

## Wikipedija *nasuprot* Hrvatskoj enciklopediji, *kvalitativan odnos slobodno i tradicionalno uređenoga enciklopedijskoga sadržaja na hrvatskom jeziku*

Ozren Kubelka

Veleučilište Vern

Petra Šoštarić

Filozofski fakultet Sveučilišta u Zagrebu

**SAŽETAK:** Slobodno uređene enciklopedije, od kojih je *Wikipedija* vodeća prema broju članaka i korisnika, danas predstavljaju nezanemariv izvor informacija. Kako je koncepcija njihova uređivanja zasnovana na anonimnosti i slobodi nosa sadržaja, od same njihove pojave traju rasprave o pouzdanosti i kvaliteti na taj način uređenih sadržaja. Ovim se radom nastojalo spoznati trenutačno kvalitativno stanje enciklopedijskih izvora nastalih slobodno uređivačkom, odnosno tradicionalnom koncepcijom. Za izvore uspoređivanih sadržaja odabrana je *Hrvatska enciklopedija*, kao tradicionalno uređen izvor, i hrvatsko izdanje *Wikipedije*, kao slobodno uređen enciklopedijski sadržaj. U radu su definirani elementi putem kojih se utvrdio njihov kvalitativni odnos, a provjereni su putem recenzentske analize u kojoj je sudjelovalo 24 recenzenata, stručnjaka u pojedinim područjima. Rezultatima istraživanja prikazan je trenutačni odnos izvora prema parametrima informativnosti, točnosti iznesenih informacija, dostatnosti, usmjerenosti i objektivnosti. Rezultati ovoga rada dio su širega istraživanja kojim je uspostavljen model za kvalitativnu usporedbu izvora informacija.

**Cljučne riječi:** *kvaliteta enciklopedijskoga sadržaja; slobodno uređen enciklopedijski sadržaj; tradicionalno uređen enciklopedijski sadržaj; Wikipedija; Hrvatska enciklopedija*

### Uvod

Razvoj suvremene tehnologije učinio je informacije dostupnijima nego ikada prije u ljudskoj povijesti. Rastuća potreba za točnom i relevantnom informacijom o bilo kojoj temi u bilo kojem trenutku, potpomognuta razvojem tehnologije, dovela je

do značajnih promjena u koncepciji enciklopedije, pa su se slobodno uređeni enciklopedijski sadržaji počeli pojavljivati već početkom 1990-ih, a od 2001. imaju više korisnika od onih tradicionalnih. Brz rast sadržaja omogućavaju im uređivačka politika i medij, a jednostavnost pristupa i uređivanja olakšava ažuriranje. Druga je strana medalje pitanje kvalitete sadržaja koje uređuju anonimni autori.

Cilj je ovoga rada usporedba kvalitete enciklopedijskoga sadržaja nastaloga slobodnim uređivanjem i onoga tradicionalno uređenoga. Kako je većina dosadašnjih istraživanja razmatrala gotovo isključivo izvore na engleskom jeziku (Giles 2005; Brandes i Lerner 2008; Fallis 2008; Okoli 2009; West i dr. 2009), namjera ovoga istraživanja bila je utvrditi kvalitativno stanje među enciklopedijskim izvorima na hrvatskom jeziku. U tu svrhu, proširujući saznanja dosadašnjih istraživanja, napravljen je i primijenjen komparativni model utvrđivanja kvalitete enciklopedijskoga sadržaja (Kubelka 2013).

Za istraživanje su odabrana dva izvora: *Hrvatska enciklopedija* Leksikografskoga zavoda Miroslav Krleža, kao tradicionalno uređen leksikografski izvor, te hrvatsko izdanje *Wikipedije* (hr.wikipedia.org), kao slobodno uređen. Obje su enciklopedije za potrebe ovoga rada obrađene u svojim elektroničkim inačicama, a u trenutku preuzimanja podataka, 15. svibnja 2012., opseg je iznosio 68 718 članaka *Hrvatske enciklopedije* i 104 356 članaka *Wikipedije*. Računalno odabran uzorak podvrgnut je automatskoj računalnoj i recenzentskoj analizi. Dobivenim rezultatima utvrđuje se odnos kvalitete dvaju izvora, a time i njihova stručna i znanstvena upotrebljivost.

U radu je također iznesen i kratak pregled djelovanja *Wikipedije*, čime je pružen uvid u razloge njezina uspjeha, ali i u elemente koje kritičari slobodnoga uređivanja smatraju spornima u pogledu kvalitete. U tome je ključni prijedor oko pouzdanosti informacija sadržanih u *Wikipediji*.

## 1. Povijest *Wikipedije*

Iako je prvo izdanje *Wikipedije* s radom započelo 2001. godine, preteče slobodnoga uređivačkoga koncepta nešto su starije. Prvi wiki-poslužitelj pokrenuo je još 1994. Ward Cunningham, s ciljem programiranja najjednostavnije moguće funkcionalne baze podataka koju svatko može uređivati. Prvu internetsku enciklopediju pokrenuo je Rick Gates u listopadu 1993. pod nazivom *Interpedia*. Gatesov projekt, nastao prije razvoja koncepta Web 2.0, bio je ispred svojega vremena te nije zaživio zbog nedostatka prikladnoga softvera, no nepuno desetljeće nakon tih vizionarskih početaka dvije ideje spojiti će se u jedan od najutjecajnijih mrežnih servisa današnjice.

