještanje jezičnih poveznica s Wikipedije na Wikidata, pa je više od 240 milijuna poveznica moglo biti uklonjeno s Wikipedije. Danas se većina jezičnih poveznica prikazanih na Wikipediji poslužuje iz Wikidata. I dalje je moguće dodati prilagođene veze u članak, što je potrebno u rijetkim slučajevima kada veze nisu dvosmjerne jer neki se članci odnose na općenitije članke na drugim jezicima, dok Wikidata namjerno povezuje samo stranice koje pokrivaju istu temu. Uvozom jezičnih veza, Wikidata je dobila ogroman skup početnih stavki koje su, može se reći, utemeljene na stvarnim stranicama Wikipedije (Vrandečić, Krotzsch, 2014., str. 78. – 85.)

2.1.1. Wikidata u brojkama

Wikidata je značajno porasla od svog pokretanja u listopadu 2012. te je s vremenom postala najčešće uređivani Wikimedijin projekt. U ovom poglavlju prikazat će se kako izgleda taj njen rast u brojkama. Prvo će se analizirati podatci iz članka samog osnivača Wikidate, Vrandečića, koji je pratio njen razvoj prvu godinu, a potom će se iznijeti današnji odnosno aktualni podatci sa stranice.

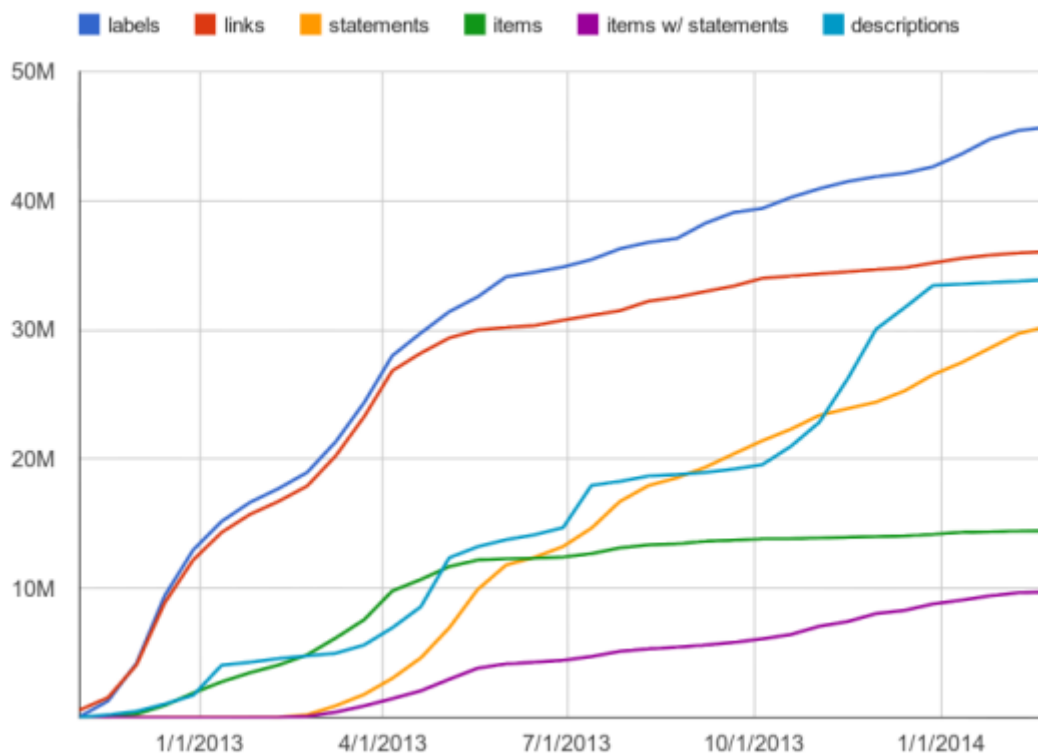
Prema podatcima iz 2013. godine, dakle na samom početku Wikidate, već se može uočiti koliki je njezin doprinos zajednici. Zabilježeno je oko pola milijuna uređivanja podataka dnevno. Otprilike 90% tih izmjena čine botovi koje su suradnici izradili za automatiziranje zadataka, a ipak gotovo milijun izmjena mjesečno naprave ljudi. Na Grafu 1. prikazan je broj ljudskih izmjena tijekom intervala od 14 dana. Ističu se doprinosi naprednih korisnika s više od deset ili sto tisuća uređivanja, a od veljače 2014., oni čine većinu varijacija. Povećanje u ožujku 2013. označava službenu objavu stranice (Vrandečić, Krotzchs, 2014., str. 78. – 81.)



Graf 1. Rast Wikidate prema broju uređivanja različitih grupa

Izvor: Vrandečić, D. & Krotzchs, M. *Wikidata: A Free Collaborative Knowledge Base*, str.

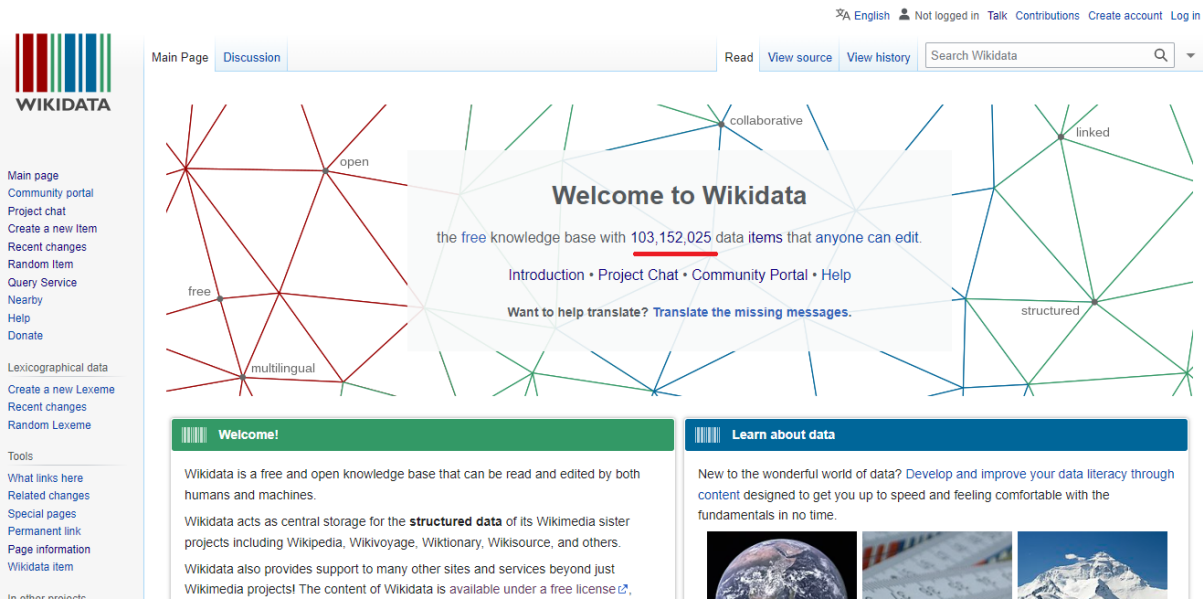
S druge strane, Graf broj 2. prikazuje rast Wikidata baze znanja od njezinog pokretanja do veljače 2014. Prema tadašnjim podacima, postoji oko 14,5 milijuna stavki i 36 milijuna jezičnih veza. Praktično svaki članak na Wikipediji je povezan s Wikidata stavkom, tako da ti brojevi rastu samo sporo. Nasuprot tome, broj oznaka, tada 45,6 milijuna, nastavlja rasti. Moglo bi se reći da već u samom početku Wikidate, postoji više oznaka nego članaka na Wikipediji. Gotovo 10 milijuna stavki ima izjave, a stvoreno je više od 30 milijuna izjava, koristeći preko 900 različitih svojstava (Vrandečić, Krotzsch, 2014., str. 78. – 81.)



Graf 2. Rast Wikidate od osnutka do 2014. godine

Izvor: Vrandečić, D. & Krotzchs, M. *Wikidata: A Free Collaborative Knowledge Base*, str.79.

