Strukturiranost i interoperabilnost mrežnih enciklopedičkih sadržaja iz tehničkoga područja

Smolčić, Ivan

Doctoral thesis / Disertacija

2020

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

https://doi.org/10.17234/diss.2020.8746

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:131:220947

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Filozofski fakultet
Sveučilišta u Zagrebu

Repository / Repozitorij:

ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI



SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET

Ivan Smolčić

STRUKTURIRANOST I INTEROPERABILNOST MREŽNIH ENCIKLOPEDIČKIH SADRŽAJA IZ TEHNIČKOGA PODRUČJA

Doktorski rad

Zagreb, 2020.



SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET

Ivan Smolčić

STRUKTURIRANOST I INTEROPERABILNOST MREŽNIH ENCIKLOPEDIČKIH SADRŽAJA IZ TEHNIČKOGA PODRUČJA

DOKTORSKI RAD

Mentori:
doc. dr. sc. Petra Bago
dr. sc. Zdenko Jecić

Zagreb, 2020.



UNIVERSITY OF ZAGREB
FACULTY OF HUMANITIES AND SOCIAL SCIENCES

Ivan Smolčić

STRUCTURE AND INTEROPERABILITY OF ONLINE ENCYCLOPEDIC CONTENT IN THE FIELD OF TECHNOLOGY

DOCTORAL THESIS

Supervisors:
Petra Bago, PhD
Zdenko Jecić, PhD

Zagreb, 2020.

MENTORI

Petra Bago rođena je 18. studenoga 1982. godine u Zagrebu. Osnovnu je školu pohađala u Zagrebu, Bernu (Švicarska) i Sofiji (Bugarska), a V. gimnaziju prirodoslovno-matematičkog usmjerenja završila je u Zagrebu.

Godine 2003. upisala je smjer informacijskih znanosti i etnologije i kulturne antropologije na Filozofskom fakultetu u Zagrebu. Od 2004. sudjelovala je u nastavi kao demonstratorica na Odsjeku za informacijske i komunikacijske znanosti. Kao suautorica 2006. godine primila je Rektorovu nagradu za studentski znanstveni rad pod naslovom *Računalna priprema, obrada i prikaz sedmojezičnog rječnika Petera Lodereckera povodom 400. godišnjice izdavanja* izrađen pod vodstvom prof. dr. sc. Damira Borasa i Nikole Ljubešića, tadašnjeg znanstvenog novaka na Odsjeku. Diplomirala je na Odsjeku za informacijske i komunikacijske znanosti i Odsjeku za etnologiju i kulturnu antropologiju 2008. godine.

Akademske godine 2008/09. upisala je poslijediplomski doktorski studij na Odsjeku za informacijske znanosti. Od 2009. godine zaposlena je na istom odsjeku kao znanstvena novakinja na projektu Hrvatska rječnička baština i hrvatski europski identitet pod vodstvom prof. dr. sc. Damira Borasa. Doktorirala je u rujnu 2014. godine tezom *Model višerazinske prezentacije tekstova starijih hrvatskih rječnika* koju je izradila pod mentorstvom prof. dr. sc. Damira Borasa i doc. dr. sc. Nikole Ljubešića.

Aktivno sudjeluje u izvođenju nastave na preddiplomskoj i diplomskoj razini na kolegijima Računalne mreže, Obrada teksta i jezika, Jezične baze podataka i Prikaz rječničkog znanja. Pohađala je više tečajeva, ljetnih škola i europskih intenzivnih programa u zemlji i inozemstvu.

MENTORI

Zdenko Jecić rođen je 20. siječnja 1967. godine u Zagrebu, gdje je maturirao u Klasičnoj gimnaziji 1985. godine te diplomirao na Arhitektonskom fakultetu Sveučilišta u Zagrebu 1993. godine. Magistrirao je na Građevinskom fakultetu Sveučilišta u Zagrebu 1999. godine obranivši magistarski rad naslova *Primjena modela pri određivanju oblika vlačno napregnutih konstrukcija*. Kao stipendist Vlade Republike Austrije boravio je 1990. godine u Grazu na usavršavanju njemačkoga jezika te u ljetnom semestru 1997. godine na znanstvenom usavršavanju na Institutu za visokogradnju Tehničkog sveučilišta u Beču.

Od 1993. godine bio je znanstveni novak na Arhitektonskom fakultetu u Zagrebu, sudjelujući na znanstvenim projektima te vodeći vježbe iz kolegija Tehnička mehanika I. i II. Godine 1998. izabran je u istraživačko zvanje mlađeg asistenta, a 1999. godine u asistenta. Doktorirao je 2008. godine obranivši doktorsku disertaciju naslova *Virtualna enciklopedija: Redefiniranje zadaće enciklopedijske djelatnosti* koju je izradio pod mentorstvom prof. dr. sc. Damira Borasa.

Od 2000. godine zaposlen je kao stručni suradnik u Leksikografskom zavodu Miroslav Krleža, od 2003. godine u istraživačkom zvanju asistenta, a od 2005. godine i u zvanju leksikografa. Od dolaska u Zavod radio je kao urednik i autor članaka iz tehničkih područja Hrvatske enciklopedije, Tehničkog leksikona te Hrvatskog općeg leksikona. Od 2014. godine pokretač je i voditelj opsežnoga projekta Hrvatske tehničke enciklopedije Leksikografskog zavoda Miroslav Krleža. Znanstveno se bavi područjem primjene informacijske tehnologije u leksikografiji, autor je više znanstvenih i stručnih radova te poglavlja u knjizi.

ZAHVALA

Zahvaljujem doc. dr. sc. Petri Bago i dr. sc. Zdenku Jeciću na mentorsvu tijekom pohađanja studija te iznimnoj pomoći pri izradi ove doktorske disertacije.

Također, zahvaljujem svim profesorima i zaposlenicima Odsjeka za informacijske i komunikacijske znanosti zagrebačkoga Filozofskog fakulteta koji su kroz doktorski studij pridonijeli stvaranju poticajne radne sredine i omogućili moj znanstveni razvoj.

Zahvaljujem Leksikografskom zavodu Miroslav Krleža, odnosno kolegama na potpori, pomoći i razumijevanju tijekom pohađanja doktorskog studija i izrade doktorske disertacije, posebno jezičnoj urednici Ivani Jović, koja je svojim primjedbama i komentarima znatno unaprijedila kvalitetu rada.

Posebnu zahvalu upućujem kolegama karlovačkoga poduzeća Infinum d. o. o. na pruženoj tehničkoj podršci pri izradi ove doktorske disertacije.

Na kraju, hvala mojoj obitelji, prijateljima i bližnjima na riječima potpore, strpljenju i razumijevanju za moje duge odsutnosti tijekom istraživačkoga rada i pisanja doktorske disertacije. Najveća hvala mojoj djevojci Jasmini koja je sve proživljavala zajedno sa mnom, zatim roditeljima Mirjani i Tomislavu, bakama Mariji i Silvestri, djedu Josipu, braći Jurju i Mislavu, sestri Katarini.

I naravno, kućnom ljubimcu Pepi koji je budno pratio svaki moj pokret.

Ivan Smolčić

PREDGOVOR

Visoka informativnost, sveobuhvatnost i organizacija znanja čine enciklopedije i njima srodne publikacije, poput leksikona ili riječnika, opsežnim djelima koja korisniku pružaju učinkovit dohvat informacija vezanih uz područje interesa. Takav postav enciklopedičkih djela, koji seže od njihovih početaka, usporedan je s funkcioniranjem suvremenih informacijskih sustava.

Stoga enciklopedička djela postavljena u mrežnome prostoru predstavljaju vrijedan resurs, koji se izgradnjom mrežnih informacijskih sustava (aplikacija) i organizacijom sadržaja može nadograditi u epistemološkom smislu diseminacije konsolidiranoga sadržaja. To se ponajprije odnosi na mogućnosti strukturiranja sadržaja mrežnih enciklopedičkih projekata na temelju njihove visoke informativnosti, odnosno faktografske vrijednosti tekstova enciklopedičkih članaka. Ta praksa već je prisutna u izvedbi nekih mrežnih enciklopedičkih izdanja, a ističu se posebno inačice Wikipedije i Encyclopaedije Britannice donošenjem strukturiranih, temeljnih podataka o promatranim pojmovima u obliku infookvira. Takav način predstavljanja mrežnog enciklopedičkog sadržaja omogućuje učinkovitije informiranje korisnika, ali i izgradnju hipertekstualne mreže koja pospješuje navigaciju pri dohvaćanju sadržaja, te postavljanje složenih upita nad bazama podataka i dohvaćanja strukturiranoga sadržaja. Takva struktura temelji se na metapodatcima koji sadržaju objavljenom u mrežnom prostoru daju semantičku vrijednost u računarskome smislu.

Zahvaljujući strukturiranom sadržaju, moguće je ostvariti njegovu interoperabilnost s drugim izvorima informacija koji su izvedeni na isti način. Time mrežni enciklopedički projekti prestaju djelovati zasebno i djeluju kao jedan informacijski sustav temeljen na mogućoj razmjeni međusobno relevantnog sadržaja. Primjer toga je projekt DBpedije kao semantičke preslike Wikipedijinoga sadržaja objavljenoga na način da u tehničkome smislu, zadovoljavajući sintaktičke i semantičke zakonitosti pri kodiranju sadržaja i stvaranja strukture, ostvaruju uvjete razmjene strukturiranoga sadržaja s drugim skupinama podataka objavljenima po istom principu.

Cilj ove doktorske disertacije je provesti kvantitativnu i kvalitativnu analizu sadržaja tekstova jedinica enciklopedičkih djela (članaka ili natuknica) iz tehničkoga područja te

iz rezultata dobivenih tom analizom uvidjeti mogućnosti strukturiranja takva sadržaja na temelju uočenih unificiranih faktografskih podataka kao predstavnika enciklopdičke informativnosti. Analiza sadržaja provest će se na ukupno deset enciklopedičkih izdanja, rezultati kojih će se donijeti za svako izdanje posebno, ali i zbirno u svrhu potvrđivanja faktografskoga standarda enciklopedičkih sadržaja objavljenih u mrežnome prostoru. Na teorijskim spoznajama o interoperabilnosti mrežnoga strukturiranog sadržaja, nastojat će se omogućiti interoperabilnost enciklopedičkoga sadržaja projekata uključenih u ovo istraživanje, strukturiranoga na temelju rezultata analize sadržaja uspostavom modela koji će omogućiti njegovu razmjenu.

SAŽETAK

Izvedbom enciklopedičkih djela kao suvremenih mrežnih projekata omogućena je nadogradnja tih informacijskih sustava u epistemološkome smislu što upućuje na potrebu redefiniranja leksikografske i enciklopedičke djelatnosti. Zadržavajući temeljne (tradicionalne) odrednice enciklopedičkoga koncepta, poput točnosti, objektivnosti i relevantnosti, mrežna enciklopedička izdanja nadilaze granice tradicionalne enciklopedike unaprjeđenjem epistemoloških značajki svoga sadržaja. Ovaj rad posebno obrađuje problematiku i vezu između strukturiranja i interoperabilnosti mrežnih enciklopedičkih sadržaja. Kako bi se donijeli zaključci o mogućnostima strukturiranja enciklopedičkoga sadržaja na temelju njegove sažetosti (informativnosti), provedena je kvantitativna i kvalitativna analiza sadržaja u cilju bilježenja unificiranih faktografskih podataka pogodnih za stvaranje strukture.

Ispitano je ukupno 455 kategoriziranih enciklopedičkih članaka iz tehničkoga područja iz 10 zasebnih mrežnih enciklopedičkih projekata na hrvatskom, engleskom i njemačkom jeziku. Kako bi se pospješilo prepoznavanje naziva u enciklopedičkim tekstovima, ispitane su vrijednosti evaluacijskih mjera alata namijenjenih automatskom prepoznavanju nazivlja (engl. *Named Entity Recognition*) na enciklopedičkim tekstovima uključenima u uzorak analize sadržaja.

Izračunom standardnih evaluacijskih mjera (preciznost, odziv, točnost, F mjera) na temelju parametara proizašlih iz označenoga teksta s pomoću alata doneseni su rezultati analize evaluacijskih mjera označavanja alata za pojedine kategorije i vrste naziva na mješovitom uzorku sastavljenom od članaka više enciklopedičkih izdanja. Na temelju analize sadržaja prikazan je način strukturiranja enciklopedičkoga sadržaja te korištenje te strukture u postizanju interoperabilnosti putem izrade posebnoga modela.

U konačnici, ova doktorska disertacija donosi zaključak vezan za sve njegove sastavnice, odnosno skreće pozornost na enciklopedičke vrline kroz postavke enciklopedičkog koncepta, ulogu strukture i mogućnosti postizanja interoperabilnosti mrežnog enciklopedičkog sadržaja.

EXTENDED ABSTRACT

The implementation of encyclopedic works as contemporary network projects enabled the upgrade of these information systems in an epistemological sense, ie the necessary redefinition of lexicographical and encyclopedic activities. While retaining the basic (traditional) determinants of the encyclopedic concept, such as accuracy, objectivity and relevance, online encyclopedic editions go beyond the traditional ones in their content creation, removed scope limits (thanks to digital content), enhanced content retrieval and user adaptability.