Kad je Jimmy Donal Wales pokrenuo *Wikipediju*, već je imao iza sebe propali projekt pod nazivom *Nupedia*, započet 2000. uz pomoć filozofa Larryja Sangera, koji je koristio svoja poznanstva u akademskom svijetu da privuče stručne suradnike.

Sanger je kao glavni urednik inzistirao na visokom stupnju obrazovanja te su mu suradnici morali faksirati potvrde o stečenom doktoratu da bi mogli sudjelovati u izradbi enciklopedije. Proces izradbe članaka nije bio jednostavan: svaki je autor morao poslati nacrt članka uredničkom odboru, koji bi ga pregledao i potom prihvatio ili odbacio. Glavni recenzent nadgledao je stvaranje članka; nakon njegova odobrenja, cijela zajednica *nupedista* dobila bi članak na uvid, a potom su slijedile dvije lekture. Zbog takva rigorozna sustava pravila, u prvoj godini postojanja *Nupedije* objavljeno je tek nekoliko desetaka članaka, što nije zadovoljilo ni pokretače ni financijere projekta, tvrtku Bomis. Do preokreta je došlo kad su Sanger i Wales otkrili *WikiWikiWeb*, internetsku stranicu koju je mogao uređivati bilo tko, i to bez upotrebe posebnog softvera. Ideju programera Warda Cunninghama Sanger i Wales odlučili su primijeniti na svojoj enciklopediji, usprkos sumnji koju su opravdano pobudile anonimnost autora i sloboda uređivanja. *Nupedia* je ugašena 2003., dvije godine nakon pokretanja *Wikipedije* kao Bomisova eksperimentalnoga projekta.

## 2. Koncept rada *Wikipedije* i istraživanja kvalitete

Uređivački koncept, zasnovan na anonimnosti suradnika i velikoj slobodi upliva u sadržaj, rezultirao je iznimno velikim brojem sakupljenih članaka, tako da danas englesko izdanje *Wikipedije* broji više od četiri milijuna članaka, a hrvatsko gotovo 140 000.

*Wikipedija* je organizirana kao publikacijski model s dvije osnovne mrežne razine: mrežom članaka i suradničkom mrežom (Korfiatis i dr. 2006). Unutar mreže članaka svaki je pojedini članak kategoriziran te referencama povezan s drugim člancima i vanjskim izvorima.

Unutar suradničke mreže postoje tri osnovne kategorije suradnika:

1. anonimni suradnici: uređuju sadržaj bez prijave.
2. prijavljeni suradnici: korisnici prijavljeni pod odabranim korisničkim imenom, bez identifikacijskih podataka.
3. suradnici s provjerom: korisnici sa statusom administratora, birokrata, stjuarta, bota, patrolera, nadzornika, unositelja i sl., kojima je povjeren pristup tehničkim dodatcima (alatima) za uređivanje sadržaja. Koriste svoje ovlasti kad procijene da je potrebno posebno zaštititi ili urediti neki sadržaj.

Aktivan suradnik *Wikipedije* svaki je prijavljeni korisnik s barem pet mjesečnih izmjena, odnosno intervencija u sadržaj. Svaki korisnik može i postavljati i uređivati sadržaj, no većinu ne zanima uređivanje: na engleskoj *Wikipediji* članke uređuje tek 0,7% od ukupnog broja prijavljenih suradnika, a na španjolskoj je 8,1% suradnika zaslužno za 90% izmjena.

Upravo je prednost koja je ostvarena konceptom slobodnoga uređivanja ono što najviše pobuđuje oprez javnosti. Slobodno uređivanje omogućava brz rast enciklopedije, ali istovremeno otvara pitanja kvalitete i sigurnosti sadržaja. Budući da je većina suradnika anonimna, nema tradicionalnoga odnosa između autoriteta i pouzdanosti informacije. Da bi se zadržala određena razina kvalitete, suradnici su pozvani da se pridržavaju sljedećih standarda: pismenosti, točnosti, neutralnosti, dostatnosti, primjene odgovarajućega stila i fokusiranosti, provjerljivosti podataka, stabilnosti te odgovarajuće ilustriranosti. Potiče se ispravljanje grešaka i poboljšavanje članaka na svim razinama, a članci, neutralnost kojih je upitna, dobivaju posebne oznake. Čestim ispravkama i uređivanjem raste status suradnika unutar zajednice te mu se omogućuje da uđe u 3. skupinu suradnika. Usprkos danas razvijenim pojmovima poput *copyrighta* i intelektualnoga vlasništva, nije u stilu *wikipedista* inzistirati na autorstvu. *Wikipedija* je, za razliku od primjerice blogova, proizvod zajednice, a članci mogu nastati suradnjom nekolicine autora.

Upravo takav brz rast broja članaka i utjecaj koji je *Wikipedija* postigla ponukali su istraživanja kvalitete. Prvim zapaženijim istraživanjem kvalitete (Giles 2005) uspoređena su 42 članka, na uzorku kojih je uspoređen broj pogrešaka u *Britannici* i engleskom izdanju *Wikipedije*. Prosječan omjer grešaka bio je 3:4 u korist *Britannice*. Iako je potvrđena viša kvaliteta tradicionalno uređenoga izvora, usporedivost rezultata pobudila je mnoge rasprave i skrenula pozornost na slobodno uređene izvore kao potencijalno relevantan izvor informacija. Svakako treba spomenuti da je metodologija rada bila u mnogim dijelovima nedoradena, što je često bilo metom kritika osporavatelja toga rada.