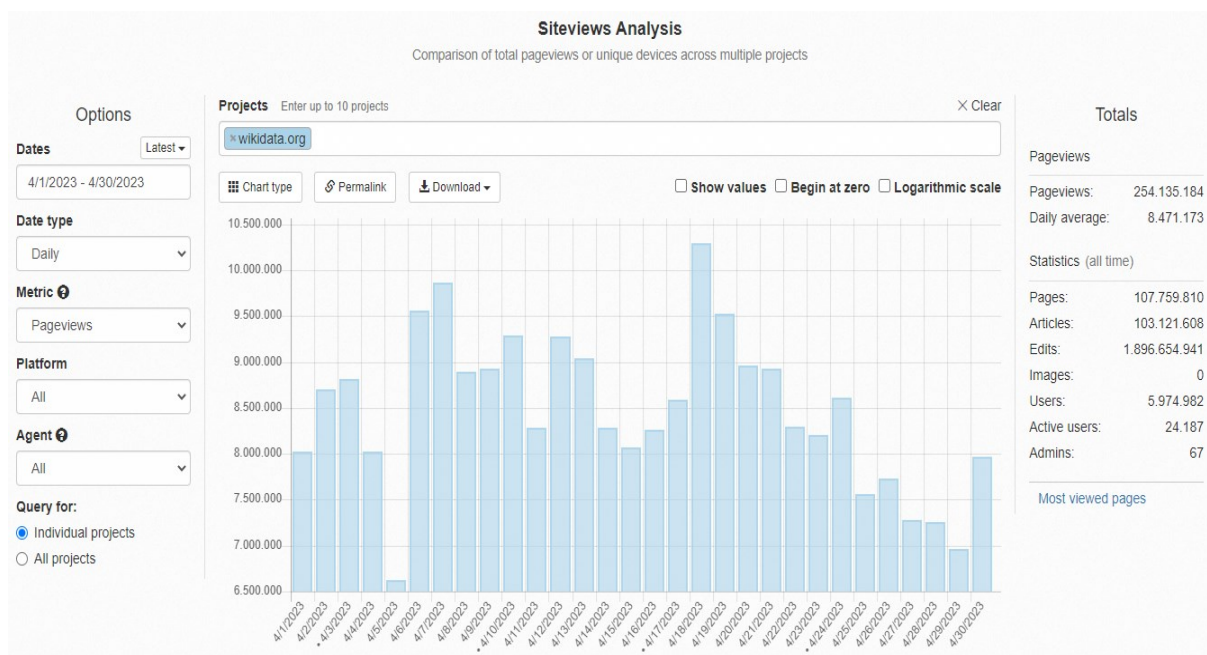
S obzirom na to da je Wikidata od svojeg postanka uvelike napredovala, danas su dostupni razni podatci, ovisno o potrebama i željama naših pretraživanja. Primjerice, na naslovnoj stranici Wikidate, kao što je prikazano na slici 2., klikom na brojku koja se nalazi ispod naslova (podcrtana crvenom bojom) otvara se „Statistika“ (engl. *Statistics*) koja iznosi tablicu s podacima pa je tako zanimljiv podatak o broju stranica sa sadržajem koji iznosi 103.154.255 ili pak broj uređivanja od nastanka projekta Wikidate koji iznosi 1.897.074.377 ("Wikidata", 2023.).



Slika 2. Izgled naslovne stranice s prikazom trenutne brojke sadržaja

Izvor: Wikidata: https://www.wikidata.org/wiki/Wikidata:Main_Page (25.04.2023.)

Također, u odjeljku *informacije o stranici* (engl. *Page information*), prikazane su glavne informacije o stranici kao i analitika stranice. Na slici 3. je tako prikazan primjer analize pregleda stranice, odnosno usporedba ukupnog broja prikaza stranica ili jedinstvenih uređaja u više projekata. S lijeve strane u stupcima moguće je odabrati datum za koji se želi analiza te potražuje li se dnevno ili mjesečno izvješće, kao i potraživanje za individualne ili sve projekte. Dok s desne strane grafa se nalaze ukupni podatci kao što su pregledi stranca, dnevni prosjek, ukupan broj članaka, korisnika, uređivanja i slično. U sredini se naravno nalaze podatci grafički prikazani. Za primjer je uzet 4. mjesec, pa broj uređivanja na Wikidati iznosi 1.896.654.941, a broj korisnika 5.974.982. Također, mogu se saznati i koji entiteti su najviše traženi, a za 4. mjesec pet najposjećenijih su bili: Q95 (Google), Q368215 (Apple Store), Q777 (Google Chrome), Q28561969 (XXXTentacion, američki reper) i Q22663 (The Pirate Bay, stranica za torrente) ("Wikidata", 2023.).



Slika 3. Analiza pregleda stranice za 4. mjesec

Izvor: Wikidata: <https://pageviews.wmcloud.org/siteviews/?platform=all-access&source=pageviews&agent=all-agents&range=last-month&sites=wikidata.org>(1.05.2023.)

2.1.2. 10 godina Wikidate

Relativno mlad projekt Wikimedija zajednice, Wikidata proslavila je svoj 10. rođendan 29. listopada 2022. godine. Ovaj jubilarni događaj obilježili su raznim lokalnim, ali i online događajima diljem svijeta. Glavni cilj proslave bio je slaviti postignuća zajednice, okupiti ljude zajedno, ali i razgovarati o Wikidati kako bi se što veći broj ljudi uključilo. U raznim dijelovima svijeta, ljudi su se okupili kako bi organizirali mnoštvo različitih rođendanskih događaja, kao što su sastanci, radionice, rasprave, prijenosi uživo i slično. Na slici 4. se može vidjeti gdje su događaji sve organizirani, a na stranici Wikidate objavljena je i tablica u kojem su navedeni svi održani događaji. Ljudi su također pripremali rođendanske darove i poruke za zajednicu te stvarali postove na blogovima, video zapise i druge medijske sadržaje. ("Wikidata Tenth Birthday", studeni 2022.)



Slika 4. Mapa događaja proslave 10. rođendana Wikidate

Izvor: Wikidata: https://www.wikidata.org/wiki/Wikidata:Tenth_Birthday (25.04.2023.)

2.2. Funkcija

Wikipedijin sestrinski projekt, Wikidata, je suradnička baza znanja (engl. *Knowledge base*)¹ koja je javno dostupna i sadrži podatke čitljive i ljudima i stroju. Glavni cilj iza razvoja projekta Wikidata je pružiti strukturirane podatke za Wikimedijine projekte kako bi se prevladale nedosljednosti podataka u jezičnim verzijama Wikipedije. Wikidata je osmišljena kako bi se svima omogućilo uređivanje, pregledavanje te korištenje podataka u potpuno višejezičnom obliku. Sadržaj Wikidata, odnosno njezini podatci, također se pohranjuju u formatu grafikona, a oni se daju u strukturiranom obliku u RDF (engl. *Resource Description Framework*) formatu, kojem se može pristupiti pomoću SPARQL5 (engl. *Simple protocol and RDF Query Language*) (Sarbas, Muller-Birn, 2019).

Kako bi se bolje razumjelo na koji način Wikidata radi kratko će se definirati ova dva jezika. Dakle, RDF (engl. *Resource Description Framework*) okvir je za prikaz informacija

¹Baza znanja je centralizirano spremište podataka koje pohranjuje podatke u bilo kojem obliku, kao što je tablični ili grafički

na Web-u. Njegov cilj je omogućiti aplikacijama razmjenu podataka na Web-u bez promjene izvornog značenja tih podataka. Za razliku od HTML-a i XML-a, svrha RDF-a nije predstavljanje podataka na strukturiran način, već daljnja obrada informacija sadržanih u tim podacima te se smatra osnovnim formatom za razvoj semantičkog weba ("Resource Description Framework - RDF", 2004.). Jezik za kreiranje upita temeljen na RDF info zapisima naziva se SPARQL ili *Simple Protocol and RDF Query Language*. Stvorio ga je W3C za kreiranje upita o RDF podacima objavljenim na webu, a jezik je sintaktički sličan jeziku SQL. SPARQL također ima mogućnost stvaranja funkcija za složenije upite i dodavanje uvjeta za filtriranje i formatiranje konačnih rezultata upita (Hitzler, Krotzsch, Rudolph, 2011.).

Kao višejezična baza znanja, Wikidata može pružiti podatke u bilo kojem kontekstu i jeziku, sve dok su dostupni. Zbog toga je, u posljednje vrijeme, došlo do povećanja interesa istraživačke zajednice za Wikidata, što ukazuje na njezinu rastuću popularnost. Mnoge su studije istraživale Wikidata iz različitih perspektiva, kao što je njena unutarnja struktura, iz perspektive podataka promatrajući njihovu cjelovitost i pokrivenost, iz inženjerske perspektive gledajući potrebne alate kao i iz perspektive aplikacije pružanjem studija slučaja korištenja Wikidata za projekte u medicini, lingvistici ili geografiji (Farda Sarbas, Muller-Birn, 2019.). Također, Wikidata postaje sve popularnija u knjižnicama kao otvorena i suradnička globalna platforma za dijeljenje i razmjenu knjižničnih metapodataka. Knjižnice su počele eksperimentirati s Wikidata kako bi poboljšale globalni doseg i pristup svojim jedinstvenim i istaknutim zbirkama i znanstvenicima. Mnogo više od tehnologije, Wikidata pruža način knjižničarima da prihvate društvenu pravdu i prakticiraju knjižničarstvo na globalnoj razini za javno dobro (Tharani, 2021.)

Iz svega navedenog, može se reći da su glavni ciljevi Wikidate slijedeći (Farda Sarbas, Muller-Birn, 2019.):

- a) biti otvorena za uređivanje bilo kome (s ili bez Wikidata računa),
- b) dopustiti suprotstavljenim idejama da koegzistiraju,
- c) pružiti podatke u jezično neovisnom obliku,
- d) biti pod kontrolom zajednice koja doprinosi,
- e) osigurati podatke s referencama, i
- f) kontinuirano se razvijati kako bi odgovorila na nove potrebe korisnika i suradnika.

2.3. Obilježja Wikidata

Glavna obilježja Wikidate, vidljiva već pri njenom definiranju, su slijedeća: ("Wikidata", 2023.)