This work addresses in detail the issues of structure and interoperability of encyclopedic content, which enable the encyclopedic content to interact and function in the global information system. An overview of the main features and the application of the structure and interoperability of encyclopedic content in the network space is provided. In order to draw conclusions about the possibilities of structuring encyclopedic content on the basis of its informative nature, a quantitative and qualitative content analysis was conducted in order to record unified factual data suitable for structure creation.

A total of 455 encyclopedic articles from 10 separate online encyclopedic projects in Croatian, English and German were examined. Entities recorded in the content analysis of encyclopedic texts were divided into three categories of encyclopedic articles: persons, organizations and general terms. The article category for persons contains a total of 33 types of entities suitable for structuring, the article category for organization marks 23 entity types, and the article category for general terms contains only seven entity types.

In order to facilitate automatic identification of named entities in encyclopedic texts, the effectiveness of Named Entity Recognition tools on encyclopedic texts in the technical field has been tested. Three tools (applications) were used for testing: CroNER and ReLDI for the Croatian samples and the Stanford CoreNLP tool for the samples in English. Based on the parameters derived from the tool-marked text, using a mixed sample of articles from multiple encyclopedic publications, standard performance measures (precision, response, accuracy, F measure) were calculated for each tool, providing evaluation of the tools for individual categories and types of names.

On the basis of the content analysis, the method of structuring the encyclopedic content and the use of that structure in achieving interoperability by developing a special model are presented. The model represents a central database that binds the structure of multiple editions and enables it to be exchanged to all related online encyclopedic editions by mapping the elements of the structure. Interoperability has been proven via experimental method by testing models and detecting changes.

Ultimately, this PhD thesis draws a conclusion regarding all its constituents, that is, points to encyclopedic virtues through the settings of the encyclopedic concept, the role of structure and the possibility of interoperability of online encyclopedic content.

KLJUČNE RIJEČI

enciklopedika; leksikografija; enciklopedički koncept; strukturiranost; interoperabilnost; automatsko prepoznavanje nazivlja; analiza sadržaja; model interoperabilnosti; metapodatci

KEY WORDS

encyclopaedic science; lexicography; encyclopedic concept; content structure; interoperability; Named Entity Recognition; content analysis; interoperability model; metadata

SADRŽAJ

1. UVOD 1

1.1. Motivacija 1

1.2. Polazišta rada 1

1.3. Ciljevi i hipoteze istraživanja 5

1.4. Sastavnice rada 5

2. EPISTEMOLOŠKE ODREDNICE ENCIKLOPEDIČKOGA SADRŽAJA – ENCIKLOPEDIČKI KONCEPT 8

2.1. Epistemološko gledište enciklopedičkoga sadržaja 9

2.2. Tradicionalne epistemološke odrednice enciklopedičkoga sadržaja 14

2.2.1. Točnost 14

2.2.2. Objektivnost 19

2.2.3. Relevantnost 21

2.2.4. Pouzdanost i vjerodostojnost 23

2.2.5. Sveobuhvatnost 23

2.2.6. Konsolidiranost 24

2.2.7. Sažetost 25

2.2.8. Organiziranost 26

2.3. Epistemološke odrednice suvremenih mrežnih enciklopedičkih projekata 27

2.3.1. Kontinuiranost i kolaborativnost 28

2.3.2. Neograničenost opsega 31

2.3.3. Dohvaćanje (pretraživost) i povezanost sadržaja 33

2.3.4. Adaptivnost 34

3. STRUKTURIRANOST I INTEROPERABILNOST MREŽNIH ENCIKLOPEDIČKIH SADRŽAJA 36

3.1. Metapodatci 36

3.2. Strukturiranost 38

3.3. Enciklopedička strukturiranost 40

3.3.1. Hrvatska tehnička enciklopedija 41

3.3.2. Proleksis enciklopedija 46

3.3.3. Encyclopaedia Britannica 48

3.3.4. Wikipedija 49

3.4. Interoperabilnost 53

3.4.1. Uvjeti postizanja interoperabilnosti 54

3.4.2. Sheme (standardi) metapodataka 55

3.4.3. Kontrolirani rječnici 56

3.4.4. Mapiranje elemenata strukture 57

3.4.5. Enciklopedička interoperabilnost 58

3.4.5.1. DBpedia 58

3.4.5.1.1. Semantički web 62

3.4.5.2. Wikidata 64

3.4.5.3. Gran Enciclopedia Aragonesa 66

4. ANALIZA SADRŽAJA MREŽNOGA ENCIKLOPEDIČKOG SADRŽAJA IZ TEHNIČKOGA PODRUČJA 68

4.1. Obuhvaćena djela 69
4.1.1. Hrvatska tehnička enciklopedija 69
4.1.2. Hrvatska enciklopedija 70
4.1.3. Proleksis enciklopedija 71
4.1.4. Istarska enciklopedija 72
4.1.5. Hrvatski biografski leksikon 72
4.1.6. Encyclopaedia Britannica 73
4.1.7. Der Brockhaus multimedial 2008 premium 74
4.1.8. Wikipedija 75
4.2. Metodologija 76
4.3. Rezultati istraživanja 78
4.3.1. Analiza sadržaja uzorka pojedinih mrežnih enciklopedičkih djela 78
4.3.1.1. Hrvatska tehnička enciklopedija 82
4.3.1.2. Hrvatska enciklopedija 87
4.3.1.3. Proleksis enciklopedija 90
4.3.1.4. Istarska enciklopedija 94
4.3.1.5. Hrvatski biografski leksikon 97
4.3.1.6. Encyclopaedia Britannica 100
4.3.1.7. Der Brockhaus multimedial 2008 premium 103
4.3.1.8. Wikipedija 106
4.3.2. Analiza sadržaja jedinstvenoga uzorka i usporedba rezultata zasebnih izdanja 112

5. ISPITIVANJE ALATA ZA PREPOZNAVANJE NAZIVLJA NA ENCIKLOPEDIČKIM TEKSTOVIMA 119

5.1. Obrada prirodnog jezika 119

5.2. Automatsko prepoznavanje nazivlja 120

5.2.1. Metodologija 123

5.2.1.1. Korišteni alati 124

5.2.2. Evaluacijske mjere 129

5.2.2.1. Preciznost (P) 130

5.2.2.2. Odziv (R) 131

5.2.2.3. Točnost (A) 131

5.2.2.4. F mjera (F1) 132

5.2.3. Rezultati istraživanja 132

5.2.3.1. Analiza uzorka 132

5.2.3.2. Ispitivanje alata 135

5.2.3.2.1. Kategorije članaka i zasebne jedinice sadržaja 135

5.2.3.2.2. Nazivi 139

6. USPOSTAVA MODELA INTEROPERABILNOSTI MREŽNOGA ENCIKLOPEDIČKOG SADRŽAJA IZ TEHNIČKOGA PODRUČJA 142

6.1. Metodologija izrade modela 142

6.2. Implementacija modela enciklopedičke interoperabilnosti 143

6.2.1. Značajke i shematski prikaz modela enciklopedičke interoperabilnosti 143

6.2.2. Uspostava modela enciklopedičke interoperabilnosti u mrežnome prostoru 147

6.2.3. Testiranje modela enciklopedičke interoperabilnosti i rezultati 150

7. ZAKLJUČAK 155

7.1. Epistemološke odrednice enciklopedičkoga sadržaja – enciklopedički koncept 155

7.2. Strukturiranost i interoperabilnost mrežnih enciklopedičkih sadržaja 156

7.3. Analiza sadržaja mrežnoga enciklopedičkog sadržaja iz tehničkoga područja 158

7.4. Ispitivanje alata za prepoznavanje nazivlja na enciklopedičkim tekstova 159

7.5. Uspostava modela interoperabilnosti mrežnoga enciklopedičkog sadržaja iz tehničkoga područja 161

8. LITERATURA 164

9. BIBLIOGRAFSKI ZAPISI ENCIKLOPEDIČKIH DJELA UKLJUČENIH U ISTRAŽIVANJE 173

10. PRILOZI 174

11. POPIS OZNAKA 193

12. POPIS SLIKA 194

13. POPIS TABLICA 198

14. POPIS PRILOGA 200

15. ŽIVOTOPIS 201

16. CURRICULUM VITAE 203

1. UVOD

1.1. MOTIVACIJA

Postavljanjem enciklopedičkih sadržaja kao mrežnih projekata omogućeno je njihovo usklađivanje s izvedbenim mogućnostima informatičkih sustava (aplikacija), što spada među najintenzivnije promjene što ih enciklopedički koncept u posljednje doba doživljava. Leksikografski zavod Miroslav Krleža kao središnja hrvatska leksikografska i enciklopedička institucija započeo je s objavljivanjem svojih edicija u mrežnome prostoru još 2008. godine. S obzirom na nedovoljno iskorištene mogućnosti, javlja se potreba za preoblikovanjem toga sadržaja, njegovom organizacijom i razvojem funkcionalnosti kako bi se kao javni sadržaj prilagodio korisnicima. Analizom enciklopedičkoga sadržaja s gledišta informacijskih znanosti moguće je ostvariti dodatna saznanja u cilju izgradnje enciklopedičke strukture koja će služili u izvedbi, organizaciji, korištenju i interoperabilnosti mrežnih enciklopedičkih sadržaja.

1.2. POLAZIŠTA RADA

Enciklopedička djela kao sintetska djela oduvijek su težila totalitetu znanja, odnosno prikupljanju, verificiranju, obrađivanju i objavljivanju ljudskih dosega u vremenu kada su nastajala. Njihova organiziranost omogućuje korisniku brzo i lako pronalaženje korisne informacije. Stoga se princip organizacije enciklopedičkih djela smatra osnovom enciklopedičkoga koncepta, a svojim značenjem nadilazi granice enciklopedike. [1] Organizacijska struktura tradicionalne enciklopedike zasniva se na uvriježenom abecednom ili konceptualnom slijedu sadržaja podijeljenom na jedinice (članke ili natuknice) različita opsega s poveznicama između jedinica sadržaja, te indeksu.[2] Dodjeljivanje naslova (oznaka) člancima, uspostavljanje veza među njima i izrada indeksa kao temelja organizacije tradicionalnih

[1] Jecić, Z.: Enciklopedički koncept u mrežnom okruženju. Studia lexicographica, 7(2013) 2(13), str. 99–115.; Jecić, Z., Boras, D., Domijan, D.: Prilog definiranju pojma virtualna enciklopedija. Studia lexicographica, 2(2008) 1(2), str. 115–126.

[2] Katz, W. A.: Introduction to Reference Work, Volume I. Basic Information Sources. New York, 1978.; Vujić, A.: Utemeljenje enciklopedijske leksikografije kao informacijske znanosti. U: Radovi zavoda za informacijske studije, knj. 1. Zagreb, 1990, str. 141–146.; Vujić, A.: Acta Lexicographica. Prema znanstvenom utemeljenju enciklopedijske leksikografije. Zagreb, 2015.

enciklopedičkih izdanja u uskoj su vezi s prikupljanjem metapodataka koji čine osnovu funkcionalnosti današnjih mrežnih enciklopedičkih djela.

Metapodatci predstavljaju informacije kojima se opisuje neki predmet ili sadržaj, a omogućuju interakciju s tim sadržajem. Općenito, informacijski sustavi prožeti su metapodatcima koji se pojavljuju u raznim oblicima. Ugrađeni u mrežne stranice, čine osnovu njihove funkcionalnosti, omogućuju povezivanje između mrežnih stranica, ključni su u obavljanju digitalnih poslovnih transakcija, na njima počivaju društvene mreže itd. Metapodatcima suvremene knjižnice, arhivi, muzeji i druge ustanove označavaju svoju građu, a razmjenom i dijeljenjem metapodataka omogućuju jednostavniji i učinkovitiji dohvat njihove zajedničke građe, odnosno pretraživanje sadržaja.[3]

Označavanje mrežnih enciklopedičkih djela ključno je prigodom njihove prezentacije, a ujedno omogućuje kontrolirano stvaranje mrežnoga izdanja urednicima, te ugodnu i učinkovitu uporabu korisnicima. Voss upućuje na organizaciju znanja najčešće korištene mrežne enciklopedije Wikipedije[4], te navodi kako kolaborativno označavanje[5] rezultira organizacijom znanja u obliku tezaurusa, odnosno kontroliranoga rječnika povezanih pojmova koji se mogu koristiti kao ključne riječi, te ujedno predstavlja kompromis između ontologijskog pristupa i najjednostavnijeg abecednog nizanja.[6] Dodatnu strukturiranost enciklopedičkoga sadržaja Wikipedia postiže izvedbom infookvira, kao tabličnih prikaza podataka vezanih uz enciklopedički članak.[7] Kontinuiranom ekstrakcijom strukturiranih podataka iz Wikipedijine građe nastaje ontologijska baza podataka DBpedia, koja se sastoji od instanci (osnovnih

[3] Haynes, D.: Metadata for information management and retrieval. London, 2004.; Baca, M.: Introduction to metadata. Los Angeles, 2008.; Riley, J.: Understanding metadata: What is metadata, and what is it for?. Baltimore, 2017.

[4] Voss, J.: Collaborative thesaurus tagging the Wikipedia way, 2006. https://arxiv.org/abs/cs.IR/0604036 (26. 7. 2018.).