Otvoren pristup i lakoća uređivanja čine *Wikipediju* osjetljivom na potencijalne prijetnje. Tekst dostupan svima izložen je kudikamo većim opasnostima nego što su to bile tradicionalne enciklopedije. Ako bismo potencijalne prijetnje sakupili u osnovne grupe, došli bismo do četiri glavna uzroka koja mogu prouzročiti kvalitativne probleme u *Wikipediji*: velik opseg sadržaja, velik broj različitih tema, nepoznato autorstvo i vandalizam (Lim i dr. 2006). Usprkos navedenim mogućim konceptualnim slabostima, rezultati dobiveni istraživanjima često su bili pozitivni spram *Wikipedijine* opće točnosti (Rector 2007; West i dr. 2007).

### 3. Kratak pregled hrvatske enciklopedike

Prva hrvatska opća enciklopedija nastala je u Osijeku pod vodstvom Ivana Zocha i Josipa Mencina – *Hrvatska enciklopedija: priručni rječnik sveobćega znanja* bila je zamišljena u šest svezaka, no objavljena su samo dva (prvi 1887., drugi 1890). Od 1941. do 1945. godine izlazila je u Zagrebu *Hrvatska enciklopedija*, koju ju izdavao Hrvatski izdavački bibliografski zavod, a uređivao Mate Ujević. Ni ona nije objavljena u cijelosti: izišlo je pet od predviđenih 12 svezaka. Tek je treća hrvatska enciklo-

pedija cjelovita: *Enciklopedija* Leksikografskoga zavoda. Izlazila je u sedam svezaka od 1955. do 1964. godine pod vodstvom Miroslava Krleže, Marka Kostrenčića i Josipa Protege. Slijedila je *Opća enciklopedija* u osam svezaka, koji su izlazili od 1977. do 1982., a glavni joj je urednik bio Josip Šentija. *Hrvatska enciklopedija* Leksikografskoga zavoda Miroslav Krleža, u 11 svezaka, izlazila je od 1999. do 2009. Uređivali su je Dalibor Brozović (sv. I–III), August Kovačec (sv. IV–VII) i Slaven Ravlić (sv. VIII–XI). Kao odgovor na potrebe globaliziranoga i informatiziranoga društva današnjice nastaje njezina mrežna verzija. Dostupna je putem interneta na [hebeta.lzmk.hr](http://hebeta.lzmk.hr) i [www.enciklopedija.hr](http://www.enciklopedija.hr).

#### 4. Kvantitativni i sadržajni odnos između izvora te odabir recenzentskoga uzorka

Kako bi se dobio statistički relevantan odabir članaka za recenzentsku analizu, provedena je kvantitativna i kvalitativna analiza obaju izvora.

Opseg sadržaja u obrađenim izvorima prikazan je tablicom 1. Iako se broj članaka u njima razlikuje, ta je razlika znatno manja nego kod engleskih izvora. Primjerice, razlika između *Britannice* i *Wikipedije* iznosi 1:35. Upravo takav manji omjer pružio je mogućnost preglednije usporedbe izvora.

Tablica 1. Usporedba opsega sadržaja

	Broj znakova	Broj riječi	Broj članaka	Broj kategorija
<i>Hrvatska enciklopedija</i> (HE)	44 252 707	6 682 619	68 718	131
hrvatska <i>Wikipedija</i> (WP)	161 454 138	23 646 020	104 356	> 1000
Odnos HE : WP	<b>1 : 3,62</b>	<b>1 : 3,53</b>	<b>1 : 1,52</b>	<b>≈ 1 : 8</b>

Na temelju kvantitativnih odnosa između izvora, utvrđuje se da je broj članaka za 50% veći u *Wikipediji*, no istovremeno broj riječi i znakova u njoj za 3,5 puta nadmašuje sadržaj *Hrvatske enciklopedije*. Upravo zbog toga nerazmjera u pripremi recenzentskoga uzorka izračunata je prosječna duljina članka za pojedini izvor, temeljem čega su članci kategorizirani kao kratki (S), srednji (M) i dugi (L). Izvedena je uniformna distribucija na svakom izvoru, a dobiveni rezultati primijenjeni su prilikom određivanja distribucije članaka po pojedinoj kategoriji.

$$P(X = x) = \frac{1}{n} \forall x \in \mathbb{N}; x=1,2,-,n$$

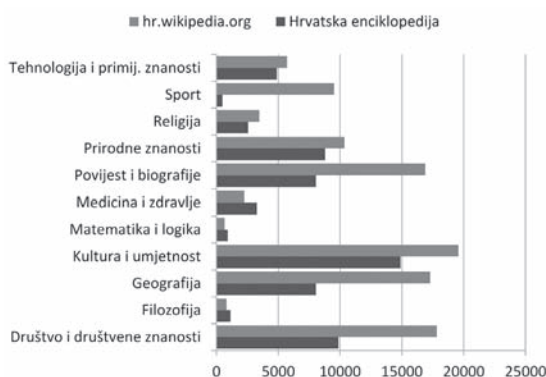
Prethodno su isključeni članci s manje od dvadeset riječi jer je na uzorku od 200 članaka te duljine zaključeno da se uglavnom radi o člancima koji nisu dovršeni (više od 83%) ili služe kao reference na druge članke. U oba slučaja radilo se o članci-

ma koji nisu pogodni za sadržajnu analizu zbog neznatna broja iznesenih informacija.