- a) **besplatna** – kako se kroz rad već spomenulo, Wikidata je besplatna baza podataka. Podatci u Wikidati su objavljeni pod *Creative Commons Public Domain Dedication 1.0.*, dopuštajući ponovnu upotrebu podataka u mnogim različitim scenarijima. Podatci se mogu kopirati, mijenjati, distribuirati i izvoditi, čak i u komercijalne svrhe, bez traženja dopuštenja. Dakle, potpuno besplatno.
- b) **suradnička** – podatci se unose i održavaju od strane urednika Wikidata, koji odlučuju o pravilima stvaranja i upravljanja sadržajem. Automatizirani botovi također unose podatke u Wikidata.
- c) **višejezična** - uređivanje, konzumiranje, pregledavanje i ponovna uporaba podataka potpuno je višejezično. Podatci uneseni na bilo kojem jeziku odmah su dostupni na svim drugim jezicima. Uređivanje na bilo kojem jeziku je moguće i poticajno.
- d) **sekundarna baza podataka** - Wikidata ne bilježi samo izjave, već i njihove izvore i veze s drugim bazama podataka. To odražava raznolikost dostupnog znanja i podržava pojam provjerljivosti.
- e) **prikupljanje strukturiranih podataka** - nametanje visokog stupnja strukturirane organizacije omogućuje jednostavnu ponovnu upotrebu podataka od strane Wikimedijinih projekata i trećih strana te omogućuje računalima da ih obrađuju i "razumiju".
- f) **podrška za wikije Wikimedie** - Wikidata pomaže Wikipediji s informativnim okvirima koji se lakše održavaju i poveznicama na druge jezike, čime se smanjuje radno opterećenje uređivanja uz poboljšanje kvalitete. Ažuriranja na jednom jeziku dostupna su svim drugim jezicima.
- g) **dostupnost** – bilo tko, bilo gdje na svijetu može koristiti Wikidata i na bilo koji broj različitih načina korištenjem sučelja za programiranje aplikacija.

2.4. Kako radi Wikidata

Kako je prikazano u prethodnim poglavljima, Wikidata je tekući projekt koji se aktivno razvija. Radi se, dakle, o središnjem spremištu za pohranu podataka kojem svi mogu pristupiti. Naravno, u budućnosti, kako se Wikidata konstantno razvija i raste, će biti dostupno više tipova podataka kao i proširenja. U nastavku će se prikazati kako su ti podaci spremljeni u repozitoriju Wikidata, odnosno na koji način se pretražuju informacije na istom.

Što se tiče informacija koje se mogu dobiti, primjerice Wikipedia je za enciklopedijski sadržaj, Wikimedia Commons je spremište medijskih datoteka, a Wiktionary pruža definicije i leksičke informacije o riječima. Za razliku od njih, u Wikidata, fokus su strukturirani podaci. To omogućuje ljudima i računalima da podjednako koriste podatke. Strukturirani podaci također otvaraju mnoštvo nevjerojatnih prilika o kojima će se nešto više reći u nastavku. Da bi se mogao prikazati način na koji se pretražuju podaci na Wikidati, odnosno njezinom repozitoriju, treba se objasniti nekoliko pojmova koji čine njezin sastavni dio, a to su: ("Wikidata", 2023.)

a) **entiteti/stavke** (engl. *items*)

Kako bi se podržalo nešto kao svo znanje dostupno na Wikipediji, treba način pohranjivanja prikaza tog znanja. Ove reprezentacije znanja nazivaju se entitetima. Entiteti su dovoljno fleksibilni da predstavljaju apstraktne koncepte poput *djetinjstva*, *gladi* i *težine*, kao i realne objekte iz stvarnog svijeta poput *televizora*, *kajaka* i *vulkana*. Svaki entitet ima vlastitu stranicu gdje se prikupljaju svi podaci o njemu i svoj jedinstveni identifikator. Ovaj identifikator se sastoji od prefiksa **Q** iza kojeg slijede redni brojevi pa on uvijek izgleda slično **Q####**. Iako je koristan za strojeve i za predstavljanje znanja na mnogo različitih jezika, ovaj identifikator nije baš prilagođen ljudima.

U Wikidata, **entiteti** se koriste za predstavljanje svih stvari u ljudskom znanju, uključujući teme, koncepte i objekte. Na primjer, "*Ljetne olimpijske igre 1988*", "*ljubav*", "*Elvis Presley*" sve su stavke u Wikidata. Za one koji su upoznati s Wikipedijom, može se na prvi pogled činiti da su entiteti Wikidateina verzija Wikipedijinih članaka. Iako su i jedni i drugi stranice za pohranu informacija o različitim konceptima ili temama ljudskog znanja, važno je imati na umu da Wikidata nije samo baza podataka sadržaja Wikipedije. Zapravo, Wikidata je sličnija Wikimedia Commons-u, spremištu medijskih datoteka, nego Wikipediji, ali umjesto

medijskih datoteka, Wikidata pruža centraliziranu pohranu za pristup strukturiranim podacima za sve Wikimedijine projekte, kao i za korištenje na vanjskim web stranicama.

b) izjave (engl. *statements*)

Služe za opisivanje detaljne karakteristike entiteta i sastoje se od **svojstva** (eng. *property*) i **vrijednosti** (eng. *value*). Svojstva u Wikipodacima imaju prefiks **P** iza kojeg slijedi redni broj. Nadalje, za osobu može se dodati svojstvo da se navede gdje se školovala, navodeći vrijednost za školu. Za zgrade se mogu dodijeliti svojstva geografskih koordinata navođenjem vrijednosti zemljopisne dužine i širine. Svojstva se također mogu povezati s vanjskim bazama podataka. Svojstvo koje povezuje stavku s vanjskom bazom podataka, kao što je baza podataka ovlaštene kontrole koju koriste knjižnice i arhivi, naziva se identifikator. Posebne veze do web-mjesta povezuju stavku s odgovarajućim sadržajem na klijentskim wikijima, poput Wikipedije, Wikiknjiga ili Wikicitata. Naravno, sve ove informacije mogu se prikazati na bilo kojem jeziku, čak i ako podaci potječu iz drugog jezika.

Za primjer pretraživanja repozitorija Wikidata uzet je grad Zagreb. Slika 5. prikazuje kako je na Wikidata zapisano da je Zagreb grad. Dakle, može se vidjeti oznaka entiteta odnosno stavke, **Zagreb** ima oznaku **Q1435**. Svojstva imaju oznaku slovo P, u ovom slučaju je to **primjerak od** (engl. *instance of*) oznake **P31** i vrijednost glavnog grada (engl. *capitalcity*) **Q5119**.

The screenshot shows the Wikidata page for Zagreb (Q1435). The page title is 'Zagreb (Q1435) oznaka stavke'. Below the title, there is a search bar and a list of labels in different languages. The 'Statements' section shows a statement 'Instance of' with the property 'P31' (instance of) and the value 'Q5119' (capital city).

Language	Label	Description	Also known as
English	Zagreb	capital city of Croatia	Zagreb, Croatia Agram City of Zagreb
Croatian	Zagreb	glavni grad Hrvatske	Agram
Bosnian	Zagreb	glavni grad Hrvatske	

Slika 5. Primjer entiteta Zagreb na Wikidata repozitoriju

Izvor: Wikidata: <https://www.wikidata.org/wiki/Q1435> (19.04.2023.)

Postoji nekoliko važnih karakteristika stavki koje pridonose uspjehu Wikidata ("Wikidata", bez dat.):

- primjetnost

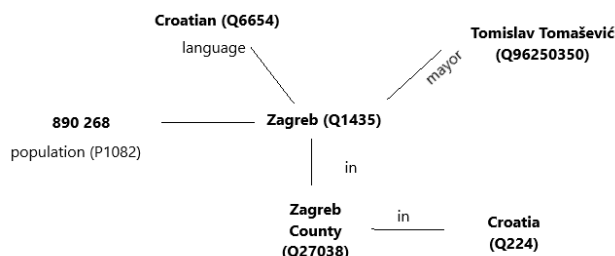
Za entitete Wikidata kaže se da su primjetni. To znači barem jednu od tri stvari: odgovarajuću stranicu na Wikimedijinoj stranici kao što je Wikipedia, Wikivoyage, Wikisource, Wikiquote, Wikiversity ili WikimediaCommons; jasno prepoznatljiv konceptualni ili materijalni entitet koji se može opisati korištenjem ozbiljnih i javno dostupnih referenci ili predmet ispunjava strukturnu potrebu.

- unikatnost

Svaki entitet treba predstavljati jasno prepoznatljiv koncept ili objekt, ili instancu jasno prepoznatljivog koncepta ili objekta. Na primjer, u Wikidata postoji entitet i za koncept planeta (Q634) i za instancu planeta, Zemlje (Q2).

- povezanost entiteta

Entiteti mogu i trebaju biti povezani jedan s drugim. Svaki sadrži jedinstveni identifikator (koji počinje s prefiksom Q) i vlastitu stranicu u glavnom imenskom prostoru Wikidata. Na tim se stranicama dodaju, uređuju i održavaju svi podaci za svaku stavku, uglavnom u obliku izjava. Stranice također dopuštaju njihovo međusobno povezivanje kako bi svi podaci na Wikidata mogli biti međusobno povezani. Veza na drugu stranicu obično se dodaje kao "svojstvo" (engl. *property*) klikom na "uredi" (engl. *edit*) u odjeljku "izjave" (engl. *statements*). Ako odgovarajuće svojstvo još nije navedeno, može se dodati klikom na "+ dodaj" (engl. *add*) na dnu odjeljka "izjave".