[5] Voss, J.: Tagging, Folksonomy & Co - Renaissance of Manual Indexing?. U: 10th international Symposium for Information Science. Cologne, 2007.; Mika, P. et al.: Learning to tag and tagging to learn: A case study on Wikipedia. IEEE Intelligent Systems, 23(2008) 5, str. 26–34.

[6] Francisco, J., Garcia, M.: The evolution of thesauri and the history of knowledge organization: Between the sword of mapping knowledge and the wall of keeping it simple. Brazilian Journal of Information Studies: Research Trends, 10(2016) 1, str. 1–11.

7 Lange, D., Böhm, C., Naumann, F.: Extracting structured information from Wikipedia articles to populate infoboxes. U: Proceedings of the 19th ACM international conference on Information and knowledge management. New York, 2010., str. 1661–1669.; Sultana, A. et al.: Infobox suggestion for Wikipedia entities. U: Proceedings of the 21st ACM international conference on Information and knowledge management. Maui, 2012., str. 2307–2310.

objekata početne razine), klasa (zbirki ili tipova objekata), atributa (pripadajućih svojstava, pojava, karakteristika ili parametara koje objekt može imati ili distribuirati) i odnosa (načina na koji se objekti odnose jedni prema drugima).[8] Starčević Stančić i Kraus donose pregled razvoja Hrvatske enciklopedije Leksikografskoga zavoda Miroslav Krleža od tiskanoga do mrežnoga izdanja, navodeći prednosti u organizaciji sadržaja koji je za razliku od abecedno nizanih natuknica po svescima u tiskanom izdanju, označavanjem mrežnoga izdanja sveden na taksonomiju od pet područja koja obuhvaćaju 126 struka u koje je raspoređen cjelokupan korpus enciklopedije.[9]

Metapodatci su ključni u ostvarivanju interoperabilnosti informacijskih sustava, u koje spadaju i mrežna enciklopedička izdanja. Haslhofer i Klas proveli su istraživanje o problemima i načinu postizanja interoperabilnosti metapodataka heterogenih repozitorija.[10] Uz objašnjenje osnovnih pojmova poput standarda metapodataka, programskih jezika i dr., te izgradnje i pojavnosti metapodataka unutar informatičkih sustava, detaljno navode heterogenosti koje ometaju uspostavljanje interoperabilnosti između dvaju sustava te rješenja, odnosno tehnike uspostavljanja interoperabilnosti metapodataka pohranjenih u udaljenim repozitorijima, poput ujednačavanja standarda (shema) metapodataka[11], kombiniranja više shema metapodataka i prilagođavanja vokabulara metapodataka prema potrebi[12] i dr. Chan i Zeng donose analizu metoda koje su se rabile za postizanje ili poboljšanje interoperabilnosti između shema (standarda, specifikacija) metapodataka i objavljenih aplikacija, u svrhu olakšavanja pretvorbe i razmjene metapodataka i omogućavanja sakupljanja metapodataka iz više domena i zajedničkoga pretraživanja.[13]

[8] Lehmann, J., Isele, R., Jakob, M.: DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. Semantic Web, 6(2015) 2, str. 167–195.

[9] Starčević Stančić, I., Kraus, C.: Hrvatska enciklopedija – od tiskanoga do mrežnoga izdanja. Studia lexicographica, 8(2014) 1(14), str. 99–116.

[10] Haslhofer, B., Klas, W.: A Survey of Techniques for Achieving Metadata Interoperability. ACM Computing Surveys (CSUR), 42(2010) 2.

[11] Rahm, E., Bernstein, P. A.: A survey of approaches to automatic schema matching. The VLDB Journal, 10(2001) 4, str. 334–350.

[12] Heery, R., Patel, M.: Application profiles: Mixing and Matching Metadata Schemas. Ariadne-Web magazine for information professionals, (2000) 25.; Baker, T., Dekkers, M., Heery, R.: What Terms Does Your Metadata Use?. Application Profiles as Machine-Understandable Narratives. U: International Conference on Dublin Core and Metadata Applications. Tokyo, 2001., str. 151–159.; Li, J., Gao, J., Dong, J.: A metadata registry for metadata interoperability. Data science Journal, 6(2007), str. 379–384.

[13] Chan, L. M., Zeng, M. L.: Metadata Interoperability and Standardization – A Study of Methodology (Part I and II). D – Lib Magazine, 12(2006) 6.

Primjer postizanja enciklopedičke interoperabilnosti je DBpedia, ontologijska baza podataka koja predstavlja semantičko ogledalo Wikipedije. Također predstavlja nadogradnju Wikipedijine organizacije znanja nastalu ekstrakcijom strukturiranih podataka (metapodataka) iz Wikipedijine građe.[14] Dio je skupine baza podataka objedinjenih terminom Linked Open Data, odnosno metode objavljivanja međusobno povezanih strukturiranih podataka pretraživih s pomoću semantičkih upita. Temelji se na standardnim mrežnim tehnologijama kao što su HTTP (engl. *Hyper Text Transfer Protocol*), RDF (engl. *Resource Description Framework*) i URI (engl. *Uniform Resource Identifier*), no umjesto da ih koristi za posluživanje mrežnih stranica namijenjenih ljudima, proširuje ih na dijeljenje informacija na način koji se automatski može čitati s pomoću računala. Tako je omogućena integracija heterogenih informatičkih sustava, odnosno njihova interoperabilnost. Svi skupovi podataka izvedeni su s pomoću RDF specifikacije za oblikovanje metapodataka čime su stvoreni uvjeti za postizanje interoperabilnosti i strojnu obradu podataka. DBpedia sa svojom širokom tematskom pokrivenošću presijeca velik dio skupova podataka i stoga tvori izvrstan povezujući čvor za takve napore u pogledu povezivosti i interoperabilnosti podataka.

Iz sadržaja Wikipedije također proizlazi Wikidata, baza strukturiranih podataka koju svatko može uređivati. Omogućuje strukturirane podatke koje mogu lako koristiti razne mrežne aplikacije. Zasniva se na zajednici koja uređuje i održava njen sadržaj. Za razliku od DBpedie koja povremeno ekstrahira podatke iz Wikipedijine građe te ovisi o njenom sadržaju, urednici Wikidate omogućuju njenu nadogradnju i ažuriranje neovisno o sadržaju Wikipedije. Također je njena struktura kodirana u RDF standardu, te implementirana u zajednicu semantičkoga weba.[15]

Sličnu mogućnost interoperabilnosti enciklopedičkoga sadržaja prikazali su Garrido, Tramullas i Martinez označavanjem digitaliziranoga teksta Gran Enciclopedie Aragonese prema standardu tematskih mapa (engl. *topic map*) s glavnim ciljem organizacije, korištenja i

[14] Pattuelli, C., Rubinow, S.: The knowledge organization of DBpedia: a case study. Journal of Documentation, 69(2013) 6, str. 762–772.; Auer S., Bizer C., Kobilarov, G.: DBpedia: A Nucleus for a Web of Open Data. U: The Semantic Web. 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007. Busan, 2007., str. 722–735.

[15] Erxleben, F. et al.: Introducing Wikidata to the Linked Dana Web. U: Proceedings of the 13th International Semantic Web Conference - Part I. Riva del Garda, 2014., str. 50–65.; Piscopo, A. et al.: What do Wikidata and Wikipedia Have in Common? An Analysis of their Use of External References. U: Proceedings of the 13th International Symposium on Open Collaboration. Galway, 2017.

razmjene sustava za organizaciju znanja.[16] Mrežna verzija standarda poznata je pod nazivom XML Topic Maps s ciljem snalaženja u strukturama znanja određenog područja i njihovog povezivanja s informacijskim izvorima. Glavna zamisao tematskih mapa sadržana je u konceptu temeljnog modela koji se sastoji od naziva, izvora i vrsta odnosa (engl. *types, occurrences, associations*).

1.3. CILJEVI I HIPOTEZE ISTRAŽIVANJA

Cilj rada je na temelju rezultata analize sadržaja strukturirati mrežni enciklopedički sadržaj te na temelju implementirane strukture razviti model interoperabilnosti enciklopedičkih sadržaja u mrežnome prostoru. Hipoteze rada su:

H1: Faktografski stil enciklopedičkih tekstova u općim i njima srodnim enciklopedičkim djelima omogućuje identifikaciju istovjetnih vrsta podataka kao temeljnih vrsta podataka ključnih za određeni tip enciklopedijskoga članka zadanih pravilima dobre prakse u području enciklopedike.

H2: Alati namijenjeni automatskom prepoznavanju nazivlja točniji su na primjeni enciklopedičkih tekstova nego na tekstovima na kojima su razvijeni, a koji su općega sadržaja.

H3: Na temelju strukturiranoga enciklopedičkog sadržaja moguća je izrada modela enciklopedičke interoperabilnosti u mrežnome prostoru.

1.4. SASTAVNICE RADA

Drugo poglavlje donosi pregled temeljnih epistemoloških odrednica enciklopedičkih izdanja koje u cjelini čine enciklopedički koncept. Na temelju epistemoloških odrednica upućuje se na ulogu enciklopedičkih djela kao nositelja obavijesti (informacija) unutar cjelokupnog informacijskog sustava. Na taj se način pozicionira i enciklopediku kao znanstvenu disciplinu unutar polja informacijskih znanosti. Kao temeljne odlike

[16] Garrido P., Tramullas, J., Martinez, F. J.: Application of Semantic Tagging to Generate Superimposed Information on a Digital Encyclopedia. U: Metadata and Semantic Research, 4th International Conference. Alcalà de Henares, 2010., str. 84–94.

enciklopedičkih djela navode se točnost kao omjer iznesenih brojčanih podataka i faktografskih činjenica, objektivnost kao mjerilo neutralnosti ili lišenosti predrasuda i ravnopravnog tretiranja činjenica, relevantnost kao zastupljenost određenih pojmova i informacija u interesu korisnika, vjerodostojnost kao mjerilo pouzdanosti pri korištenju enciklopedija kao izvora informacija. Donosi se pregled i drugih temeljnih osobina enciklopedičkoga sadržaja poput sveobuhvatnosti, konsolidiranosti i sažetosti informacija, te organiziranosti cjelokupnoga sadržaja. Pojavom enciklopedičkih izdanja kao mrežnih projekata otvorile su se nove mogućnosti u unapređenju enciklopedičkoga sadržaja u epistemološkom smislu. Kao odlika mrežnoga enciklopedičkog sadržaja, uz zadržavanje onih tradicionalnih, ističu se kontinuiranost rada na sadržaju i kolaborativnost s korisnicima, neograničenost opsega zahvaljujući novim medijima kao nositeljima sadržaja, mogućnosti dohvaćanja i povezivanja sadržaja na mreži, te mogućnosti prilagodbe sadržaja u cilju njegova učinkovita korištenja (adaptivnost).

Treće poglavlje u epistemološkom smislu nadovezuje se na drugo te donosi definicije pojmova strukturiranosti i interoperabilnosti i pregled njihove zastupljenosti u izvedbi mrežnih enciklopedičkih projekata. Kao temelj tih dviju odlika mrežnih enciklopedičkih sadržaja navode se metapodatci kao dijelovi informacijskih sustava koji uz opisivanje, definiranje i lociranje sadržaja, omogućavaju automatizaciju prigodom upravljanja njime. Opisana je njihova izvedba, podjela i mogućnosti. Zasebno su obrađeni pojmovi strukturiranosti i interoperabilnosti s gledišta informacijskih sustava. Strukturiranost je opisana kroz implementaciju strukturnih elemenata i postignutim vezama među njima. Interoperabilnost se opisuje kroz tehnike postizanja interoperabilnoga sadržaja više sustava, što se odnosi na nadvladavanje problematike vezane uz razmjenu elemenata strukture više sustava (sintektička interoperabilnost) i zadržavanje njihovih kontekstualnih vrijednosti (semantička interoperabilnost). Kao primjeri strukturiranosti mrežnoga enciklopedičkog sadržaja navode se i opisuju Wikipedija, Hrvatska tehnička enciklopedija, Encyclopaedia Britannica i Proleksis enciklopedija, dok je interoperabilnost enciklopedičkoga sadržaja postignuta putem projekata DBpedije i Wikidate, koje su vezane uz projekt Wikipedije, te prilagodbe sadržaja Gran Encyclopedie Aragonese.

Četvrto, peto i šesto poglavlje donose rezultate izvornih istraživanja ove disertacije te se odnose na dokazivanje ili opovrgavanje hipoteza. Četvrto poglavlje donosi opsežnu analizu ukupno deset mrežnih enciklopedičkih izdanja. Uzorak čini ukupno 455 enciklopedičkih članaka podijeljenih u tri kategorije. Također, prikazani su rezultati kvantitativne i kvalitativne analize sadržaja uzorka svakoga zasebnoga izdanja, Hrvatske tehničke enciklopedije, Hrvatske

enciklopedije, Proleksis enciklopedije, Istarske enciklopedije, Hrvatskog biografskog leksikona, Encyclopaedije Britannice, Der Brockhaus mulitmedial 2008 premiuma te hrvatske, engleske i njemačke inačice Wikipedije koje su tretirane kao zasebna izdanja. Analiza sadržaja obuhvaćala je klasifikaciju naziva u četiri osnovne kategorije naziva koje se pojavljuju u enciklopedičkim tekstovima iz tehničkoga područja: osobe, organizacije, lokacije i vremenska obilježja. Dodjeljivanjem kontekstualnih obilježja svakomu nazivu definirani su unificirani faktografski podatci pogodni za stvaranje strukture.