Radi objektivne primjene kriterija u recenziji su uspoređivani isključivo članci istih kategorija po duljini.

Drugi korak u pripremi izveden je usklađivanjem kategorijske strukture oba izvora. Već je spomenut nerazmjer u broju kategorija između enciklopedija, te je temeljem *Hrvatske enciklopedije* izveden opći kategorijski okvir. Sve kategorije (strukture) posljednjega izdanja *Hrvatske enciklopedije*, njih 131, za potrebe analize razvrstane su unutar 11 općih kategorija (područja): kultura i umjetnost, povijest i biografije, medicina i zdravlje, tehnologija i primijenjene znanosti, geografija, religija, prirodne znanosti, matematika i logika, filozofija, sport te društvo i društvene znanosti. Nakon toga je ista struktura kategorija i potkategorija primijenjena na članke hrvatske *Wikipedije*, i to upotrebom metode nadziranoga računalnoga učenja, i to metodom linearnoga SVM<sup>1</sup>-a, a za određivanje težine značajki statistika TF-IDF<sup>2</sup>.

Algoritam za primjenu metode »naučio« je prepoznavati kategorije na uzorku svih članaka *Hrvatske enciklopedije* i primijenio kategorizaciju na *Wikipediju*. Točnost predviđanja algoritma iznosila je 84%, a provjerena je tzv. matricom zabune, na uzorku od 250 članaka. Točnost primjene metode donekle je kvario rezultat kategorije *Sport*, sadržaj koje je relativno nezastupljen u *Hrvatskoj enciklopediji*, te je stoga algoritmu pružen relativno malen set podataka za »učenje«. Isključenjem ove kategorije iz daljnjega razmatranja točnost metode iznosila je približno 90%, što je bilo zadovoljavajuće za njezinu primjenu.



Slika 1. Broj članaka u pojedinoj kategoriji prema izvorima

<sup>1</sup> Metoda SVM (engl. *Support Vector Machine*) zasniva se na kreiranju hiperravnina u višedimenzionalnom prostoru koje se koriste za razdvajanje dviju klasa.

<sup>2</sup> TF (engl. *Term Frequency*) i IDF (engl. *Inverse Document Frequency*) statističke su metode određivanja težinskih karakteristika nekog termina u dokumentima.

Grafikonom na slici 1. prikazana je zastupljenost članaka prema općim kategorijama u odabranim izvorima.

Kad je na objema enciklopedijama primijenjena ista kategorijska struktura i kad su članci razvrstani prema dužinskim kategorijama, izvedeno je sparivanje članaka za recenziju. Tablica 2. prikazuje slučajnim odabirom sparenih 500 članaka, odnosno 250 parova u općim kategorijama.

Tablica 2. Broj članaka prema općim kategorijama u recenzentskom uzorku

Enciklopedijsko područje	Broj članaka / parova
Društvo i društvene znanosti	60/(30)
Filozofija	40/(20)
Geografija	60/(30)
Kultura i umjetnost	60/(30)
Matematika i logika	60/(30)
Medicina i zdravlje	40/(20)
Povijest i biografije	40/(20)
Prirodne znanosti	60/(30)
Religija	20/(10)
Tehnologija i primijenjene znanosti	60/(30)
<b>UKUPNO</b>	<b>500/(250)</b>

Nakon provedene recenzentske analize isključeno je 5% rubnih rezultata prema kriterijima većih i manjih činjeničnih pogrešaka te ti članci nisu razmatrani niti u drugom dijelu recenzije koji se odnosi na dostatnost, usmjerenost i objektivnost članaka.

## 5. Elementi ocjene kvalitete sadržaja

Kod odabira elemenata za kvalitativnu usporedbu, sagledane su dvije ključne različitosti enciklopedijskih izvora. Jedna se odnosi na spomenut uređivački koncept, a druga na medij objave. Kod potonjega valja naglasiti kako slobodno uređene enciklopedije koriste internetski medij za objavu, dok je *Hrvatska enciklopedija* objavljena najprije kao tiskano izdanje. Iz toga proizlazi nemogućnost potpune usporedbe prema svim sadržanim elementima. Ako bismo uzeli vanjske poveznice kao element usporedbe, tiskano bismo izdanje postavili u nepovoljan položaj jer je tehnički nemoguće ostvariti onakvu referentnost kakvu omogućavaju *hiperlinkovi* internetskoga izdanja. Još je značajnija razlika u duljini članaka i sadržajnom opsegu. Internet omo-

gućuje neusporedivo nižu cijenu objave nego tisak, iz čega proizlazi znatna razlika i u opsegu sadržaja.