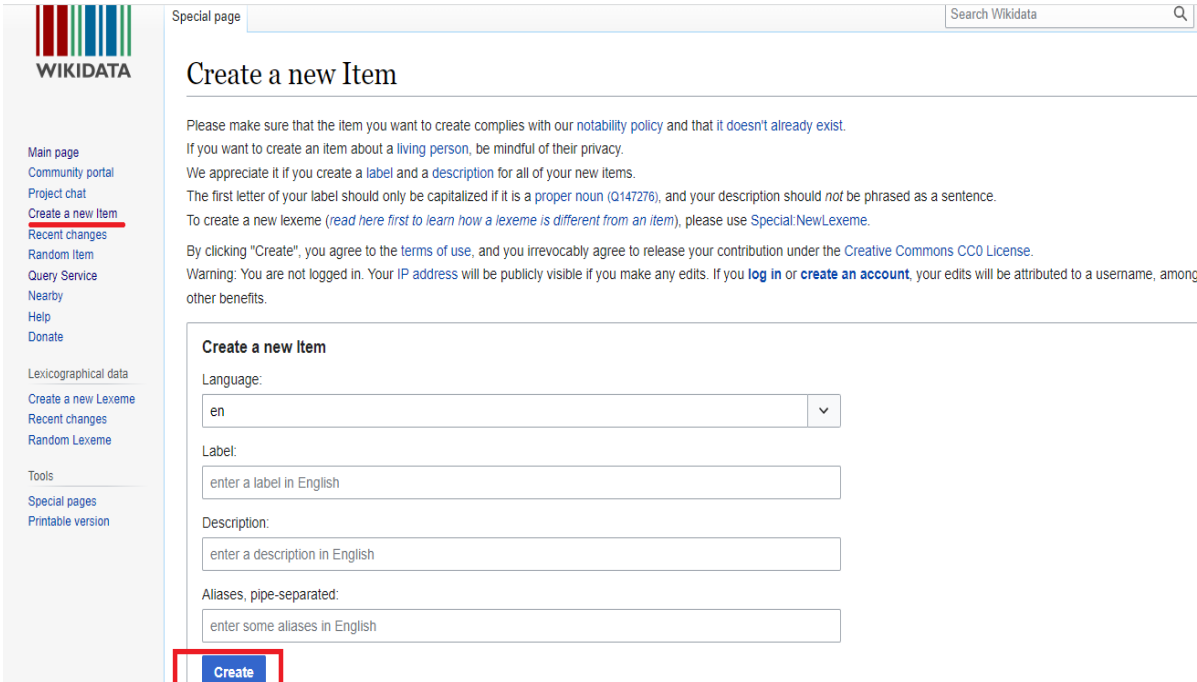
Slika 6. Primjer povezanosti podataka na Wikidati

Izvor: autorovo djelo, prema podacima sa Wikidata

2.5. Unos podataka na Wikidatu

Što se tiče unosa podataka na Wikidatu, radi se o jednostavnom procesu. Kako je već rečeno kroz rad, bilo tko može dodavati i mijenjati postavljeni sadržaj. Proces postavljanja podataka je slijedeći:

- na stranici, kao što je prikazano na slici 7., u lijevom gonjem kutu, podcrtano crvenom bojom, potrebno je kliknuti „stvoriti novi entitet“ (engl. *create a new item*)
- otvara se novi prozor u kojem je potrebno upisati podatke, kao što je prikazano na slici
- kada se unesu podatci, pritiskom na tipku „stvoriti“ (engl. *create*) otvara se novi prozor gdje se pritiskom na „dodati izjavu“ (engl. *add statement*) može dodati još podataka i na taj način povezati sadržaj s ostalim entitetima i relacijama u bazi podataka
- u konačnici se dobije pridružena oznaka Qxxx



Special page

Create a new Item

Please make sure that the item you want to create complies with our [notability policy](#) and that it *doesn't already exist*.
If you want to create an item about a [living person](#), be mindful of their privacy.
We appreciate it if you create a [label](#) and a [description](#) for all of your new items.
The first letter of your label should only be capitalized if it is a [proper noun](#) (Q147276), and your description should *not* be phrased as a sentence.
To create a new lexeme ([read here first to learn how a lexeme is different from an item](#)), please use [Special:NewLexeme](#).
By clicking "Create", you agree to the [terms of use](#), and you irrevocably agree to release your contribution under the [Creative Commons CC0 License](#).
Warning: You are not logged in. Your [IP address](#) will be publicly visible if you make any edits. If you [log in](#) or [create an account](#), your edits will be attributed to a username, among other benefits.

Create a new Item

Language:

Label:

Description:

Aliases, pipe-separated:

Slika 7. Podatci koji se moraju unijeti na Wikidata

Izvor: autorovo djelo

2.6. Sukobi mišljenja na Wikidati

S obzirom da uređivati na Wikidati može bilo tko, dolazi do potencijalnog problema, odnosno neslaganja dva urednika oko izjave. Primjerice, Wikipedia rješava takve sukobe golemim skupom pravila i mehanizama, pri čemu je glavni alat stranica za razgovor (engl. *talk page*). Svaki članak na Wikipediji dolazi sa stranicom za razgovor, a za sukobe stranica za razgovor je prvo mjesto za raspravu o problemima i rad na postizanju konsenzusa. Baš poput Wikipedije, Wikidata nudi stranicu za razgovor za svaku stavku i svako svojstvo, ali zbog raznolikosti jezika koji se koriste u Wikidatu, ovaj alat se baš i ne koristi. Ako dva urednika bez zajedničkog jezika žele razgovarati o nekoj izjavi, moraju nekako premostiti jezičnu barijeru. To predstavlja ozbiljnu prepreku održivosti projekta. Wikidata ipak pokušava ublažiti ovaj problem drugim mehanizmima. Što je najvažnije, u Wikidatu se ne bave istinom. Baš kao i u Wikipediji, glavna mjera za uključivanje izjave u projekt nije istinitost, već provjerljivost. To se provodi kroz dva mehanizma: reference i mogućnost pohranjivanja iskaza koji su proturječni. Dok su mnoge izjave prilično nekontroverzne, primjerice "*Mozart je rođen u Salzburgu*", za mali broj izjava ključno je da imaju dobar izvor. Ovo podiže raspravu o izjavi od toga je li izjava doista činjenično točna do toga može li se potkrijepiti prihvatljivim izvorom, što uvelike smanjuje, ali očito ne eliminira, potrebu za raspravama (Vrandečić, 2013., str. 92.).

Wikidata može sadržavati proturječne izjave. Primjerice, populacija Mumbaija može se navesti i kao 12,5 milijuna, prema objavljenim brojevima popisa indijskog Zavoda za statistiku, i kao 20,5 milijuna, prema procjeni Ujedinjenih naroda. Nema potrebe da zajednica Wikidata odluči o pravom broju, može jednostavno ponuditi oboje, zajedno s odgovarajućim izvorima. Sukob oko uređivanja dogodio se i oko podatka o Jeruzalemu. Tako se opis Jeruzalema (Q1218) često mijenja od spominjanja njegovog statusa glavnog grada Izraela ili Palestine, oba ili nijednog. Budući da se ova uređivanja odnose samo na engleski opis, a ne na strukturirane podatke, sa sigurnošću možemo pretpostaviti da jezik nije prepreka koja koči rješavanje ovog konkretnog problema. Za razliku od Wikipedije, Wikidata ne nudi mogućnost pomirenja i objašnjenja ovih razlika u prozi, dajući odgovarajuću težinu različitim stajalištima. Odgovornost je na čitatelju i korisniku Wikidata da odluče kojim izvorima vjerovati (Vrandečić, 2013., str. 92.).

2.7. Prednosti i mane Wikidatae

Nakon definiranja i malo šireg uvida u područje rada Wikidate, može se reći da postoje neki aspekti koji se svakako ističu, a mogu predstavljati ili prednost ili manu za korisnika. Prije svega, kao što je već naglašeno, Wikidata predstavlja besplatnu i otvorenu bazu znanja koju mogu čitati i uređivati i ljudi i strojevi. Prednost je svakako što je besplatna i otvorena, ali mogućnost uređivanja od strane bilo koga predstavlja određenu dozu rizika. S druge strane Wikidata sadrži različite vrste podataka naprimjer tekst, slike, količine, koordinate, zemljopisne oblike, datume, što za neke korisnike stvara zbrku, ali i koristi SPARQL, koji može predstavljati problem za neke korisnike, pogotovo ako ga nikada prije nisu koristili, odnosno ako na niti jedan način nisu upoznati s njim ("KD nuggets", 2018).

Jedna od prednosti Wikidata je oslanjanje na zajednicu odnosno društvo da uređuje i održava njen sadržaj. To je dovelo do toga da je od početka nastanka projekta broj registriranih urednika na Wikidati rastao sve više i više, što znači da i broj entiteta odnosno stavki raste. Ovo je svakako prednost u odnosu na slične projekte, kao što je DBpedia, gdje se podaci povremeno izvlače iz Wikipedije i prvo moraju biti izmijenjeni na internetskoj enciklopediji kako bi se ispravili. Još jedna prednost je ta da se svi podaci u Wikidata mogu otvoreno ponovno koristiti i dijeliti bez potrebe za ikakvim pripisivanjem, budući da su objavljeni pod CC0 licencom. Zbog toga Wikidata može djelovati kao izvor za gotovo neograničen broj sustava i aplikacija temeljenih na informacijama ("KD nuggets", 2018).