Peto poglavlje odnosi se na ispitivanje evaluacijskih mjera alata namijenjenih automatskom prepoznavanju nazivlja na enciklopedičkim tekstovima. Uz definiranje osnovnih pojmova poput obrade prirodnog jezika (engl. *Natural Language Processing*, NLP) i automatskog prepoznavanja nazivlja (engl. *Named Entity Recognition*, NER), opisani su alati (aplikacije) korišteni u ovom istraživanju za prepoznavanje naziva u enciklopedičkim tekstovima na hrvatskom i engleskom jeziku. Određen je uzorak tekstova iz više izdanja korištenih i u analizi sadržaja, koji su ručno označeni kako bi se ti rezultati mogli usporediti s izlaznim rezultatima alata nakon obrade tekstova. Evaluacija alata određena je na temelju izračuna standardnih mjera u tu svrhu (preciznost, odziv, točnost i F mjera). Analiza je obuhvaćala ispitivanje evaluacijskih mjera prepoznavanja naziva (entiteta) pojedinih kategorija članaka, uključujući i pojedine enciklopedičke članke, te prepoznavanja pojedinih naziva u cjelokupnom uzorku.

Šesto poglavlje donosi metodologiju uspostavljanja interoperabilnoga mrežnog enciklopedičkog sadržaja iz tehničkoga područja na temelju strukturiranoga sadržaja. Na temelju rezultata analize sadržaja Hrvatske tehničke enciklopedije, Proleksis enciklopedije te Encyclopaedije Britannice pristupljeno je strukturiranju tih izdanja u cilju korištenja njihove strukture za postizanje međusobne interoperabilnosti putem posebno izvedenoga mrežnog modela. Prikazan je shematski prikaz modela i značajke, zatim implemenatcija i funkcionalnosti te na kraju ispitivanje (testiranje) i vrednovanje u smislu interoperabilnosti.

2. EPISTEMOLOŠKE ODREDNICE ENCIKLOPEDIČKOGA SADRŽAJA – ENCIKLOPEDIČKI KONCEPT

Epistemologija kao znanstvena disciplina bavi se genezom i dosezima znanja, odnosno spoznaje. Odgovara na pitanja: Što je znanje? Kako ljudi postižu znanja? Njezina zadaća je ocjenjivanje institucija ili projekata, poput enciklopedija, koji stvaraju i prenose znanja. Informacije se vrednuju prema njihovoj točnosti, objektivnosti, vjerodostojnosti, relevantnosti itd. Upravo su navedena svojstva opisana u radu kao zaštitni znak enciklopedičkih djela – enciklopedički koncept, koji omogućuje pouzdanu uporabu enciklopedija kao izvora informacija. Enciklopedički koncept čini više svojstava na kojima se temelje enciklopedička djela, te se na taj način izdvajaju od ostalih vrsta publikacija. Tradicionalna svojstva (odlike) enciklopedičkih djela su točnost, objektivnost, relevantnost, vjerodostojnost (pouzdanost), sveobuhvatnost, konsolidiranost, sažetost, organiziranost, dok su odlike suvremenih, mrežnih enciklopedičkih projekata i kontinuiranost, kolaborativnost, neograničenost opsega, dohvaćanje (pretraživost) i povezanosti sadržaja, strukturiranost, interoperabilnost.[17]

Enciklopedički koncept odnosi se na uvriježeno poimanje enciklopedičkih djela, odnosno ona zatvorenoga tipa, sa stručnim uredništvom, koja angažiraju stručnjake iz svih promatranih područja. Nasuprot tome stoji najveća enciklopedija svih vremena Wikipedija, koja svojom masovnom kolaboracijom ipak ne može ispuniti uvjete klasičnih, profesionalno izvedenih enciklopedija.

Enciklopedije kao sveobuhvatni projekti posjeduju velik epistemološki potencijal kao kvalitetan, jednostavan i brz izvor informiranja. Ulaze u novo doba svoga razvoja, postaju virtualne, razvijaju se kontinuirano uz pomoć urednika, suradnika i korisnika, te su zasnovane na naprednim rješenjima informacijske i komunikacijske tehnologije. Pod utjecajem digitalizacije i načina prezentacije mijenja se paradigma enciklopedičkoga rada.

[17] Pojedine sastavnice enciklopedičkoga koncepta obrađene su u radovima Jecić, Z.: Enciklopedijski koncept u mrežnom okruženju. Studia Lexicographica, 7(2014) 2(13), str. 99−115.; Smolčić, I., Tolj, J., Jecić, Z.: Epistemological Value of Contemporary Encyclopedic Projects. INFuture2017. Integrating ICT in Society. Zagreb, str. 141−149.

2.1. EPISTEMOLOŠKO GLEDIŠTE ENCIKLOPEDIČKOGA SADRŽAJA

Enciklopedika kao dio informacijskih znanosti dijeli epistemološke probleme informacijske znanosti, u koje spadaju struktura znanosti i znanja, organizacija i prikaz znanja, vremenitost znanja i zastarijevanje znanja (tj. teorije o vremenskoj strukturi znanja), teorije o vrstama i tipovima znanja te načinima stjecanja znanja i dr.[18] Osim navedenoga, znanstveno osnovano djelovanje na izradi enciklopedičkih djela koja pokrivaju pojedine struke i discipline širi enciklopedički djelokrug rada na gotovo sva zamisliva područja ljudske djelatnosti. Tehnološkim razvojem i njegovom implementacijom u sustavu informacijskih znanosti, u najnovije doba dolazi do promjena u ljudskom prihvaćanju novih znanja. Znanje se tretira kao simbolički proizvod te je za njegov rast i razvoj presudna instrumentalizacija simboličke djelatnosti, odnosno podjela na funkcije znanja.[19] Spoznajna funkcija predstavlja djelatnost imenovanja i poimanja predmeta, komunikacijska djelatnost prikaza, diseminacije (razdiobe) i distribucije znanja, informacijska djelatnost pretraživanja i organizacije znanja, a funkcija pamćenja djelatnost pohranjivanja, akumulacije i zaštite zapisa. Znanje se kao simbolički proizvod razvija kao cjelina, a navedene funkcije koje ga tvore uzajamno su vezane.

Instrumentalizacijom znanja kao simboličkog proizvoda nazire se kako enciklopedička djela ne možemo poistovjetiti sa spoznajnom funkcijom koja je predodređena drugim znanostima i znanstvenim disciplinama. Ta hipoteza bi enciklopediku i znanost učinila nespojivima. Prema Vujiću, sustavno (leksikografsko, enciklopedičko) bavljenje znanostima jest i samo znanost.[20] Enciklopedika se bavi promatranim znanostima i njihovim objektima na način vlastita poimanja predmeta istraživanja, u okviru vlastita sustava i plana istraživanja, na temelju procedura vlastite metodologije, na način posebne znanosti kao sintetske znanosti. Enciklopedika uspostavlja vlastita pravila prikaza znanja, osobitu enciklopedijsku epistemologiju. Svojim metodama donosi autentično svjedočanstvo o razvoju i stanju znanosti,

[18] Tuđman, M.: Obavijest i znanje. Zavod za informacijske studije, Zagreb, 1990.

[19] Funkcije znanja dijele se na spoznajnu, komunikacijsku, informacijsku i funkciju pamćenja. Posljedica instrumentalizacije simboličke djelatnosti jest ta da se znanje konstantno mijenja po svojoj strukturi i organizaciji, odnosno zahtijeva stalnu metodologijsku prilagodbu. Više u: Tuđman, M.: Prikazalište znanja. Hrvatska sveučilišna naklada, 2003., str. 15.

[20] Vujić, A.: Acta Lexicographica. Prema znanstvenom utemeljenju enciklopedijske leksikografije. Leksikografski zavod Miroslav Krleža, Zagreb, 2015., str. 88.

koje postiže razvojem i unapređenjem vlastitih enciklopedičkih metoda.[21] Odatle proizlazi njezin zahtjev autonomnosti kao zasebne znanstvene discipline.

Zatvoreni informacijski krug kojim se opisuje postupak prikupljanja podataka, njihove obrade u korisniku razumljiv oblik te njihova uporaba uobičajeno se sastoji od planiranja, prikupljanja i obrade podataka, proizvodnje gotovih obavijesnih (informacijskih) proizvoda (npr. članaka, knjiga) i razdiobe (diseminacije). Enciklopedička djela poistovjećuju se sa svim koracima informacijskoga kruga, kao pomno osmišljena sintetska djela cjelokupne ljudske djelatnosti. Sveobuhvatnost predstavlja sintezu znanja kojim se enciklopedika bavi. U prvom redu odnosi se na najširi izbor tema koje enciklopedije kao djela obuhvaćaju, ali i djelomično na njihovu obradu (sadržaj), koja više potpada pod enciklopedičku relevantnost. Enciklopedička djela u tradicionalnome smislu nerijetko su višesveščana izdanja koja opsegom sadrže tisuće stranica teksta i desetke tisuća jedinica sadržaja. Sveobuhvatnost je u uskoj vezi s opsegom članaka, odnosno dopuštenim brojem redaka (znakova)[22], koji je uglavnom ograničavajući čimbenik tiskanih izdanja. Mrežna izdanja nisu podložna ograničenom opsegu, što postaje njihova prednost, a uredniku dopušta slobodniji pristup pri izradi takvih enciklopedijskih članaka. To ne znači da se mrežna izdanja ne pridržavaju drugih stavki enciklopedičkoga koncepta, ali neke je lakše odrediti i postaviti, poput relevantnosti.

Opća enciklopedička djela, kao što su opća enciklopedija (njem. *Allgemeinenzyklopädie, Universalenzyklopädie;* engl. *general encyclopedia*), opći leksikon ili opći enciklopedijski rječnik obrađuju sva područja ljudskog znanja. Kao takva, odraz su suštine enciklopedičkoga rada koji teži sveobuhvatnom i cjelovitom prikazu znanja, pa se pojam enciklopedija često poistovjećuje s općom enciklopedijom. Cilj univerzalnog pristupa svim temama ravnopravno ipak nije moguće ostvariti, jer se svako od enciklopedičkih djela obraća određenoj publici (tržištu), što je u najmanju ruku datost jezika na kojem je djelo pisano. Tako i opća enciklopedička djela opširnije obrađuju nacionalne teme ili teme područja kojem enciklopedija pripada, što se nužno očituje čak i u internacionalnim izdanjima koja se iz tih djela izvode za potrebe svjetskoga tržišta. Pojedine opće enciklopedije nose nacionalnu

[21] Vujić, A.: Razvitak enciklopedistike i enciklopedijsko vrednovanje. Radovi Leksikografskog zavoda Miroslav Krleža, sv. 1, 1991., str. 25–45.

[22] Jedan redak sadrži 60 mjesta za znakove. Takav način označavanja i bilježenja opsega tekstualnoga sadržaja uvriježen je u tradicionalnoj enciklopedici.

sastavnicu čak i u naslovu (npr. Encyclopaedia Britannica[23]), a neka djela imaju nacionalnu sastavnicu ciljano pojačanu u toj mjeri da istodobno predstavljaju i opće i nacionalne enciklopedije (primjeri Brockhaus Enzyklopädie[24], The World Book Encyclopedia[25], Hrvatska enciklopedija[26], Hrvatski opći leksikon[27]). Enciklopedička djela obrađuju i određene struke ili znanosti, poput strukovnih enciklopedija (engl. *subject encyclopedia*, njem. *Spezialenzyklopädie, Realenzyklopädie*), strukovnih leksikona i strukovnih rječnika, posvećenih jednom području, znanstvenoj disciplini, struci, djelatnosti, fenomenu, predmetu interesa i dr. Općenito uzevši, strukovna enciklopedička djela pružaju detaljnije informacije o nekoj temi nego odgovarajuća djela općega tipa, pa su i namijenjena određenim populacijskim skupinama koje su interesno bliže povezane s dotičnom temom. Primjeri takvih djela su Tehnička enciklopedija[28], Tehnički leksikon[29], Medicinska enciklopedija[30], Pomorska enciklopedija[31], McGraw-Hill encyclopedia of science and technology[32], International Encyclopedia of the Social Sciences[33], Encyclopedia of Nanoscience and Nanotechnology[34], World Scientific Encyclopedia of Nanomedicine and Bioengineering[35] i dr.

Enciklopedije i leksikografska djela postaju važan posrednik u postizanju novih znanja kao tercijarne publikacije koje se sastoje od različitih vrsta kazala i vodiča do primarnih i sekundarnih dokumenata, iako specijalizirana enciklopedijska izdanja, poput stručnih i

[23] Vidi poglavlje 4. 1. 6.

[24] Vidi poglavlje 4. 1. 7.

[25] The World Book Encyclopedia, World Book Inc., Chicago, 1994, 22 sveska.

[26] Vidi poglavlje 4. 1. 2.

[27] Hrvatski opći leksikon : A–Ž. Leksikografski zavod Miroslav Krleža, Zagreb, 1996.