Imajući u vidu razlike, izdvojeni su i određeni usporedivi elementi, kojima će se utvrditi kvalitativan odnos izvora:

- broj iznesenih činjenica u članku; ovim kriterijem zadatak recenzenta bio je pobrojiti iznesene činjenice. Činjenica je u uputama definirana kao svaki navod koji je relevantan za izloženu temu, a koji se s jedne strane može nepobitno i neupitno utvrditi, a s druge empirijski i eksperimentalno dokazati.
- broj netočnih činjenica s podjelom na veće i manje činjenične pogreške; većima su smatrane one pogreške koje značajno utječu na točnost pojedinog članka, a manjima one navodom kojih se ne umanjuje značajno kvaliteta članka. Ista tvrdnja može u jednom članku biti manja, a u drugom veća činjenična pogreška s obzirom na svoju važnost za samu temu.

Ti su elementi pobrojani u člancima uzorka i njihov je odnos uspoređen. Drugi dio recenzentske analize odnosio se na pojedinačno ocjenjivanje članaka prema kriterijima:

- dostatnosti, koji se odnosi na količinu informacija iznesenih u članku, a koje bi trebale biti dovoljne za razumijevanje teme. Primjena kriterija dostatnosti razlikuje se u primjeni na različite dužinske kategorije članka.
- usmjerenosti, kojim se opisuje povezanost izloženih činjenica s temom članka. Digresije u druga područja umanjuju usmjerenost članka.
- objektivnosti, koji je u smjernicama za izradu članaka na *Wikipediji* naveden kao jedno od osnovnih načela. U slučaju članaka koji obrađuju kontroverzne teme potrebno je iznijeti sva gledišta, dok je iznošenje jednostranih i pristranih sudova neprihvatljivo.

Za svaki od ta tri kriterija članku je dodijeljena ocjena od jedan do tri, gdje je jedinica predstavljala u potpunosti udovoljavanje kriteriju, dvojka djelomično udovoljavanje, a trojka potpuno nezadovoljavanje kriterija.

## 6. Recenzentski tim

Za provedbu ovoga dijela istraživanja okupljen je tim od 24 recenzenta, stručnih za pojedino područje recenzije. Kriteriji prilikom njihova odabira bili su formalno obrazovanje u području recenzije i aktivan istraživački ili praktični rad vezan uz određeni sadržajni segment. U svrhu dobivanja što pouzdanijih rezultata te kako bi se smanjila odstupanja u interpretaciji kriterija, za svako područje recenzije birano je više recenzenata. Da bi se smanjila pristranost zbog sklonosti određenom tipu enciklopedijskih



izvora, odabrani su recenzenti različite dobi. Prosjek godina iznosio je 39,5, s tim da je desetero recenzenata bilo mlađe od 35 godina, a starijih od 49 godina bilo je troje. Time je dobiven tim koji je upoznat i s tradicionalnim tipom enciklopedija i sa suvremenim, a koji se obrazovao prije nagloga razvoja slobodno uređenih enciklopedijskih sadržaja.

## 7. Opća informativnost sadržaja

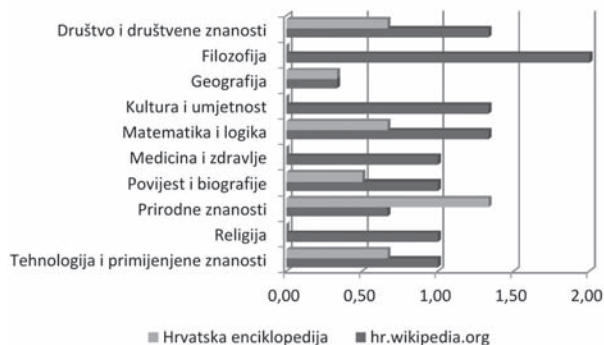
U uzorcima obiju enciklopedija pobrojane su iznesene činjenice. Njih 3015 pripadalo je *Hrvatskoj enciklopediji*, a 3315 *Wikipediji*. Iako je njihov broj u slobodno uređenoj enciklopediji nešto veći, gledajući na informativnost kao na odnos iznesenih činjenica prema broju upotrijebljenih riječi, informativnost *Wikipedije* je manja: kod *Wikipedije* je prosječno trebalo 18 riječi kako bi se iznijela jedna činjenica, dok su u tradicionalnoj enciklopediji one bile koncentriranije, prosječno sadržane u 15 riječi. Taj je rezultat u skladu s kasnijom ocjenom usmjerenosti kojom su recenzenti ocijenili članke.

Presjek podudarnih činjenica iznosio je 19,77%, što znači da je tek petina iznesenoga sadržaja podudarna.

## 8. Čestoća većih i manjih pogrešaka u izvorima

Ako bismo na isti način interpretirali podatke kako je to napravio Giles u svojoj usporedbi *Britannice* i engleske *Wikipedije*, rezultat pronađenih većih pogrešaka iznosio bi 1:2,25, odnosno na svaku pronađenu pogrešku u *Hrvatskoj enciklopediji* pronađene su 2,25 pogreške u *Wikipediji*. Odnos manjih pogrešaka iznosi 1:1,77.