Utjecaj zajednice, odnosno društva, proteže se čak i na tehnički razvoj web stranice i temeljnog softvera Wikidate. Ona se temelji na otvorenom razvojnom procesu koji poziva na doprinose, a sama stranica nudi mnoga proširenja za dodatke koje su izradili korisnici. Zajednica je dizajnirala i razvila različite značajke sučelja, npr. za ugrađivanje slika i uređivanje na više jezika. Vanjski alati prikupljaju ove informacije, analiziraju skup podataka radi kršenja ograničenja i objavljuju popis kršenja na Wikipodacima kako bi urednici mogli provjeriti jesu li valjane iznimke ili pogreške (Vrandečić, Krotzsch, 2014., str. 5).

Iako Wikipedia predstavlja jedno od najvažnijih web-mjesta današnjice, Wikidata je, premda dosta mlađa, već postala važna platforma za integraciju informacija iz mnogih izvora. Pored ovih primarnih podataka, Wikidata također prikuplja velike količine slučajnih metapodataka o vlastitoj evoluciji i utjecaju na Wikipediju. Wikidata stoga ima potencijal postati glavni resurs za istraživanje i razvoj novih i poboljšanih aplikacija. Wikidata, kao besplatna baza

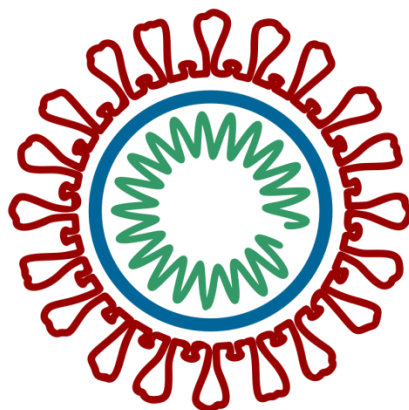
znanja koju svatko može uređivati, može postati najznačajniji izvor za svaku vrstu pretraživanja (Vrandečić, Krotzsch, 2014., str.5).

2.8. Wikidata projekti

Zajednica Wikidate se neprestano razvija i raste. Kao rezultat toga, pokrenuti su razni projekti koji omogućavaju lakše korištenje odnosno služenje podacima. U nastavku će se, za primjer, ukratko prikazati dva relativno nova Wikiprojeka, a to su Wikiprojekt COVID19 te Wikiprojekt Music.

2.8.1. Wikiprojekt COVID19

Pandemija korona virusom, odnosno COVIDOM 19, zatekla je cijeli svijet. Na život u novom normalnom, tri godine nakon, svi su se već priviknuli, no u početku nije bilo lako. Naročito kada su vijesti donosile zastrašujuće podatke o broju zaraženih i umrlih. Upravo jedan od najvećih doprinosa Wikidata zajednice je istaknuti WikiProject COVID-19, a njegov logo prikazan je na slici ispod.



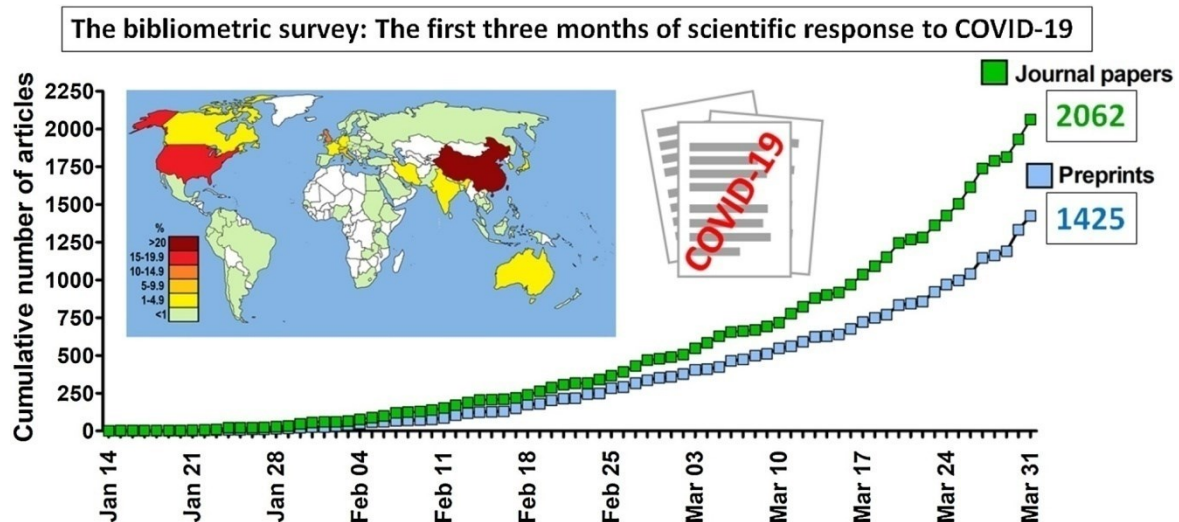
Slika 8. WikiProject COVID-19 logo

Izvor: ByGeraki - Ownwork, CC BY-SA 4.0,

<https://commons.wikimedia.org/w/index.php?curid=88248113> (19.04.2023.)

Dok je diljem svijeta aktualna borba protiv pandemije COVID-19, urednici Wikidata su bili iznimno korisni u prikupljanju i modeliranju korisnih informacija vezanih uz COVID-19 pa

je tako nastao WikiProject COVID19. Tako su prikupljali od relevantnih istraživačkih članaka i kliničkih ispitivanja, do epidemioloških statistika raznih regija, do leksema za riječi i izraze koji su uobičajeni povezano s pandemijom i još mnogo toga. Primjerice, na slici je prikazan grafički pregled znanstvenih publikacija o COVIDU-19 tijekom prva tri mjeseca 2020. godine. Puno rada WikiProjecta COVID-19 ulaže se u stvaranje i održavanje ovih stavki, jačajući ulogu Wikidata kao korisnog otvorenog izvora u pandemiji ("Wikidata", 2023.)



Slika 9. Pregled znanstvenih publikacija o COVIDU-19 tijekom prva tri mjeseca 2020. godine

Izvor: ByNowakowska, Joanna; Sobocińska, Joanna; Lewicki, Mateusz; Lemańska, Żaneta; Rzymiski, Piotr (2020). *When science goes viral: Their search response during three months of the COVID-19*. Biomedicine & Pharmacotherapy:

<https://commons.wikimedia.org/w/index.php?curid=91572702>(19.04.2023.)

Glavni zadatci Wiki Projekta su ("Wikidata Wiki Project COVID19", 2023.):

- dodavanje epidemioloških informacija, kao što su broj umrlih, broj zaraženih, broj preboljenih, broj medicinskih pretraga, karta distribucije, broj cijepljenja, tablični podatci..
- pomoć u prikupljanju informacija o literaturi o COVIDU-19 putem Scholie
- lista želja za bota – pronaći i programski dodati podatke za entitete koji nisu zemlje (tj. kontinenti ili države), pronaći i programski dodati podatke o cijepljenjima ili

testovima (trenutno nije obuhvaćeno), postaviti bot na server (trenutno je na osobnom računalu)

Trenutno (prema posljednjim podacima u travnju 2023.) postoji 2430 članaka u okviru Wikiprojekta COVID-19. Uključujući stranice bez članaka, kao što su stranice za razgovor, preusmjeravanja, kategorije, stranice projekta itd., u projektu ima 9032 stranice. ("WikiProject COVID-19", 2023.)

2.8.2. WikiProject Music

WikiProject Music dom je urednika koji pomažu u dodavanju podataka o izvođačima, glazbenim izdanjima, pjesmama, nagradama i nastupima. Također, uvoz iz i povezivanje Wikidata s mnogim glazbenim bazama podataka i uslugama strujanja još je jedan fokus projekta ("WikiProject Music", 2023.)



Slika 10. Logo WikiProject Music

Izvor: Wikidata:

https://www.wikidata.org/wiki/Wikidata:WikiProject_Music#/media/File:Wikidata-music-logo-en.svg (22.04.2023.)

Glavni ciljevi WikiProject Musica su slijedeći ("WikiProject Music", 2023.)

- definirati skup svojstava koje mogu koristiti glazbene *infoboxove*
- mapirati i uvesti sve relevantne podatke koji su trenutno raspoređeni u Commonsu, Wikipediji i Wikisourceu
- uspostaviti metode za interakciju s tim podacima iz različitih izvora

3. ODNOS WIKIPEDIE I WIKIDATE

O popularnosti Wikipedije nije potrebno previše govoriti. Svatko tko se služi internetom barem je jednom posjetio stranicu Wikipedije i koristio se njezinim podacima. Kao besplatna i otvorena enciklopedija znanja, predstavlja pregršt mogućnosti i opcija. U nastavku će se ukratko definirati Wikipedia sa svojim glavnim karakteristikama i manama, ali i ono važnije, njezin odnos sa sestrinskim projektom Wikidatom, te prikaz glavnih razlika između ova dva Wikimedijina projekta, kao jedan od ciljeva rada.

3.1. Ukratko o Wikipediji

Wikipedija (engl. *Wikipedia*), čiji je logo prikazan na slici 11., je besplatna, online enciklopedija kojom se svatko može služiti i koju svatko može uređivati. Projekt je Wikimedia Zaklade, a od svog pokretanja 15. siječnja 2001., izrasla je u najveću svjetsku referentnu stranicu, privlačeći više od milijardu posjetitelja mjesečno. Trenutačno sadrži više od 61 milijun članaka na više od 300 jezika, uključujući 6.657.661 članak na engleskom jeziku, sa 125.164 aktivnih suradnika u proteklom mjesecu. Svrha Wikipedije je koristiti čitateljima tako što sadrži informacije o svim granama znanja, a sastoji se od sadržaja koji se može slobodno uređivati, a njezini članci također imaju mnogo poveznica koje čitatelje usmjeravaju na više informacija. Napisana je u suradnji s uglavnom anonimnim volonterima, svatko s pristupom internetu i tko trenutno nije blokiran, može pisati i uređivati članke na Wikipediji, osim u ograničenim okolnostima gdje je uređivanje ograničeno kako bi se spriječio vandalizam ("Wikipedia", 2023.)