[28] Tehnička enciklopedija. Jugoslavenski leksikografski zavod / Leksikografski zavod Miroslav Krleža, Zagreb, 1963–1997, 13 svezaka.

[29] Tehnički leksikon, Zagreb: Leksikografski zavod Miroslav Krleža, 2007.

[30] Medicinska enciklopedija, I. izdanje. Jugoslavenski leksikografski zavod / Leksikografski zavod Miroslav Krleža, Zagreb, 1957–65, 10 svezaka.

[31] Pomorska enciklopedija, I. izdanje. Jugoslavenski leksikografski zavod / Leksikografski zavod Miroslav Krleža, Zagreb 1954–64, 8 svezaka.

[32] McGraw-Hill Encyclopedia of Science & Technology, XI. izdanje. McGraw-Hill Education, New York, 2012, 20 svezaka.

[33] International Encyclopedia of the Social Sciences, II. izdanje. Gale, Farmington Hills, 2008.

[34] Encyclopaedia of Nanoscience and Nanotechnology, I. izdanje. American Scientific Publishers, Valencia, California, 2004, 10 svezaka.

[35] World Scientific Encyclopedia of Nanomedicine and Bioengineering I. World Scientific Publishing Company, Singapore, 2016, 4 sveska.

znanstvenih, zasigurno predstavljaju sekundarne izvore znanja[36]. Tu tvrdnju pospješuje teza kako enciklopedijska izdanja mogu sadržavati neposredne rezultate znanstvenoistraživačkoga rada, budući da svoj sadržaj članaka katkad crpe iz znanstvenih periodičnih publikacija, knjiga, povijesnih dokumenata i dr.

Prvi je korak k enciklopedičkoj organizaciji znanja podjela tog znanja u pregledne dijelove ili jedinice sadržaja, tj. članke kao osnovne informacijske pakete ili sadržajne jedinice enciklopedičkih djela. Kako bi se traženi dijelovi znanja (članci) mogli pronaći među njihovim mnoštvom, potrebno ih je u daljnjem koraku organizacije znanja označiti, tj. dodijeliti im odabranu riječ ili sintagmu koja opisuje o čemu je riječ u članku. Ta se riječ ili sintagma u enciklopedici naziva naslovom članka ili natuknicom, a u informacijskim znanostima i deskriptorom, opisnicom ili opisnikom. Natuknica se odabire iz niza istoznačnih termina pa postaje normirana riječ ili sintagma. Takav način planiranja rada vodi k izradi abecedarija, popisa natuknica poredanih abecedno, po strukama, prema veličini (opsegu) i dr. Zbog toga enciklopedijska djela krasi visoka organiziranost zapisa.

Prikupljanje podataka u tradicionalnoj enciklopedici odvija se poglavito iz objavljenih radova deklariranih stručnjaka i znanstvenika. Autore većih enciklopedijskih članaka (pregledni članci pojedinih struka, znanstvenih disciplina, biografski članci povijesnih ličnosti i dr.) redovito predstavljaju kompetentni stručnjaci pojedinih područja. Rukopise članaka uređuju stručni urednici, leksikografi (enciklopedičari) ili druge osobe vične enciklopedičkom radu, kao i stručne u području koje obuhvaća temu članka. Uređivanje podrazumijeva oblikovanje rukopisa u enciklopedičke članke, tj. prilagođavanje stila i jezika pisanja enciklopedičkim načelima, svojevrsnu sterilizaciju teksta uklanjanjem suvišnih izraza i fraza svojstvenih osobnom stilu pisanja, prestrukturiranje slijeda iznošenja sadržaja, skraćivanje teksta ili njegovo dopunjavanje dodatnim činjenicama ili aspektima obrade teme, terminološko usklađivanje, umetanje uputnica i sl. Pod uređivanjem članaka podrazumijeva se i verifikacija svih iznesenih podataka.

Posebna enciklopedijska epistemološka vrlina jest diseminacija pouzdanih, razumljivih, sistematiziranih članaka korisnicima, odnosno razdiobi sređenih podataka ponovo u informacijski sustav. Možemo ju nazvati i primarnom zadaćom enciklopedijskih izdanja,

[36] Sadrže preglede i skraćene prikaze radova objavljenih u primarnim publikacijama. Ponekad je sadržaj tih članaka teško razlikovati od sadržaja primarnih publikacija.

budući da enciklopedije prenose postojeća znanja, a rijetko stvaraju nova. Naravno, korištenjem enciklopedičkih djela korisnici mogu doći do novih saznanja. Općenito, sakupljanje mnoštva postojećega znanja na jednom mjestu nerijetko dovodi do novih otkrića. Pri stvaranju novoga znanja, ili propitivanja postojećega, znanstvenici, ali i drugi korisnici informacijskog sustava istražuju i upotrebljavaju objavljeno (poznato) znanje. Sukladno tome javlja se potreba za učinkovitijim sredstvima za indentifikaciju toga, postojećega znanja te povratnim (dohvatni) sustavom. Swanson[37] u vezi s tim govori o dopunjujućem neinteraktivnom objavljenom znanju (engl. *Complementary Noninteractive Literatures*), odnosno o znanju koje nije povezano, ne upućuje jedno na drugo, te djeluje odvojeno. To znanje, sadržano u raznim publikacijama, dopunjuje jedno drugo i njegovom interakcijom (sintezom) moguće je stvaranje novoga znanja. Autor je razvio i softver pod nazivom Arrowsmith (Swanson & Smalheiser, 1997), namijenjen pronalaženju takvih radova.[38] Takav sustav predstavlja i MEDLINE (engl. *Medical Literature Analysis and Retrieval System Online*)[39], bibliografska baza radova iz medicine, farmacije, stomatologije, veterine, biologije, biokemije i dr. Nacionalna knjižnica medicine Sjedinjenih Američkih Država pokrenula ju je 1964. godine kao dohvatni sustav dopunjujućih neinteraktivnih radova. Sadrži više od 26 milijuna zapisa iz više od 5 500 znanstvenih radova objavljenih od 1950. godine.

Mrežne enciklopedije kao svojevrsne baze podataka, uz potrebnu informatičku potporu postaju mjesta umreženoga znanja. Svojom sveobuhvatnošću i strukturom, bibliografskom komponentom, te povezivanjem s drugim informacijskim izvorima, kao sintetska djela imaju potencijal uspostavljanja veza između odvojenih, ali potencijalno međusobno relevantnih pojmova, znanstvenih radova, pa i čitavih znanstvenih disciplina. Moderna mrežna enciklopedijska djela su mjesta interdisciplinarnosti. Tako postavljena u mrežnom okruženju, kao virtualna djela[40], diseminacijom svoje građe otvaraju put k novim spoznajama. Svojstvo diseminacije se ponajprije odnosi na mrežna enciklopedijska izdanja koja svojim epistemološkim vrlinama preuzimaju i mijenjaju problematiku enciklopedijskoga rada, ali i

[37] Swanson, D. R.: Undiscovered public knowledge. Library Quarterly, 56(1986) 2, str. 103–118.

[38] Više u: Swanson, D. R., Smalheiser, N. R.: Undiscovered public knowledge: a ten year update. U: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. Portland, 1996, str. 295–298. Navedeni su primjeri sinteze radova u biomedicini čije su nove spoznaje dovele do kliničkih ispitivanja i pokazale se opravdanima.

[39] https://www.nlm.nih.gov/pubs/factsheets/medline.html (18. 10. 2017.)

[40] Podrobnije objašnjeno u: Jecić, Z., Boras, D., Domijan, D.: Prilog definiranju pojma virtualna enciklopedija. Studia lexicographica, 2(2008) 1(2), str. 115–126.

donose pogodnosti u vidu pretraživanja zapisa, njegove pohrane (očuvanja), ažurnosti i dr. Njihov epistemološki značaj nadilazi tiskana izdanja, koja se digitaliziraju i postaju vidljiva i dostupnija javnosti.

2.2. TRADICIONALNE EPISTEMOLOŠKE ODREDNICE ENCIKLOPEDIČKOGA SADRŽAJA

Enciklopedika se bavi prikupljanjem, očuvanjem, obradom i prezentacijom znanja u obliku dokumenata i publikacija. Također možemo definirati enciklopedije kao djela koja sakupljaju, kondenziraju i organiziraju već postojeće ljudsko znanje, te poput priručnika služe za klasifikacju i deklasifikaciju, ujedno i indeksiranje velike količine znanja. U tim procesima karakteriziraju ih određena svojstva koja u epistemološkom smislu predstavljaju vrline koje u konačnici pridonose jednostavnom, brzom i kvalitetnom prenošenju sintetiziranih znanja određenoga područja. Općenito, suvremena enciklopedička djela odlikuju se točnošću, objektivnošću, relevantnošću, pouzdanošću (vjerodostojnošću), sveobuhvatnošću, konsolidiranošću, sažetošću, organiziranošću, stavkama koje u konačnici čine temeljni ili tradicionalni postav enciklopedike neovisan o vremenskoj komponenti nastajanja enciklopedičkih djela ili utjecaja medija kojim se takav sadržaj prenosi.

2.2.1. TOČNOST

Točnost je jedan od najvažnijih čimbenika enciklopedičkoga rada, te korisniku pruža sigurnost pri korištenju enciklopedija kao izvora informiranja. Prema Jeciću[41], očituje se u omjeru iznesenih brojčanih podataka i faktografskih činjenica. Npr. biografski članci obvezatno sadržavaju godinu, datum i mjesto rođenja i smrti (uz iznimne slučajeve kada je gotovo nemoguće pronaći točne podatke), redovito se navode godine osnivanja tvrtki, razdoblja djelovanja, podatci o broju stanovnika, zaposlenih u nekim djelatnostima, fizikalnim zakonitostima i dr. Pojedini podatci se višestruko provjeravaju, iz više izvora, nerijetko i iz primarnih izvora.

[41] Jecić, Z.: Enciklopedički koncept u mrežnom okruženju. Studia lexicographica, 7(2013) 2(13), str. 102.

Postoji nekoliko dimenzija kvalitete informacija (podataka)[42]: točnost, potpunost, razumljivost, njihov optjecaj. Enciklopedička djela, pogotovo mrežna izdanja, u tom su smislu visokokvalitetna te pozitivno (korisno) utječu na prihvaćanje znanja. Točnost nekog informacijskog izvora proporcionalna je njegovoj pouzdanosti (vjerodostojnosti), odnosno informacijski izvor je pouzdan ako je većina podataka koje sadrži točna. Jasno je kako svaki informacijski izvor, pogotovo sveobuhvatni poput enciklopedije, ne može biti u potpunosti lišen netočnosti, ali svakako može biti visoke pouzdanosti.

Istraživanja kvalitativnog odnosa slobodno i tradicionalno uređenoga enciklopedijskog sadržaja na hrvatskom jeziku između Wikipedije i Hrvatske enciklopedije pokazuju visoku točnost sadržaja, poglavito Hrvatske enciklopedije kao tradicionalne, zatvorenoga tipa. U istraživanju Kubelke i Šoštarića[43] metodologija se sastojala od određivanja kvantitativnoga i sadržajnoga odnosa između izvora, odabira recenzentskoga uzorka i elemenata ocjenjivanja te na temelju toga određivanja opće informativnosti sadržaja, čestoće većih[44] i manjih[45] pogrešaka u izvorima i ocjenjivanja dostatnosti, usmjerenosti i objektivnosti. U korelaciji s ukupnim opsegom objavljenoga sadržaja određena je prosječna pojava pogrešaka u određenom broju članaka. Tablica 1 donosi usporedbu opsega sadržaja dvaju promatranih izdanja.

Tablica 1. Usporedba opsega sadržaja hrvatske Wikipedije i Hrvatske enciklopedije (Kubelka, Šoštarić, 2011)

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

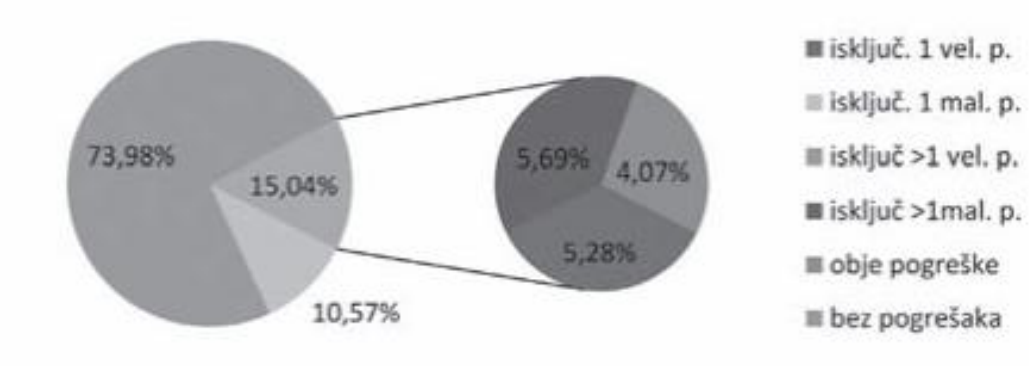
[42] Podatci su određeni simboličkim sustavom i simboličkom djelatnošću, odnosno funkcijama znanja. Razvojem simboličke djelatnosti izgubili smo neposrednost predmeta, izravan put do činjenica, te nam je glavni način spoznaje posredovanjem. Posredovanje se odvija s pomoću knjižnica, arhiva, muzeja, dokumentacijskih centara, bibliografskih, leksikografskih i statističkih zavoda i dr., odnosno informacijske infrastrukture. To nas dovodi do niza posredovanih društvenih odnosa (političkih, ekonomskih, kulturnih, osobnih) o kojima smo i epistemološki ovisni, budući da utječu na točnost, objektivnost, relevantnost podataka i dr.