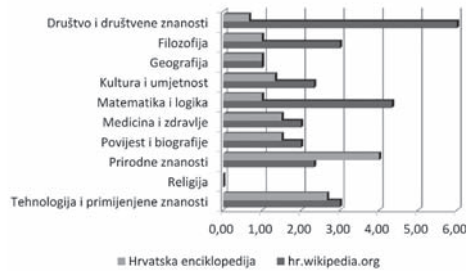
Takva interpretacija nije dovoljno precizna jer nije u korelaciji s ukupnim opsegom objavljenoga sadržaja te može navesti na pogrešan zaključak. Precizniju interpretaciju pruža pregled prosječne pojave pogrešaka u određenom broju članaka. Grafikonu na slikama 2. i 3. prikazuju te podatke.



Slika 2. Veće činjenične pogreške u općim temama (prosjek od 10 članaka)

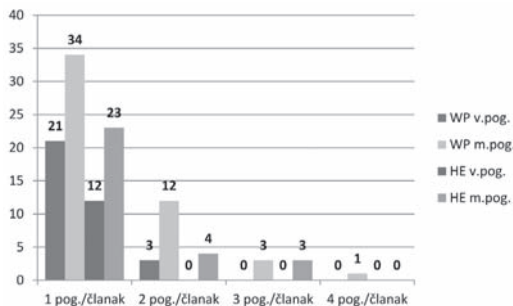
Ako se detaljnije sagleda pojava većih činjeničnih pogrešaka, vidjet će se da je u svim kategorijama pronađeno više većih pogrešaka u *Wikipediji*. Najviše pogreška pronađeno je u kategoriji *Filozofija*, u kojoj se javljaju prosječno dvije pogreške na deset članaka. Kategorija *Geografija* bilježi jednak broj većih pogrešaka, u prosjeku 0,25 na deset članaka. Jedina kategorija u kojoj je pronađeno više pogrešaka u *Hrvatskoj enciklopediji* jesu *Prirodne znanosti*, s omjerom 1,25 nasuprot 0,75.

Odnos manjih činjeničnih pogrešaka nešto je manji. Izuzetak je kategorija *Društvo i društvene znanosti*, u kojoj je taj odnos relativno velik, 1:3. Uz *Društvo*, u još šest kategorija pronađena je veća točnost podataka u tradicionalnom izdanju. Jedino u kategoriji *Religija* nisu pronađene manje pogreške, a *Prirodne znanosti* ponovno su jedina kategorija u kojoj je točnost podataka veća u slobodno uređenoj enciklopediji.



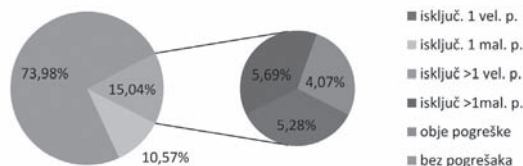
Slika 3. Manje činjenične pogreške u općim temama (prosječno od 10 članaka)

Dodatna interpretacija dobivenih rezultata prikazana je grafom na Slici 4., na kojem je uspoređen apsolutan broj pogrešaka prema vrsti i prema pojavi u člancima. Tako možemo vidjeti da je u uzorku od 250 članaka pronađena 21 veća pogreška u *Wikipediji*, nasuprot 12 u *Hrvatskoj enciklopediji*. Odnos manjih pogrešaka je 34:23. Pojava dvije i više većih pogrešaka nije zabilježena u tradicionalno uređenoj enciklopediji, dok su u uzorku članaka iz slobodno uređenog izvora pronađena tri članka s dvije veće pogreške. Ostale vrijednosti ukazuju na pojavu više manjih pogrešaka, gdje je kvalitativno stanje uglavnom bolje kod *Hrvatske enciklopedije*.



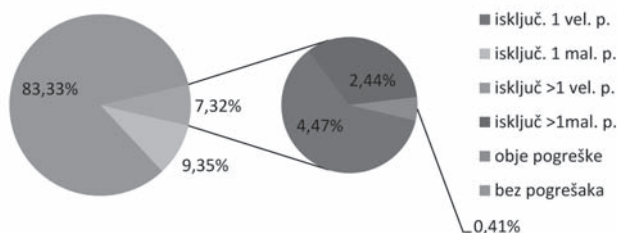
Slika 4. Broj pogrešaka u člancima prema izvorima

Pouzdanost pojedinog izvora može se sagledati udjelom članka s pogreškama i onima bez njih. Slika 5. prikazuje udjel pogrešaka u ukupnom sadržaju hrvatskoga izdanja *Wikipedije*. U približno 74% sadržaja nije pronađena niti jedna vrsta pogreške, a u 11% tek jedna manja. Kako pojava manjih pogrešaka ne utječe na informativnost članka, zbroj ovih dvaju segmenata možemo uzeti kao zadovoljavajuće informativan, a zbroj iznosi oko 85%. Pojava većih pogrešaka pronađena je u nešto više od 5% članaka, dok je pronalazak obiju vrsta pogrešaka zabilježen u 4% članaka. Pojavu više od jedne manje pogreške također smo smatrali dovoljnom za smetnju po informativnost, a takvih je pronađeno u 5,69% obrađenih članaka.



Slika 5. Hrvatska *Wikipedija*; udjel pogrešaka u uzorku

Kod *Hrvatske enciklopedije* rezultati su sljedeći: u nešto više od 83% članaka nije uočena niti jedna pogreška, dok je u 9,35% pronađena manja. Njihovim zbrojem dobivamo segment od više od 92% sadržaja koje možemo smatrati u potpunosti informativnima. Pojava veće činjenične pogreške pronađena je u približno 4,5% članaka, a ostale u znatno manjem broju, usporedimo li *Hrvatsku enciklopediju* s *Wikipedijom*. Tako je pronađen samo jedan članak koji je sadržavao obje pogreške, što u postotku iznosi 0,4. Više od jedne manje pogreške sadržavalo je 2,44% članaka.