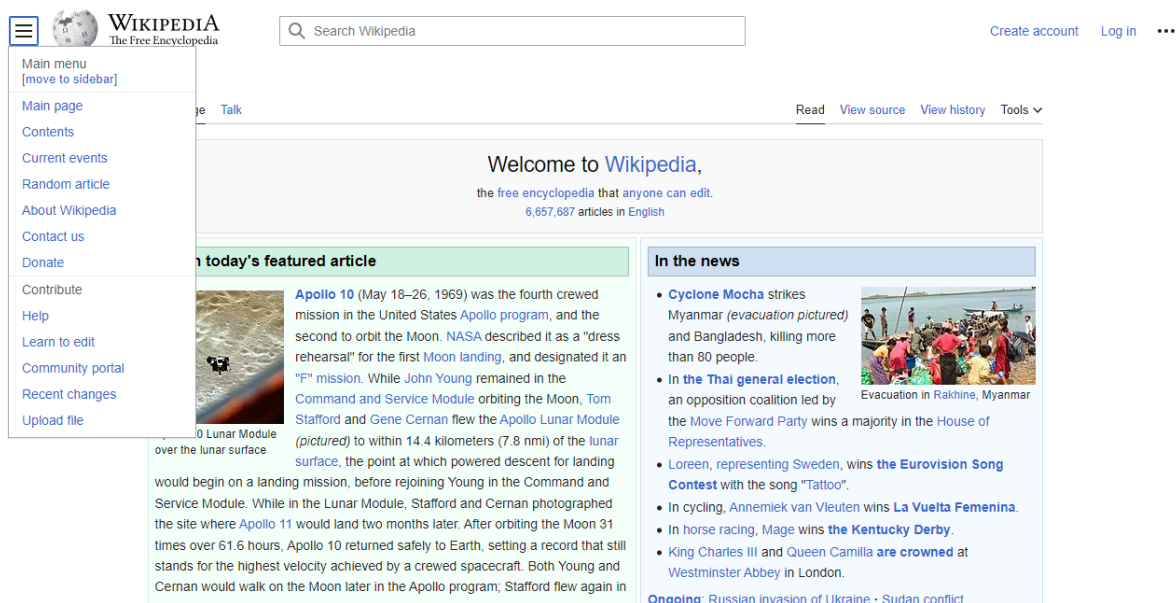
Slika 11. Logo Wikipedije

Izvor: <https://www.wikipedia.org/> (10.05.2023.)

Drugim riječima, projekt Wikipedija predstavlja najveću bazu podataka na svijetu, čiji je domaćin neprofitna Zaklada Wikimedia, koja tisućama volontera diljem svijeta pruža infrastrukturu za stvaranje i uređivanje stranica. Kao otvorena, besplatna i lako dostupna, postala je "gotova referenca" za današnje studente. Kao dokaz, istraživanja objavljena u posljednjih pet godina, pokazuju da je to jedan od najčešće korištenih izvora informacija ili pak polazište za istraživanje, kako za studente tako i za profesore. Činjenica da je grade korisnici za korisnike, čini je dinamičnim nastavnim alatom koji se može koristiti samostalno ili u učionici, online prostoru gdje učenici doprinose i primaju informacije. Studenti mogu pregledati povijest stvaranja i uređivanja Wikipedijinih članaka, ali i samostalno uređivati stranice. Popularnost Wikipedije i njezina neprofitna misija čine je očitim pedagoškim alatom za podučavanje informacijske pismenosti u akademskim knjižnicama (Dowell, Bridges, 2019., str. 81 - 83).

Wikipedia se razlikuje od tiskanih referenci na važne načine. Stalno se stvara i ažurira, a enciklopedijski članci o novim događajima pojavljuju se u minutama, a ne mjesecima ili godinama. Međutim, postoji mnogo pravila i smjernica koje je razvila zajednica Wikipedije, iako urednici ne moraju biti upoznati s njima prije nego što daju doprinos. Svatko može uređivati Wikipedijin tekst, reference i slike. Važnije je što je napisano nego tko je napisao. Sadržaj mora biti u skladu s pravilima Wikipedije, uključujući mogućnost provjere objavljenih izvora. Urednička mišljenja, uvjerenja, osobna iskustva, istraživanja koja nisu recenzirana, klevetnički materijal i sadržaj koji krši autorska prava neće biti sačuvani. Wikipedijin softver olakšava uklanjanje pogrešaka, a iskusni urednici promatraju i patroliraju u potrazi za pogrešnim izmjenama ("Wikipedia", 2023.).

Wikipedija ima jednostavno sučelje za korištenje. Na slici 12. prikazana je naslovna stranica Wikipedije, na engleskom jeziku. Na naslovnoj stranici mogu se vidjeti dnevni istaknuti članci, događaji koji su se dogodili na taj dan, zanimljivosti te aktualne novosti. Primjerice, od jedna od novosti je članak da je predstavnic Švedske, Loreen, pobijedila na Eurosongu sa pjesmom „Tattoo“. S lijeve strane nalazi se ikona sa tri crtice koja predstavlja sadržaj Wikipedije. Klikom na tu ikonu otvara se njezin glavni sadržaj koji otvara druge poveznice poput sadržaja, trenutnih događaja, o Wikipediji, kako se može uređivati i slično. Na samom vrhu, u desnom kutu se nalaze ikone za kreiranje računa (engl. *create account*) odnosno prijavljivanje na Wikipediju (engl. *log in*). U svakom slučaju, Wikipedija predstavlja jednostavnu stranicu kako za korištenje tako za uređivanje ("Wikipedia", 2023.).



Slika 12. Naslovna stranica Wikipedije

Izvor: Wikipedia: https://en.wikipedia.org/wiki/Main_Page (18.05.2023.)

3.2. Utjecaj Wikidata na Wikipediju

Kako je već spomenuto u radu, projekt Wikidata je stvoren imajući na umu neke od problema s kojima se suočavala Wikipedija. Drugim riječima, Wikidata je Wikimedijino spremište strukturiranih podataka (Sahoo, Abdulhamid, Peel, 2021). Dakle, jedna od prvih funkcija Wikidata je da služi kao centralizirano središte za međujezične poveznice na Wikipediji. Wikidata-ovi identifikatori neovisni o jeziku eliminiraju potrebu za lokaliziranim verzijama stavki i svojstava, tako da se mogu pohraniti sve poveznice na članke Wikipedije u različitim jezičnim verzijama. Zbog toga, za svaki članak na Wikipediji postoji entitet na Wikidati, ali ne i obrnuto, tj. svaki entitet na Wikidati možda nema odgovarajući članak na Wikipediji (Piscopo, Vougiouklis, Kaffee, Phethean, Hare, Simperl, 2017).

Dakle, Wikidata je izvorno zamišljena kao strukturna okosnica Wikipedije, kako bi se poboljšala dosljednost između različitih izdanja online enciklopedije i smanjio urednički posao. Stoga su dva projekta međusobno povezana na više razina. Podaci se aktivno dijele i ponovno koriste između dviju platformi, na primjer Wikidata se koristi za popunjavanje infokutija Wikipedije strukturiranim podacima. (Piscopo i sur., 2017). Međutim, iako

Wikidata brzo raste otkako je osnovana 2012., još uvijek je daleko iza članaka na Wikipediji u smislu količine podataka koje pohranjuju. Konkretno, engleska Wikipedija ima puno polustrukturiranog sadržaja u obliku infokutija, popisa i uvodnih odjeljaka. Za razliku od Wikipedija na drugim jezicima, još nije došlo do značajne masovne migracije ovog sadržaja na Wikidata. Iako se ove informacije čuvaju samo na engleskom, umjesto da budu u višejezičnim strukturiranim podacima Wikidata, ne mogu ih lako koristiti druge Wikipedije s većom integracijom Wikidata ili vanjske web stranice koje koriste Wikidata podatke. Premda se ovim bave volonteri koji uvoze podatke ručno ili uz pomoć skripti, u Wikimedijinim projektima nema roka, tako da ovaj proces može trajati dugo (Sahoo i sur., 2021).

Kategorije Wikipedije igraju značajnu ulogu u organiziranju članaka po temama. Naime, one čine hijerarhiju koja grupira srodne članke u veće zbirke. S druge strane, Wikidata osigurava odgovarajući entitet odnosno stavku za svaku kategoriju i omogućuje definiranje pripadnosti drugih entiteta određenoj kategoriji pomoću SPARQL upita ili određivanjem klasa i svojstava (Feddoul, Loffler, Schindler, 2022).