[43] Kubelka, O., Šoštarić, P.: Wikipedija nasuprot Hrvatskoj enciklopediji, kvalitativan odnos slobodno i tradicionalno uređenoga enciklopedijskoga sadržaja na hrvatskome jeziku. Studia lexicographica, 5(2001) 2(9), str. 119–134.

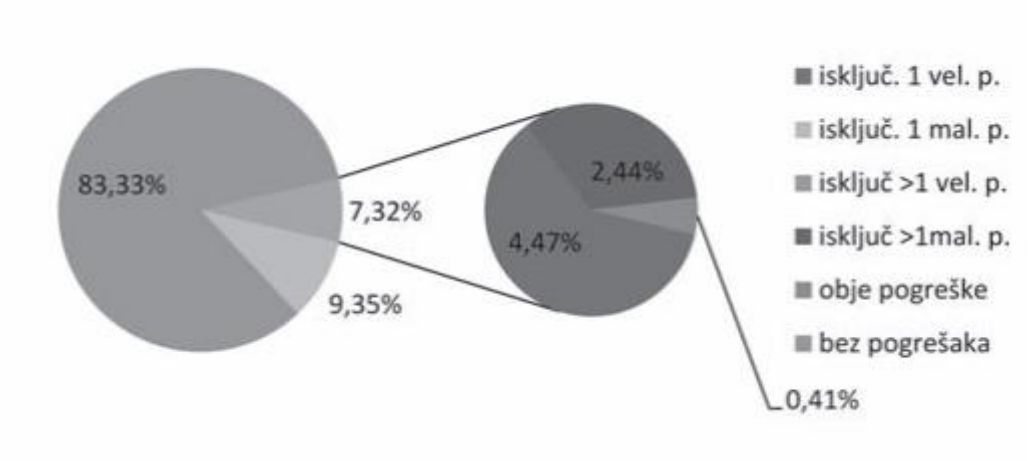
[44] Netočne tvrdnje koje ozbiljno utječu na kontekst članka, znatno utječu na kvalitetu članka i razumijevanje teme koja se obrađuje.

[45] Netočne tvrdnje koje ne umanjuju znatno točnost članaka te nisu ključne u narušavanju kvalitete prenesene informacije i razumijevanja sadržaja u cjelini.

Općenito, nešto više pogrešaka pronađeno je u Wikipediji, što potvrđuje veću pouzdanost tradicionalnoga izdanja. Točnost je proporcionalna pouzdanosti, koja je prikazana slikama dolje.



Slika 1. Udio pogrešaka u uzorku hrvatske Wikipedije (Kubelka, Šoštarić, 2011)



Slika 2. Udio pogešaka u uzorku Hrvatske enciklopedije (Kubelka, Šoštarić, 2011)

Dokazano je kako u nešto više od 83% članaka Hrvatske enciklopedije nije uočena ni jedna pogreška, dok je u 9,35% članaka pronađena manja pogreška. Time je više od 92% sadržaja u potpunosti informativan. Wikipedija na hrvatskom jeziku sadrži oko 85% sadržaja visoke informativnosti, što za projekt otvorenoga tipa predstavlja dobar rezultat.

Istraživanja su proveli i strani autori. Griles[46] je usporedio točnost Wikipedije na engleskom jeziku i izdanja Encyclopaedije Britannice. Članci su preuzeti s mrežnih stranica obaju izvora, njih 42, te poslani stručnjacima na provjeru. Uočeno je samo osam većih pogrešaka, po četiri u svakome izdanju. Pronađeno je više manjih pogrešaka koje ne mijenjaju

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

znatno kvalitetu sadržaja. Wikipedija je sadržavala 162 manje pogreške, a Britannica 123. Uredništvo Britannice osporavalo je istraživanje jer su uočene manjkavosti u metodologiji, poput usporedbe tekstova iz krivih izdanja, razlikovanja većih od manjih pogrešaka i dr.[47] Unatoč tome, istraživanja favoriziraju Britannnicu, od koje Wikipedija sadrži sveukupno 32% više pogrešaka. Britannica sadrži prosječno 3,0 pogrešaka po ispitivanome članku, a Wikipedija 3,9. Pritom standardna devijacija[48] pogrešaka za Britannicu iznosi 2,4, a za Wikipediju 3,5.[49]

Sljedeća istraživanja vezana za enciklopedičku točnost upućuju na kvalitetu Wikipedije u odnosu na druga dostupna mrežna enciklopedička izdanja. Casebourne et al.[50] proveli su opsežno istraživanje kvalitete Wikipedije u odnosu na enciklopedije na engleskom, španjolskom i arapskom jeziku, odnosno mrežne enciklopedičke projekte Encyclopaediju Britannicu, Enciclonet[51] te Mawsoah[52] i Arab Encyclopaedije.[53] Analiza kvalitete članaka, od kojih i točnost informacija, provedena je na člancima iz skupine humanističkih, društvenih, prirodoslovnih i medicinskih znanosti, gdje je općenito zabilježena veća točnost Wikipedijina sadržaja. Međutim, Rector[54] upozorava kako Wikipedija ne predstavlja pouzdan izvor informacija povijesnih članaka, s ukupnom točnošću od 80% u usporedbi s drugim srodnim izvorima, koji uključuju Encyclopaediju Britannicu, Dictionary of American History[55] te American National Biography Online[56], koji ostvaruju točnost sadržaja u rasponu 95 do 96%. Ovako oprečni podatci upućuju na veću neujednačenost kvalitete Wikipedijina sadržaja s gledišta točnosti u odnosu na profesionalno uređeni enciklopedički sadržaj, iako je također dokazano kako kao kolaborativni pothvat predstavlja pouzdan izvor s visokom točnošću

[47] Fatally flawed. Refuting the recent study on encyclopedic accuracy by the journal Nature. Encyclopaedia Britannica, 2006.

[48] Standardna devijacija statistički je pojam koji označava mjeru raspršenosti podataka u skupu, odnosno odstupanje od prosječne vrijednosti (aritmetičke sredine).

[49] Magnus, P. D.: Epistemology and the Wikipedia. North American Computing and Philosophy Conference. New York, 2006.

[50] Casebourne, I. et al.: Assessing the accuracy and quality of Wikipedia entries compared to popular online encyclopaedias: A comparative preliminary study across disciplines in English, Spanish and Arabic. Brighton, 2012.

[51] http://www.enciclonet.com/ (22. 1. 2020.)

[52] http://www.mawsoah.net/ (22. 1. 2020.)

[53] http://www.arab-ency.com (22. 1. 2020.)

[54] Rector, L.: Comparison of Wikipedia and Other Encyclopaedias for Accuracy, Breadth, and Depth in Historical Articles. Reference Services Review, 36(2008) 1, str. 7–22.

[55] Dictionary of American History. Charles Scribners & Sons, New York, 2002, 10 svezaka.

[56] https://www.anb.org/ (22. 1. 2020.)

informacija. Primjerice, Kräenbring et al.[57] ukazuju kako je Wikipedija izrazito točan i sveobuhvatan izvor informacija vezanih uz lijekove koji su predmet preddiplomskih medicinskih studija, s točnošću od 99,7% u odnosu na informacije donesene u specijaliziranim udžbenicima iz područja farmakologije.

Netočne informacije korisnike mogu navesti na prihvaćanje uvjerenja[58] koja ih dovode u epistemološki nepovoljnu poziciju. Epistemološko je mišljenje kako netočnost spada u najozbiljniju epistemološku posljedicu.[59] Pri tome rizici nisu samo epistemološki. Ljudi na temelju takvih informacija mogu donijeti neke odluke koje će primijeniti u praksi. Naravno, činjenica kako neki informacijski izvor, uključujući enciklopedije, sadrži određeni udio netočnih podataka ne mora nužno negativno utjecati na korisnike. Ako korisnici mogu odvojiti točne od netočnih podataka, smanjuje se opasnost od epistemološke posljedice netočnosti.[60] Informacije se mogu učiniti provjerljivima, ponajprije organizacijom (sistematizacijom) istih, odnosno sadržaja publikacije. Osim što će takva organizacija znanja olakšati korisnicima pronalaženje daljnjih dokaza relevantnih promatranoj temi, olakšat će im konzultaciju s drugim izvorima znanja. Urednici mogu održavanjem metapodataka[61] u kontekstu u kojem su informacije stvorene i diseminirane također olakšati korisnicima provjeru navedenih tvrdnji. Time raste i vjerodostojnost takvih izvora znanja.

Tradicionalne enciklopedije svojom organiziranošću dijele znanje na manje, pregledne dijelove. Zasnivaju se na razdiobi nepregledne količine znanja u pregledne dijelove (informacijske pakete, odnosno članke), njihovu označavanju (dodjeljivanju naslova), njihovu nizanju logičkim slijedom, uspostavljanju veza među njima te izradi indeksa. Organizacija svih današnjih mrežnih enciklopedija počiva na hipertekstu, moćnom alatu koji ujedinjuje cjelokupnu građu i čini ju dostupnom i bržom, te predstavlja važan epistemološki čimbenik

[57] Kräenbring J. et al.: Accuracy and Completeness of Drug Information in Wikipedia: A Comparison with Standard Textbooks of Pharmacology. PLoS ONE, 9(2014) 9.

[58] Postavki, tvrdnji ili naučavanja o pojavama, obilježjima pojava i odnosima među njima. Vidi natuknicu: uvjerenje. Hrvatska enciklopedija, mrežno izdanje. Leksikografski zavod Miroslav Krleža, 2020. http://www.enciklopedija.hr/Natuknica.aspx?ID=63540 (4. 5. 2020.)

[59] Fallis, D.: Toward an Epistemology of Wikipedia. Journal of the American Society for Information Science and Technology, 59(2008) 10, str. 1662–1674.

[60] Fallis, D.: On verifying the Accuracy of Information: Philosophical Perspectives. Library Trends, 52(2004) 3, str. 463–487.

[61] Podatci koji pomažu u identifikaciji, opisu i lociranju nekih podataka, napose umreženih elektroničkih izvora. Uključuju kratki opis dokumenta, npr. ključne riječi, što pretražnicima omogućuje lakše pronalaženje toga dokumenta na odgovarajući upit. Vidi poglavlje 3. 1.

znanja postavljenoga na internetu. Faktografski stil pisanja enciklopedičkih djela i razumljivost tekstova dodatno olakšavaju korisnicima verifikaciju tvrdnji. Ako je korisnicima olakšan put do verifikacije podataka njihova interesa, vjerojatnije je kako će do nje i doći. U suprotnome, morat će vjerovati enciklopediji kao izvoru znanja. Ako se na taj način potakne korisnike mrežnih enciklopedičkih izdanja na suradnju s uredništvima, doći će do dodatnog ispravljanja netočnih podataka i povećanja pouzdanosti i vjerodostojnosti enciklopedičkih projekata.

2.2.2. OBJEKTIVNOST

Prema Jeciću[62], pod objektivnošću se podrazumjeva enciklopedički princip na osnovi kojega se pojedine spoznaje i činjenice tretiraju ravnopravno, lišene svih predrasuda na osnovi vjerske, nacionalne, ideološke, rasne ili seksualne pripadnosti, a koje bi se inače mogle očitovati iznošenjem pristranih stajališta i ocjena, neiznošenjem pojedinih relevantnih činjenica, a u ekstremnom slučaju i iznošenjem neistina. Objektivnost predstavlja temeljnu odliku znanstvenoga rada, te brani autoritet znanosti u društvu. Razrađeno je više koncepata znanstvene objektivnosti: objektivnost kao vjernost činjenicama, objektivnost kao nevrijednosni ideal, objektivnost kao suzdržavanje od osobnih predrasuda, sklonosti ili utjecaja.[63]

U svim slučajevima postoje razlozi vjerovanju kako znanost ne može postići potpunu objektivnost. Znanost treba promatrati kao subjektivnu, te propitivati ili sumnjati u njenu objektivnost. Ta subjektivnost nije samo individualna, produkt pojedinca, već ju može uvjetovati skupina ljudi, znanstvenika, koji dijele ista stajališta. Ako neku tvrdnju i smatramo objektivnom, ne postoji metoda s pomoću koje bismo u potpunosti dokazali njenu objektivnost, koja uvijek ostaje upitna. Naravno, enciklopedika pokazuje određen stupanj vezanosti uz društvene uvjete svoga djelovanja. Ti uvjeti mogu biti ekonomski, politički, ali se češće javljaju kroz znanstvene, stručne te dijelom i ideološke preferencije enciklopedista ili leksikografa kao pojedinca. Time se postavlja pitanje postoji li uopće objektivna enciklopedika. Međutim, enciklopedika svakako teži objektivnosti. U suprotnosti logici i eksperimentalizmu koji

[62] Jecić, Z.: Enciklopedički koncept u mrežnom okruženju. Studia lexicographica, 7(2013) 2(13), str. 103.