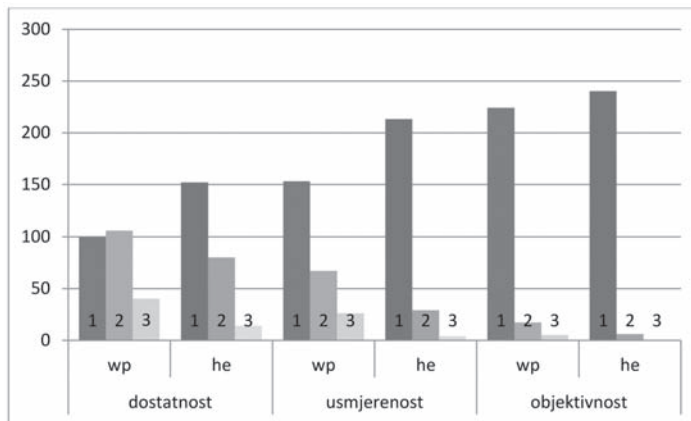


Slika 6. *Hrvatska enciklopedija*; udjel pogrešaka u uzorku

## 9. Dostatnost, usmjerenost i objektivnost sadržaja u izvorima

Recenzenti su osim ispravnosti činjenica ocjenjivali i članke prema kriterijima dostatnosti, usmjerenosti i objektivnosti, a koji su opisani u poglavlju *Elementi ocjene kvalitete sadržaja*.

Slika 7. prikazuje dodijeljene ocjene u navedenim kriterijima. Prema kriteriju dostatnosti, u kojem su recenzenti bili najkritičniji prema iznesenom sadržaju, najbolju ocjenu dobilo je 40% članaka *Wikipedije* i 62% članaka *Hrvatske enciklopedije*. Odnos djelomičnog udovoljenja zadanom kriteriju iznosi 40:32 posto. Nedovoljnu ocjenu dobilo je 16% *Wikipedijinih* članaka, nasuprot 5% iz tradicionalnoga izvora.



Slika 7. Distribucija ocjena dostatnosti, usmjerenosti i objektivnosti

Usmjerenost je kategorija u kojoj su oba izvora zabilježila bolje rezultate nego u prethodno opisanoj kategoriji. Tako je potpunu »fokusanost« zadržalo 62% slobodno uređenih članaka i visokih 87% tradicionalnih. Ovom kriteriju nije udovoljilo 10% članaka iz *wiki* izvora i 1% tradicionalnih.

Najbolje su rezultate obje enciklopedije postigle prema kriteriju objektivnosti – 91% članaka *Wikipedije* iznosi nepristrano stajalište, a takvih je gotovo 98% u *Hrvatskoj enciklopediji*. S druge strane neobjektivno je svega 2% članaka iz *wiki*-izvora i ni jedan tradicionalno uređen članak.

## 10. Subjektivna ocjena i recenzentski komentari

Uz predstavljene kvalitativne kriterije, recenzentima je u upitniku omogućen i unos vlastite subjektivne ocjene. Uputa recenzentima bila je da se ne rukovode prethodno danim ocjenama ni jedne kategorije već da odluče između članaka koji im se osobno čine prikladnijima enciklopedijskoj formi. Imali su i mogućnost usporedne članke proglasiti podjednakima. Time se željelo dobiti još jednu dimenziju u

ocjeni, iako ona nije uzeta u konačnici kao faktor ukupne kvalitete. Tablica 3. prikazuje odabir recenzenata prema subjektivnom dojmu.

Tablica 3. Odabir kvalitetnijega članka prema općem subjektivnom dojmu recenzenata

Opća kategorija	Odabir WP	Odabir HE	Podjednako
Društvo i društvene znanosti	6	22	2
Filozofija	6	10	4
Geografija	5	9	16
Kultura i umjetnost	2	14	13
Matematika i logika	5	17	8
Medicina i zdravlje	5	12	3
Povijest i biografije	3	13	2
Prirodne znanosti	8	15	7
Religija	0	8	2
Tehnologija i primijenjene znanosti	8	11	10
<b>UKUPNO</b>	<b>48</b>	<b>131</b>	<b>67</b>

Većina odabranih članaka pripala je *Hrvatskoj enciklopediji*, njih 53%, dok je 19,5% odabranih članaka iz *Wikipedije*. U 27% slučajeva recenzenti se nisu odlučili između članaka i proglasili su ih podjednakima.

Promotrimo li opće kategorije, osim *Religije*, nema posebnog odstupanja i u svima su recenzenti odabrali većinom tradicionalno uređene članke.

## Zaključak

Ovim je istraživanjem statistički opisan kvalitativni odnos *Hrvatske enciklopedije*, kao tradicionalno uređenoga enciklopedijskoga izvora, te hrvatske *Wikipedije*, kao slobodno uređenoga enciklopedijskoga izvora.