3.3. Glavne sličnosti i razlike

Iz svega navedenog kroz prethodna dva poglavlja, mogu se izdvojiti odnosno sumirati, glavne sličnosti i razlike između ova dva Wikimedijina projekta. Glavna, vidljiva na prvi pogled, razlika je upravo format informacija. Znanje Wikidata kodirano je u modelima parova svojstvo - vrijednost. Iako ovaj format možda nije privlačan ljudima za čitanje, on omogućuje strojevima da obrađuju podatke i lako izvlače prilagođene informacije. Nasuprot tome, članci Wikipedije možda su bolji za čitanje ljudi, ali prirodni jezik, odnosno prozni tekst, na kojem su napisani teško je automatski obraditi. Što se tiče samih sučelja odnosno prikaza stranica, moglo bi se reći da je Wikipedija, na prvi pogled, jednostavnija za korištenje u smislu pronalaska kategorija, podataka i slično. S druge strane, ukoliko želimo saznati neki konkretni podatak, primjerice koliko stanovnika ima jedan grad, Wikidata je bolji izbor zbog konkretnog odgovora, odnosno samog podatka, a ne proznog teksta. Sličnost je i postojanje stranice za razgovor (engl. *talk page*), dakle urednici obadviju stranica mogu komunicirati putem stranica za razgovor, međutim, one se na Wikidata rijetko koriste s obzirom na jezičnu raznolikost sustava (Piscopo i sur., 2017).

Jedna od razlika je i u tome što je Wikidata definirana kao sekundarna baza podataka, koja prikuplja tvrdnje iz primarnih izvora, umjesto da navodi činjenice o stvarnom svijetu, dok je Wikipedija poznata kao besplatna enciklopedija. Svaka Wikidata izjava, barem teoretski, mora biti podržana provjerljivim izvorom koju definiraju politike i smjernice na projektu. Odnosno, sve izjave koje ne sadrže referencu smatraju se neprovjerenim i trebaju biti uklonjene. Iznimka su neosporne tvrdnje, koje predstavljaju opće poznato, primjerice planet Zemlja., kada izjava povezuje entitet s vanjskim izvorom informacija, npr. ID u bazi podataka i kada je izvor izjave sama stavka, primjerice za knjigu i njenog autora. No, Wikidata dopušta i različite verzije iste tvrdnje, pod uvjetom da su potkrijepljene referencama i definirane kvalifikatorima. Ovaj pristup smanjuje mogućnost izbijanja ratova odnosno sukoba pri uređivanju, udaljavajući se od Wikipedijinog mehanizma utemeljenog na konsenzusu za rješavanje sukoba između suprotnih stajališta. Štoviše, Wikidatin liberalni pristup olakšava izražavanje različitih gledišta, rješavajući još jednu problematičnu točku svog sestrinskog projekta (Piscopo i sur., 2017).

Također i sve informacije u člancima na Wikipediji moraju biti provjerljive i potkrijepljene pouzdanim izvorom, inače bi ih zajednica mogla ukloniti. Wikipedia svoju politiku provjerljivosti definira detaljnije od Wikidata. Tri faktora određuju pouzdanost izvora: sam izvor, njegov izdavač i njegov autor. Kao opće pravilo, svi izvori koji su prošli neku vrstu postupka objavljivanja i uredničkog nadzora trebaju se smatrati mjerodavnima. Ažurirani izvori se smatraju boljim odnosno adekvatnijim od starijih. Budući da Wikipedija ne smije sadržavati izvorna istraživanja, savjetuje se korištenje sekundarnih izvora umjesto primarnih, koji mogu biti otvoreni za različita tumačenja. Znanstveni izvori i izvori vijesti općenito se smatraju mjerodavnima, a potrebno je izbjegavati pristrane izvore. To može uključivati stranice dobavljača ili primjerice političkih i vjerskih udruga, ili čak izvore vijesti u nekim slučajevima. Slično kao i na Wikidata, samoobjavljeni izvori trebaju podržavati samo izjave o svojim autorima, a sadržaj koji su generirali korisnici nije prihvatljiv (Piscopo i sur., 2017).

Iako je neosporno da su ova dva Wikimedijina projekta povezana, kao i da postoje određene sličnosti u načinu njihova rada, ipak postoje i neke duboke razlike. One se prvenstveno odnose na izvore, ali i različite strategije s obzirom na vrstu i strukturu sadržanih informacija. Također, potrebno je naglasiti da veza između njih, do sada, nije dovoljno stručno istražena odnosno pokrijepljena dodatnim istraživanjima. S obzirom na mogućnosti koje pružaju, ostavljaju prostora za daljnja istraživanja i unaprjeđivanja, za cjelokupnu dobrobit i napredak društva.

4. WIKIMEDIA COMMONS I WIKIDATA

Još jedan projekt Wikimedia Zaklade usko je povezan sa Wikidatom. Radi se o Wikimedia Commonsu, odnosno besplatnom poslužitelju multimedijalnih sadržaja za sve. U nastavku će se iznijeti glavne karakteristike Wikimedia Commonsa i ono važnije, na koji način je povezan s Wikidatom, odnosno prikaz njihovog uzajamnog utjecaja.

4.1. Općenito o Wikimedia Commonsu

Wikimedia Commons je repozitorij medijskih datoteka koji svima stavlja na raspolaganje javne i slobodno licencirane obrazovne medijske sadržaje kao što su slike, zvuk i video isječci, na njihovom vlastitom jeziku. Osnovan je 7. rujna 2004. godine te djeluje kao zajednički repozitorij za razne projekte Zaklade Wikimedia, kojeg stvaraju i održavaju volonteri. Trenutna brojka učitanih medijskih datoteka je 93,039,423.

Wikimedia Commons koristi istu wiki tehnologiju kao Wikipedia i svatko je može uređivati. Za razliku od medijskih datoteka učitanih u druge projekte, datoteke učitane na WikimediaCommons mogu se ugraditi na stranice svih Wikimedijinih projekata bez potrebe da ih se tamo posebno učita. Kao i drugi projekti Wikimedia Zaklade i WikimediaCommons je besplatan. To znači da svatko može slobodno kopirati, koristiti i mijenjati bilo koje datoteke sve dok slijede uvjete koje je naveo autor. To se odnosi najčešćena navođenje izvora i autora na odgovarajući način i puštanje kopija/poboljšanja pod istom slobodom drugima. Uvjeti licence za svaku pojedinačnu medijsku datoteku mogu se pronaći na njihovoj stranici opisa ("Wikimedia Commons", 2023.)

Glavni cilj Wikimedia Commons je osigurati repozitorij javnih i slobodno licenciranih obrazovnih medijskih datoteka za sve, ali i koji djeluje kao zajednički repozitorij za razne projekte Zaklade Wikimedia. Pod obrazovnim medijskim sadržajem misli se na sadržaj koji educira, odnosno pruža informacije ili znanje. Međutim, većina Wikimedijinih projekata i dalje dopušta lokalne prijenose koji nisu vidljivi drugim projektima ili jezicima, ali ova je opcija prvenstveno namijenjena za korištenje materijala (kao što je sadržaj poštene upotrebe) koji dopuštaju lokalne politike projekta. S druge strane, Wikimedia Commons ne dopušta učitavanje pod neslobodnim licencama, uključujući licence koje ograničavaju komercijalnu upotrebu materijala ili zabranjuju izvedene radove. Drugim riječima, Wikimedia Commons

ugošćuje samo slobodno licencirane medije i briše kršenja autorskih prava. Licence koje su prihvatljive uključuju Creative Commons Attribution i Attribution/ShareAlike licence, drugi besplatni sadržaj i licence za besplatni softver te javnu domenu ("Wikimedia Commons", 2023.)



Slika 13. Logo WikimediaCommons

Izvor: Wikimedia Commons: https://commons.wikimedia.org/wiki/Main_Page (10.05.2023.)

Kako bi dodavali sadržaj odnosno datoteke na Wikimedia Commons, potrebno je prijaviti se odnosno registrirati. Međutim, ako se samo želi urediti stranice, nije se potrebno prijavljivati, iako se to, naravno, preporučuje. Datoteke na repozitoriju Wikimedia Commonsa organizirane su u kategorije i galerije. Pregled kategorija je dostupan na naslovnoj stranici, a raspoređene su u slike, zvukove i videa, koji klikom na neku od njih, otvaraju dodatne kategorije i podkategorije. Pretraživanje je vrlo jednostavno, dovoljno je unijeti pojam u tražilicu u desnom kutu i otvara se bezbroj mogućnosti.

4.2. Povezanost sa Wikidatom

S obzirom na sadržaj koji pruža Wikimedia Commons, postavlja se pitanje, na koji način osigurati da bilo tko na svijetu pronade sliku koju želi bez obzira na jezik koji govori. Odgovor leži u strukturiranim podacima, a upravo je tu i povezanost sa Wikidata projektom.

Strukturirani podaci na Commons-u višejezične su informacije o medijskoj datoteci koju mogu razumijeti ljudi, s dovoljnom dosljednošću da je također mogu obraditi i strojevi. Upravo tu na scenu stupa Wikidata, besplatna baza znanja, koja sadrži podatke o entitetima gotovo svega u svijetu. Drugim riječima datoteke na Wikimedia Commons mogu se opisati entitetima iz Wikidata. Posebni identifikatori pomažu razlikovati različite stvari, iako imaju isto ime. Svi Wikidata entiteti su višejezični, stoga nema veze na kojem jeziku se pretražuje. Na primjer, kiwi (Q13194) je samo voće kivi, a ne vrsta ptice, te za isti identifikator izbacuje na bilo kojem jeziku, primjerice ruski киви (Q13194). Entiteti su također povezani linkom jer Wikidata opisuje odnose između stvari u svijetu. Tako se, primjerice, može pronaći hrana koja je dobila ime po određenoj osobi ili pak različite vrste sarmi iz različitih kuhinja diljem svijeta. Strukturirani podaci na Commons-u korisni su i kreatorima sadržaja i korisnicima sadržaja te poboljšavaju pristup, mogućnost pretraživanja, istraživanje i pružaju nove načine korištenja sadržaja ("Wikimedia Commons", 2023.)