[63] Reiss, J., Sprenger, J.: Scientific Objectivity. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/sum2016/entries/scientific-objectivity/ (18. 9. 2019.)

zagovaraju znanstvenu objektivnost stoji paradigma, tj. skup znanja i vještina koji u određenom trenutku čine znanost.

Vrijednosti toga skupa nisu neutralne, već obojene nizom uvjerenja i drugih subjektivnih okolnosti što utječu na znanstveni rad. Pojave paradigmi dovode do znanstvenih revolucija[64], odnosno prihvaćanja novih teorija i saznanja koja su uvjetovana pratećom paradigmom. One utječu na metodologije istraživanja i predodžbu rezultata, usmjeruju znanstvenika na postavljanje teza u okvirima paradigme kojoj pripada. Njihova promjena nije temeljena isključivo na argumentima, već i na društvenim uvjetima. Enciklopedijski rad ne može nadići postojeće spoznaje, ali može skrenuti pozornost na njihovu diskutabilnost te potaknuti nove rasprave. Ta odlika predstavlja put k očuvanju enciklopedijske objektivnosti.[65]

Enciklopedika kao znanstvena disciplina metodološki je postavljena da promatra paradigme drugih znanosti koje nastaju i izmjenjuju se tijekom vremena, ne upada u njih, te donosi njihov pregled i kritiku. Naravno, nije u potpunosti lišena društvenoga utjecaja. Enciklopedička objektivnost proporcionalna njenoj sveobuhvatnosti, budući da promatranje i sinteza cjelokupnoga ljudskoga znanja otvara put objektivnosti. Enciklopedisti sakupljaju cjelokupnu građu vezanu uz pojedinu temu, te na temelju njene analize donose sažet, visokoinformativan, neutralan sadržaj članka. Pojedine teme obrađuju se u više natuknica, struka i znanstvenih disciplina, s više gledišta.

I u nedemokratskim, ideološkim vremenima na našim prostorima, enciklopedika je hrabro, svojom objektivnošću, pogotovo specijalističkih izdanja, napravila presedan u domaćoj publicistici.[66] Iako ni ta djela nisu ostala imuna na ideološke implikacije, zahvaljujući enciklopedijskoj metodologiji koja podrazumijeva točnost, vjerodostojnost i objektivnost, poslužila su kao kontrola drugim izvorima informacija te javnim, stručnim i znanstvenim publikacijama.

[64] Kada se nakupi dovoljno dokaza protiv neke teorije, ona zapada u krizu. Pojavom nove, nadmoćnije teorije, dolazi do zamjene paradigmi, odnosno znanstvene revolucije. Više u Kuhn, T. S.: Struktura znanstvenih revolucija. Hrvatsko sociološko društvo, Zagreb, 2003.

[65] Svojstvo objektivnosti upućuje na enciklopediku kao kritičku znanost, koja se definira kao sposobnost znanosti da integrira vrijednosti u dinamičku strukturu razvitka znanosti. Više u: Vujić, A.: Acta Lexicographica. Prema znanstvenom utemeljenju enciklopedijske leksikografije. Leksikografski zavod Miroslav Krleža, Zagreb, 2015., str. 95–99.

[66] Tomović, S.: Ideološke implikacije u enciklopedistici. Radovi Leksikografskoga zavoda „Miroslav Krleža“, sv. 1, 1991., str. 45–51.

2.2.3. RELEVANTNOST

Prema Jeciću[67], relevantnost kao enciklopedički princip podrazumijeva razlučivanje bitnoga od nebitnoga, obradu samo onih pojmova za koje se smatra kako će korisniku pružiti kvalitetnu informaciju o temi koja ga zanima. U informacijskim znanostima, što se odnosi i na enciklopediku, općenito su dvije zastupljene kategorije relevantnosti, ona koja pripada sustavu (engl. *System or Algoritmic relevance*) te ona koja pripada čovjeku (engl. *Human based relevance*, *User based relevance*).[68] Korisnici izvode relevantnost temeljenu na odnosima konteksta i informacija ili informacijskih objekata, tzv. subjektivnu relevantnost.

Sustav stvara relevantnost temeljenu na algoritamskim postupcima, uspoređuje upite koje je postavio korisnik i informacije u sustavu, tzv. objektivnu relevantnost.[69] Čovjek izvlači iz rezultata upita informacijske objekte ili informacije koje su za njih relevantne, te može shvatiti puno više od podudaranja između imenskih fraza, koje sustavi rabe za usporedbu. Tu se stvara novi odnos, između relevantnosti članaka i spoznajne moći pojedinca. Čovjek pronalazi relevantnost iz sadržaja koju nijedan sustav ne može prepoznati, ali ta mogućnost je uvjetovana njegovom stručnošću. Viša stručnost znači djelotvornije pronalaženje relevantnosti. Bez stručnosti pojedinac pronalazi relevantnost kao stroj, jednako kao sustav.[70]

Enciklopedika kao stručna djelatnost i znanstvena disciplina koja se često u širem smislu smatra dijelom leksikografije pristupa ovom problemu na najvišoj razini, budući da za autore svojih članaka redovito pronalazi visokostručne ljude iz područja znanosti, gospodarstva, kulture, sporta i dr. I sama leksikografska obrada tih članaka pridonosi odstranjivanju nerelevantnih tvrdnji kako bi se korisnicima predočio razumljiv proizvod (članak) visoke relevantnosti, neovisan o potrebama i spoznajnim mogućnostima korisnika. Enciklopedička relevantnost podložna je uvjetovanostima i promjenama, ovisi o vrsti djela i skupini korisnika kojima je namijenjena. Poput točnosti i objektivnosti, enciklopedička relevantnost je mjera, te

[67] Jecić, Z.: Enciklopedički koncept u mrežnom okruženju. Studia lexicographica, 7(2013) 2(13), str. 106.

[68] Saracevic, T.: The Notion of Relevance in Information Science. Everybody knows what relevance is. But what is it really?. Morgan & Claypool, 2017. (e - izdanje)

[69] Uvođenje pojma objektivnosti uz relevantnost ipak nije točno. Sustavi za pretraživanje (npr. Google tražilica) proizvod su informatičara, koji stvaranjem algoritama mogu na više načina nekim informacijskim objektima dati prednost pri pretraživanju. Na kraju, oni su ograničeni vlastitim znanjem, djeluju unutar vlastite paradigme.

[70] Više o tretitanju relevantnosti kao ljudske kategorije u: Saracevic, T.: Relevantnost i kako se istraživala. Vjesnik bibliotekara Hrvatske, 50(2007) 1/2, str. 1–26.

su moguća odstupanja unutar enciklopedičkoga principa relevantnosti.[71] Relevantnost ne proizlazi nužno iz historiografskih i znanstveno utemeljenih činjenica, nego je djelomično utemeljena na subjektivnoj procjeni autora ili urednika, koja je najčešće odraz percepcije u javnosti.

Međutim, poput enciklopedičke objektivnosti, stručnost i mogućnost sveobuhvatne obrade određenog pojma donosi enciklopedijama visoku relevantnost. Primjerice, natuknica *Hrvatska* objavljena u općoj Hrvatskoj enciklopediji sadrži, uz definiciju pojma, glavne cjeline: položaj, prirodna obilježja, biljni pokrov, životinjski svijet (fauna), regionalna geografska obilježja, stanovništvo i naselja, narodnosni sastav, vjerski sastav, jezici nacionalnih i jezičnih manjina, gospodarstvo, promet, novac, povijest, politički sastav, nacionalni blagdani, kulturne, znanstvene i druge javne ustanove; uz pripadajuće podnaslove radi još bolje sistematizacije teksta. Ova teorija relevantnosti naišla je i na kritike, uključujući leksikografiju, za koju se navodi kako se temelji na znanju stručnjaka, a zanemaruje potrebe laika koji taj jezik koriste u praksi.[72]

U suprotnosti stoji predmetna (tematska) relevantnost, koja može biti izvedena iz rezultata sustava za dohvaćanje informacija, odnosno provođenja aktivnosti dohvaćanja resursa informacijskog sustava koji su relevantni za informacije iz iste zbirke resursa, ili potpuno neovisna o bilo kojem sustavu[73], izvedena od bilo koje skupine informacijskih objekata. Ovakvo tumačenje izvedeno iz pragmatične teorije o znanju može se poistovjetiti s pozitivističkom epistemologijom.[74] Međutim, podvojena su mišljenja oko ove teze, budući da čovjek može postići tu istu predmetnu relevantnost sofisticiranije, barem za sada, nego samo usporedbom upita i rezultata, kako to obavlja stroj. Mrežna enciklopedijska izdanja, osim kroz relevantnost sadržaja koji donose, treba promatrati i kao svojevrsne baze podataka, koje mogu,

[71] Smolčić, I., Jermen, N., Jecić, Z.: Analiza enciklopedičke relevantnosti na primjeru hrvatske tehničke baštine. Studia Lexicographica, 9(2015) 1(16), str. 133–160.

[72] Hjørland, B.: The foundation of the concept of relevance. Journal of the American Society for Information Science and Technology, 61(2010) 2, str. 217–237.

[73] Vidi prethodnu bilješku. Autor spominje termin *Subject knowledge view of relevance* koji je jednak predmetnoj relevantnosti, te ga smatra temeljnim sustavom relevantnosti.

[74] Sukladno pozitivizmu, pozitivistička epistemologija pretpostavlja kako samo činjenice izvedene iz znanstvenih istraživanja (metoda) mogu proizvesti vjerodostojne tvrdnje. Znanstvenik ne utječe na rezultate istraživanja, već ostaje neutralan.

ako su informatički dobro izvedene, poslužiti kao mjesta upita koji se uspoređuju s njezinom bogatom građom. Time smo došli do polja sustavne, tematske relevantnosti.

2.2.4. POUZDANOST I VJERODOSTOJNOST

Ako enciklopedije predstavljaju oslonac u smislu izvora informacija, bit će pouzdane. To znači kako će pouzdana enciklopedija težiti točnosti, objektivnosti, relevantnosti i drugim epistemološkim značajkama koje će korisnicima omogućiti bezbrižno korištenje takvoga informacijskoga izvora. Ako je pouzdana, tijekom vremena će korisnicima postati i vjerodostojna. Prema Giedyminu[75], općenito postoje dva načina kako odrediti pouzdanost informacijskog izvora. Ako je informacijski izvor pružao pouzdane informacije u prošlosti, postoji veća mogućnost kako ih pruža i u sadašnjosti. Također, s gledišta pouzdanosti, upućuje li nešto kako izvor trenutačo ne pruža točne informacije te postoji li vanjski utjecaj na njega. Pri određivanju vjerodostojnosti informacijskoga izvora, korisnici moraju obratiti pozornost i na stil, sadržaj, metodologiju stvaranja podataka.

Tradicionalno, vjerodostojnosti enciklopedijskih djela pridonosi ugled glavnog urednika pojedinog izdanja, angažiranje stručnjaka iz pojedinih područja kao autora tekstova, a nerijetko i suradnja s institucijama prepoznatima kao autoriteti u znanstvenom i kulturnom području (akademije, instituti). Vjerodostojnosti mrežnih enciklopedijskih izdanja pridonosi interakcija s korisnicima koji svojim komentarima redakciji također mogu utjecati na sadržaj članaka i upozoravati na moguće pogreške i manjkavosti. Zasigurno su tradicionalno uređene mrežne enciklopedije pouzdanije i vjerodostojnije od srodnih informacijskih izvora u mrežnom okruženju. Za primjer treba navesti Wikipediju koja svojom kolektivnom naravi dolazi u sukob sa svojstvima vjerodostojnosti i pouzdanosti. Dodatni problem stvara njen stil, gdje autori na razne načine, npr. postavljanjem mnoštva poveznica, postižu lažni privid vjerodostojnosti ne utječući na sadržaj članaka.

2.2.5. SVEOBUHVATNOST

Po svojim temeljnim postavkama i pozicioniranju u odnosu na druge vrste publikacija, enciklopedije i njima srodna djela poput leksikona ili atlasa spadaju među tercijarne tipove

[75] Giedymin, J.: Reliability of informations. British Journal for the Philosophy of Science, 13(1963) 52, str. 287–302.

publikacija. Prema tome bave se sakupljanjem, verifikacijom i integracijom informacija iz raspoloživih primarnih izvora. Sadrže konsolidirano i akumulirano znanje u vremenu svoga nastajanja. Iako, kad je riječ o specijaliziranim enciklopedičkim izdanjima, poput Tehničke enciklopedije (1963–97)[76] Leksikografskog zavoda Miroslav Krleža ili Stanford encyclopedia of Philosophy[77], ona mogu predstavljati građu izvedenu na znanstvenim temeljima, nerijetko donoseći članke koji predstavljaju znanstvene radove.