Tim stručnih suradnika, neutralnih u pitanju sklonosti određenom tipu enciklopedije, analizirao je 250 parova članaka na istu temu primjenjujući kriterije informativnosti, točnosti, dostatnosti, usmjerenosti i neutralnosti. Prema odabranim kriterijima bolje je ocjene dobila tradicionalno uređena *Hrvatska enciklopedija*, usprkos sve većem utjecaju *Wikipedije*. No svakako treba naglasiti da se radi o trenutnom stanju te da slobodno uređen koncept posjeduje kvalitetu u područjima koja ovim radom nisu obuhvaćeni. Tako jedan od temeljnih elemenata enciklopedike, dostupnost, nije ovim radom obrađen, no može se pretpostaviti da predstavlja »jaču stranu« slobodnoga koncepta. Također bi trebalo razmotriti i načela kojima je taj

koncept dobio na dinamici sadržajnih promjena, iznimno važnom elementu u današnje vrijeme brzih promjena u svijetu tehnologije i znanosti.

Pretpostavka je da će u budućnosti zaživjeti neki od hibrida koji će kombinirati jake strane obaju koncepata. Dakle, sve su izgledniji enciklopedički projekti u kojima će se masovnosti slobodnoga koncepta pridruživati jake karakteristike baštinjene tradicionalnim uređivanjem, poput primjerene strukturiranosti i pouzdanosti podataka. Bez obzira na entuzijazam velikoga broja današnjih internetskih korisnika i njihovu spremnost na slobodno dijeljenje znanja, kao i na institucije koje profesionalno sudjeluju u stvaranju i uređivanju enciklopedijskih sadržaja, rezultati objiju strana, iako udruženi, nedostatni su. Navedena količina informacija premašuje mogućnosti ručne obradbe, ako se želi održati zadovoljavajuća razina kvalitete. Stoga je važno nastaviti rad u smjeru veće automatizacije modela kvalitativne usporedbe kako bi se omogućili praćenje i promjena kvalitativnoga odnosa pri unapređenju njihovih konceptualnih i tehnoloških okvira.

**WIKIPEDIA OPPOSITE THE CROATIAN ENCYCLOPAEDIA, A QUALITATIVE  
RELATION BETWEEN A FREELY AND A TRADITIONALLY EDITED  
ENCYCLOPAEDIC TEXT IN CROATIAN**

**Ozren Kubelka**

University of Applied Sciences Vern

**Petra Šoštarić**

Faculty of Humanities and Social Sciences, Zagreb

**ABSTRACT:** Freely edited encyclopaedias, of which *Wikipedia* has the largest number of articles and users, today represent a source of information that can not be neglected. Since the concept of their editing is based on anonymity and the freedom of data input, from their very beginning there have been discussions about the reliability and quality of contents edited in such a way.

This paper tries to find out what is the current qualitative state between encyclopaedic sources created with a concept of free editing and those with a traditional editing. For sources of compared contents we chose the *Croatian Encyclopaedia* as a traditionally edited source and the Croatian edition of *Wikipedia* as a freely edited encyclopaedic content. The paper defines the elements used to determine their qualitative relation, which were tested in a review analysis in which 24 reviewers, experts in specific fields, participated. The results of the research show the current relation of sources according to the parameters of informativeness, accuracy of presented information, sufficiency, direction and objectivity.

The results of this work are a part of a wider research with which a model for a qualitative comparison of information sources is established.

**Keywords:** *The quality of encyclopaedic content; freely edited encyclopaedic content; traditionally edited encyclopaedic content; review analysis; Wikipedia; the Croatian Encyclopaedia*

#### LITERATURA I IZVORI

**Brandes, Ulrik, Lerner, Jürgen:** Visual analysis of controversy in user-generated encyclopedias. *Information Visualization* 7(2008), str. 34.

**Fallis, Don:** Toward an epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology* 59(2008), str. 1662.

**Giles, Jim:** Special Report Internet encyclopedias go head to head. *Nature* (2005)348, str. 900–901.

<http://www.lzmk.hr/hr>

**Kubelka, Ozren:** *Komparativni model za utvrđivanje kvalitete enciklopedijskoga sadržaja*, doktorski rad, 2013.

**Korfiatis, Nikolaos Th., Poulos, Marios, Bokos, George:** Evaluating authoritative sources using social networks: an insight from Wikipedia. *Online Information Review* 30(2006) 3, str. 252–262.

**Lih, Andrew:** *The Wikipedia Revolution*. Hyperion e-books, 2009.

**Lim, Ee-Peng, Vuong, Ba-Quy, Lauw, Hady Wirawan, Sun, Aixin:** Measuring Qualities of Articles Contributed by Online Communities, u: *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, 2006.

**Njari, Denis.** Prva hrvatska enciklopedija – osječki doprinos hrvatskoj enciklopedistici. *Hrvatistika* 3(2009) 3.

**Rector, Lucy Holman:** Comparison of Wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review* 36(2007) 1, str. 7–22.

**Sunstein, Cass R.:** *Infotopia: How Many Minds Produce Knowledge*. Oxford University Press. 2006.

[wikipedija.org](http://wikipedija.org)

**Wu, Kewen, Zhu, Qinghua, Zhao, Yuxiang, Zheng, Hua:** Mining the Factors Affecting the Quality of Wikipedia Articles, u: *International Conference of Information Science and Management Engineering*, 2010.