5. ZAKLJUČAK

Na temelju danih informacija u radu, mogu se sumirati glavne zaključne točke o temi. Wikidata, je suradnička baza znanja koja je javno dostupna i sadrži podatke čitljive i ljudima i stroju. Glavni cilj iza nastanka projekta Wikidata je bio pružiti strukturirane podatke za Wikimedijine projekte kako bi se prevladale nedosljednosti podataka u jezičnim verzijama Wikipedije. Međutim, kao što se moglo vidjeti kroz rad, Wikidata danas predstavlja puno više. Kao višejezična baza znanja, Wikidata može pružiti podatke u bilo kojem kontekstu i jeziku. Mnoge su studije istraživale projekt Wikidata iz različitih perspektiva, pa se može zaključiti da je Wikidata jako korisna kao početna baza za istraživanje za projekte u medicini, lingvistici ili geografiji. Također, Wikidata postaje sve popularnija u knjižnicama kao otvorena i suradnička globalna platforma za dijeljenje i razmjenu knjižničnih metapodataka. Zbog toga je, u posljednje vrijeme, došlo do povećanja interesa istraživačke zajednice za Wikidata, što ukazuje na njezinu rastuću popularnost. Wikidata je tekući projekt koji se aktivno razvija. Naravno, u budućnosti, kako se Wikidata konstantno razvija i raste, će biti dostupno više tipova podataka kao i proširenja.

U radu je prikazan i odnos Wikidate sa svojim sestrinskim projektima, Wikipedijom, kao besplatnom enciklopedijom i Wikimedia Commonsom, repozitorijem medijskog sadržaja. Danas je svima naročito poznata Wikipedia te predstavlja jedno od najvažnijih web-mjesta današnjice, međutim, Wikidata je, premda dosta mlađa, već postala važna platforma za integraciju informacija iz mnogih izvora. Iako je neosporno da su ovi Wikimedijini projekti povezani, kao i da postoje određene sličnosti u načinu njihova rada, ipak postoje i neke duboke razlike. One se prvenstveno odnose na izvore, ali i različite strategije s obzirom na vrstu i strukturu sadržanih informacija. Također, potrebno je naglasiti da veza između njih, do sada, nije dovoljno stručno istražena odnosno pokrijepljena dodatnim istraživanjima. S obzirom na mogućnosti koje pružaju, ostavljaju prostora za daljnja istraživanja i unaprjeđivanja, za cjelokupnu dobrobit i napredak društva. Zaključno se može reći da Wikidata, kao besplatna baza znanja koju svatko može uređivati, može postati najznačajniji izvor za svaku vrstu pretraživanja.

6. LITERATURA

Dowell, M.L.; Bridges, L. M. (2019). *A Perspective on Wikipedia: Your Students Are Here, Why Aren't You?*, The Journal of Academic Librarianship, Volume 45, Issue 2, str. 81-83. Preuzeto 12.05.2023. s

<https://www.sciencedirect.com/science/article/abs/pii/S0099133319300187?via%3Dihub>

Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., Vrandečić, D. (2011). *Introducing wikidata to the linked data web*. In: ISWC, str. 50. – 65.

Farda Sarbas, M. & Müller-Birn, C. (2019). *Wikidata from a Research Perspective - A Systematic Mapping Study of Wikidata*. Preuzeto 5.05.2023 s <https://arxiv.org/pdf/1908.11153.pdf>

Feddoul, L., Loffler, F., Schindler, S. (2022) *Analysis of Consistency between Wikidata and Wikipedia Categories*. Preuzeto 15.05.2023. s <https://ceur-ws.org/Vol-3262/paper4.pdf>

Hitzler, P., Krotzsch, M., Rudolph, S. (2011). *Foundations of Semantic Web Technologies*. Philadelphia, PA: Chapman & Hall / CRC

Pellissier Tanon, T., Suchanek, F. (2019). *Querying the Edit History of Wikidata*. In: et al. The Semantic Web: ESWC 2019 Satellite Events. Lecture Notes in Computer Science, vol 11762. Springer, Cham. Preuzeto s 10.05.2023. s https://doi.org/10.1007/978-3-030-32327-1_32

Piscopo, A., Vougiouklis, P., Kaffee L.A., Phethean, C., Hare, J., Simperl, E. (2017). *What do Wikidata and Wikipedia Have in Common? An Analysis of their Use of External References*. Association for Computing Machinery, New York, NY, USA, Article 1, 1–10. Preuzeto 7.05.2023. s https://eprints.soton.ac.uk/412922/1/opensym_wd_vs_wp_2_.pdf

Tharani, K. (2021.) *Much more than a mere technology: A systematic review of Wikidata in libraries*, The Journal of Academic Librarianship, JAI, Volume 47, Issue 2. Preuzeto 6.05.2023. s <https://www.sciencedirect.com/science/article/pii/S0099133321000173>

Vrandečić, D., Krötzsch, M. (2014). *Wikidata: a free collaborative knowledge base*. Commun. ACM 57(10), str. 78–85

Vrandečić, D. (2013). *The rise of Wikidata*. IEEE Intelligent Systems 28, 4 (2013), str. 90–95.

Wikidata: https://www.wikidata.org/wiki/Wikidata:Main_Page (18.4.2023.)

Wikipedia: The free Encyclopedia: <https://en.wikipedia.org/wiki/Wikipedia:About> (11.05.2023.)

Wikimedia Commons: <https://commons.wikimedia.org/wiki/Commons:Welcome> (10.05.2023.)

Resource Description Framework – RDF (2004). *Concepts and Abstract Syntax*, World Wide Web Consortium. Dostupno na: <https://www.w3.org/TR/rdf-concepts/> (5.05.2023.)

POPIS SLIKA

Slika 1. Zdenko „Denny“ Vrandečić – osnivač Wikidate.....	3
Slika 2. Izgled naslovne stranice s prikazom trenutne brojke sadržaja.....	7
Slika 3. Analiza pregleda stranice za 4. mjesec.....	8
Slika 4. Mapa događaja proslave 10. rođendana Wikidate.....	9
Slika 5. Primjer entiteta Zagreb na Wikidata repozitoriju.....	13
Slika 6. Primjer povezanosti podataka na Wikidati.....	14
Slika 7. Podatci koji se moraju unijeti na Wikidata.....	15
Slika 8. WikiProject COVID-19 logo.....	18
Slika 9. Pregled znanstvenih publikacija o COVIDU-19 tijekom prva tri mjeseca 2020. godine.....	19
Slika 10. Logo WikiProject Music.....	20
Slika 11. Logo Wikipedije.....	22
Slika 12. Naslovna stranica Wikipedije.....	24
Slika 13. Logo WikimediaCommons.....	28

POPIS GRAFOVA

Graf 1. Rast Wikidate prema broju uređivanja različitih grupa.....	5
Graf 2. Rast Wikidate od osnutka do 2014. godine.....	6

SAŽETAK

Ovaj rad bavi se prikazom projekta Wikidata, njenim razvojem, karakteristikama te glavnim prednostima i manama. Wikidata kao dio Wikimedija zaklade, jedna je od projekata koji pružaju besplatan pristup informacijama. Ova baza je razumljiva strojevima i ljudima jer se sastoji od strukturiranih podataka koji su međusobno povezani. Projekt je relativno novi dio Wikimedija, a u cilju ima povezati entitete u strukturiranu bazu podataka kako bi što više informacija bilo integrirano te kako bi pristup informacijama bio znatno lakši. Upravo prilikom unosa podataka, informacije prolaze kroz mehanizme provjerljivosti kako bi se ne bi dogodili sukobi autora, koji se događaju. Obilježjima poput višejezičnosti, besplatnog pristupa i strukturiranih podataka, Wikidata unaprjeđuje sestrinske projekte Wikimedije, poput Wikipedije i WikimediaCommonsa, te je zato u radu objašnjen uzajamni utjecaj Wikidatae i Wikipedije, a naglašene su i glavne sličnosti i razlike u projektima. Osim toga, objašnjeni su strukturirani podatci u WikimediaCommonsu, te potreba njihovog nastajanja i povezanosti sa Wikidatom.

Ključne riječi: *Wikidata, strukturirani podatci, informacije, Wikipedija, WikimediaCommons*

SUMMARY

This paper deals with the presentation of Wikidata, its development, characteristics and main advantages and disadvantages. Wikidata, as part of the Wikimedia Foundation, is one of the projects that provide free access to information. This database is understandable by machines and people because it consists of structured data that are interconnected. The project is a relatively new part of Wikimedia, and its goal is to connect entities into a structured database so that as much information as possible is integrated and access to information is much easier. Precisely when entering data, the information goes through verifiability mechanisms so that conflicts of authors, which do occur, do not occur. With features such as multilingualism, free access and structured data, Wikidata improves sister projects of Wikimedia, such as Wikipedia and Wikimedia Commons, and therefore the mutual influence of Wikidata and Wikipedia is explained in the paper, and the main differences in the projects are emphasized. In addition, structured data in Wikimedia Commons and the need for their creation and connection with Wikidata were explained.

Keywords: *Wikidata, structured data, information, Wikipedia, Wikimedia Commons*