Prednosti sveobuhvatnoga enciklopedičkoga sadržaja proizlaze iz sređenoga i razumljivoga uvida u mnogobrojne primarne izvore informacija, što ih čini pogodnima za početak istraživanja neke teme i najprihvatljivijima za većinu krajnjih korisnika. Nedostatci se mogu očitovati, zbog načina rada, u slabijoj ažurnosti zbog ovisnosti o primarnim izvorima te utjecaju autora i urednika na relevantnost informacija. Međutim, mrežna enciklopedička izdanja omogućuju trenutačnu ažurnost i kolaborativnost s korisnicima, što otklajnja te nedostatke. U prilog sveobuhvatnosti mrežnih enciklopedičkih izdanja ide i neograničenost opsega, što se posebno ističe u slučaju Wikipedije, no ponekad na uštrb drugih enciklopedičkih epistemoloških vrlina, poglavito relevantnosti.

2.2.6. KONSOLIDIRANOST

Enciklopedička djela donose sređene ili konsolidirane informacije. Kao i kod konsolidiranja informacija (obavijesti) kao problematike vezane za informacijske znanosti[78] koja se sastoji od niza procesa, rad vezan uz nastajanje enciklopedičkih članaka kao nositelja informacija temelji se na nizu radnji uredništava prije objavljivanja sadržaja.

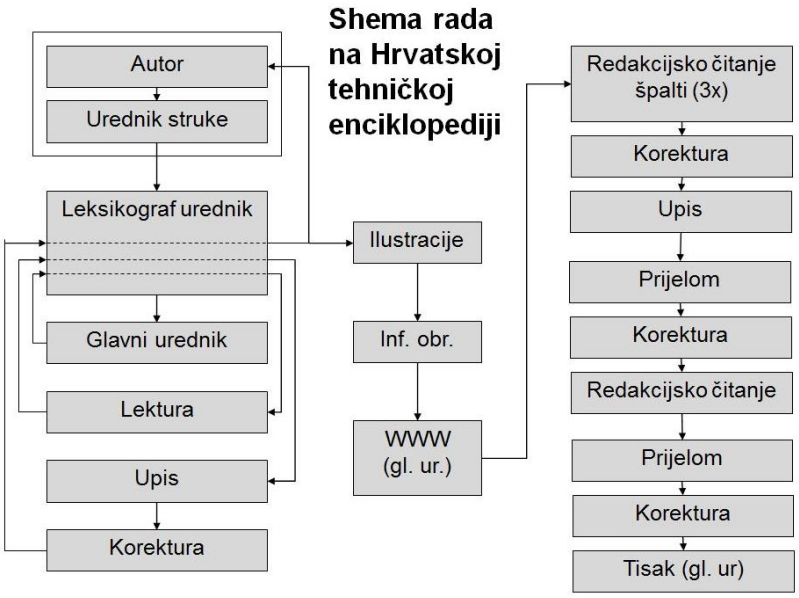
Kako bi enciklopedičko djelo donijelo konsolidiranu informaciju, provode se radnje odabira tematike kojim se pojedino izdanje bavi, odabira relevantnih primarnih izvora informacija, verifikacije izvora sadržaja, sinteze sadržaja više izvora informacija kao prestrukturiranja u cilju djelotvornog korištenja, fizičkog uobličavanja sadržaja u oblik i medij koji će omogućiti njihovo bolje korištenje, jezičnu obradu (lektoriranje) i korekciju sadržaja u cilju otklanjanja grešaka i unapređenja informacijskih paketa, širenja (diseminacije) sadržaja, a kod mrežnih izdanja i dobivanja povratnih informacija od korisnika što omogućuje obavljanje

[76] http://tehnika.lzmk.hr/tehnicka-enciklopedija-1966-1997/ (7. 8. 2019.)

[77] https://plato.stanford.edu/ (7. 8. 2019.)

[78] Vidi: Tuđman, M., Boras, D., Dovedan Z.: Uvod u informacijsku znanost. Zagreb, 1993., str. 49–52.

potrebnih preinaka i prilagodbi sadržaja. Shematskim prikazom (slika 3) prikazani su koraci pri radu na sadržaju Hrvatske tehničke enciklopedije, koji rezultiraju konsolidiranim sadržajem u konačnici.



Slika 3. Shematski prikaz rada na projektu Hrvatske tehničke enciklopedije

Princip konsolidiranosti trajno se provlači kroz enciklopedička djela, poglavito kroz odabir relevantnih informacija (relevantnost), verifikaciju podataka (točnost), visoku faktografsku vrijednost (sažetost) te u slučaju mrežnih enciklopedičih izdanja, diseminaciju putem mrežnih aplikacija.

2.2.7. SAŽETOST

Sažetost je prepoznatljiva odlika tekstova enciklopedičkih članaka, a odnosi se na povećanu faktografsku vrijednost. Enciklopedički sadržaj nastoji u što manje teksta pružiti što više informacija. Ta je karakteristika izraženija kod tradicionalnih (tiskanih) enciklopedičkih izdanja zbog ograničenosti opsega, dok enciklopedička djela koja su ponajprije zamišljena kao mrežna izdanja u nekim slučajevima odstupaju od ovoga pravila.

Sažetost enciklopedičkih izdanja zatvorenoga tipa, koja kontroliraju uredništva i koja se temelje na profesionalnom pristupu, vidljiva je usporedbom rezultata kvantitativne analize sadržaja (prilozi 1, 5, 8 i 9) provedene ovim istraživanjem između Wikipedije na hrvatskom (W_{hr}) i engleskom jeziku (W_{eng}) kao enciklopedičkoga djela otvorenoga tipa te Hrvatske

tehničke enciklopedije (HTE) i Hrvatskoga biografskog leksikona (HBL) kao profesionalno izvedenih enciklopedičkih djela (tablica 2).

Tablica 2. Omjer broja naziva i opsega uzorka dijela mrežnih enciklopedičkih izdanja zastupljenih u istraživanju

Izdanje	Opseg (redaka)[79]	Broj utvrđenih naziva	Omjer broja naziva i opsega
W_{hr}	5021	3669	0,73
W_{eng}	12 267	7998	0,65
HTE	5345	4561	0,85
HBL	1438	2169	1,51

Profesionalno uređeni enciklopedički sadržaj Hrvatske tehničke enciklopedije i Hrvatskoga biografskog leksikona faktografski je bogatiji u odnosu na Wikipediju koja se uređuje slobodno. To se pogotovo vidi u slučaju Hrvatskoga biografskog leksikona koji donosi čak 1,51 naziva[80] po retku teksta. Zbog takvoga pristupa enciklopedički tekstovi uglavnom su lišeni osobnih stilskih obilježja i literalno su šturi. Predstavljaju objektivan izvor informacija, a unatoč novim medijima na kojima se donosi enciklopedički sadržaj i fleksibilnosti u smislu opsega, sažetost je ostala prepoznatljiva enciklopedička vrlina.

2.2.8. ORGANIZIRANOST

Organizacija enciklopedičkoga sadržaja među ključnim je stavkama enciklopedičkoga koncepta budući da omogućava korištenje enciklopedičkoga djela kao opsežnog izvora informacija. Nelinearna organizacija enciklopedičkih djela podrazumijeva fragmentaciju sadržaja na manje, pregledne dijelove, u tradicionalnoj enciklopedici nanizane nekim logičnim slijedom. Uz razdiobu nepregledne količine znanja na pregledne dijelove (jedinice sadržaja ili članke, natuknice), organizaciju enciklopedičkoga sadržaja čini označavanje jedinica sadržaja, odnosno dodjeljivanje naziva člancima (lematizacija), nizanje jedinica sadržaja nekim logičnim

[79] Vidi bilješku 22.

[80] S gledišta obrade prirodnog jezika, naziv (entitet) je točno određena pojavnica ili niz pojavnica u tekstu koje se mogu vezati uz samo jedan konkretni pojam u stvarnosti. Ti su pojmovi imena i nazivi, prirodne pojave, vremenske i brojčane oznake poput datuma i valutnih iznosa te drugih jedničnih oznaka i dr. Ispitivani su nazivi nasjzastupljeniji u enciklopedičkim tekstovima vezanima uz područje tehnike, koji se odnose na osobe, organizacije, lokacije i vremenska obilježja, te su nositelji faktografskoga standarda (informativnosti). Vidi poglavlje 5. 2.

slijedom (uglavnom abecednim), uspostavljanje veza između jedinica sadržaja (umrežavanje članaka dodjeljivanjem poveznica na druge članke) te izrada indeksa kao sustavnog, najčešće abecednim redom sastavljenog popisa imena osoba, naziva, predmeta, tema, mjesta i sl. spominjanih u cjelokupnom sadržaju (tekstu) s ciljem navođenja i omogućavanja njihova pronalaženja ili prepoznavanja iskazanim brojem stranica.

Indeks ili kazalo važna je komponenta organizacijske strukture tiskanih enciklopedičkih djela, a napose onih u kojima pojedini članci obuhvaćaju više pojmova ili naziva. Tako je u enciklopedijskim rječnicima i leksikonima indeks manje važan, dok se npr. u velikim općim enciklopedijama ili enciklopedijskim atlasima indeks javlja kao nužnost, kojom se omogućuje pronalaženje pojedinog pojma ili termina čak i u velikim, preglednim člancima, koji mogu istodobno obrađivati nekoliko pojmova ili čak nekoliko desetaka pojmova. Osim svoje organizacijske funkcije, kazalo ima ulogu i u međusobnoj usporedbi različitih enciklopedičkih djela, pri čemu je broj obrađenih pojmova u kazalu nerijetko dobar pokazatelj opsežnosti djela, za razliku od broja članaka. Tiskana izdanja takav sustav čini teže vidljivom razgranatom linearnom organizacijom znanja.

Organizacija kao temeljna postavka uvriježena prigodom planiranja i izvedbe tradicionalnih i mrežnih enciklopedičkih djela temelj je strukturiranosti modernih enciklopedičkih izdanja koja iz te organizacije proizlazi.

2.3. EPISTEMOLOŠKE ODREDNICE SUVREMENIH MREŽNIH ENCIKLOPEDIČKIH PROJEKATA

Enciklopedije kao sveobuhvatni projekti posjeduju velik epistemološki potencijal kao kvalitetan, jednostavan i brz izvor informiranja. Mijenja se paradigma enciklopedičkoga rada koji ulazi u novo doba svoga razvoja, gdje su enciklopedije virtualne, kontinuirano ih razvijaju urednici, suradnici i korisnici, te su zasnovane na naprednim rješenjima informacijske i komunikacijske tehnologije. Sukladno tome, uz tradicionalne epistemološke enciklopedičke odrednice, enciklopedije kao mrežne projekte pružaju mogućnosti poput kontinuiranosti rada na sadržaju i kolaboratvnosti sa svojim korisnicima, objavljivanja neograničene količine sadržaja uz razvijanje sustava za njegovo dohvaćanje i pretraživanje te u konačnici dinamične prilagodbe sadržaja potrebama korisnika.

2.3.1. KONTINUIRANOST I KOLABORATIVNOST

Podatci navedeni u enciklopedijama zastarijevaju, time i smanjuju epistemološki učinak tih djela. Sama potreba za kontinuiranim praćenjem dosega i promjena nastalih u svim područjima ljudske djelatnosti vidljiva je redovitim objavljivanjem ažuriranih izdanja određenih djela. Encyclopaedia Britannica aktualna je od 1768. godine te je ostvarila ukupno 15 izdanja prije postavljanja u mrežni prostor. Projekt Brockhaus Enzyklopädije započet je 1796. godine, a objavljeno je 21 izdanje. Međutim, tiskane enciklopedije nerijetko su već zastarjele i u trenutku svoga objavljivanja. Vremenski odmak između pojedinih izdanja nekog djela može iznositi niz godina.[81]

Širenje netiskanih enciklopedija omogućeno je razvojem CD-ROM diskova i širenjem uporabe osobnih računala. Kada je 1986. godine Grolierova Academic American Encyclopedia izdana na CD-u pod nazivom Electronic Encyclopedia[82], postala je prvom elektroničkom enciklopedijom distribuiranom na prenosivom mediju. Tiskanu inačicu te enciklopedije sačinjavao je 21 svezak s 10 000 stranica teksta, dok je u CD-ROM inačici baza sadržavala samo tekst, i to više od 9 milijuna riječi i 30 000 članaka. Korištenje enciklopedije omogućavala je programska podrška za dohvat i pretraživanje informacija. Njome je omogućeno pretraživanje naslova članaka i cijelog teksta. Nakon elektroničkih izdanja uglavnom manjih enciklopedija i onih namijenjenih mladima, postupno su se i veće enciklopedije odlučivale okušati u tom području. Nakon višegodišnje neodlučnosti, Britannica je naposljetku objavljena u elektroničkom obliku 1994., isprva kao mrežna, internetska enciklopedija (Britannica online[83]), a tek 1995. našla se na tržištu ta verzija izdana i na CD-u[84].

[81] Primjerice, u hrvatskoj enciklopedici početni svesci Tehničke enciklopedije tiskani su davne 1963. (1. sv.), 1966. (2. sv.) itd., što znači kako od toga razdoblja hrvatska tehnika nije dobila osvježena enciklopedijska izdanja koja se bave isključivo znanjima i dostignućima u području tehnike i tehničkih znanosti. Objavljeno je izdanje Tehničkoga leksikona Leksikografskoga zavoda Miroslav Krleža 2007., koji više djeluje kao informativni priručnik znanja o tehnici. Nasuprot tomu, digitalna izdanja mogu redovito objavljivati naslove, ne moraju čekati dan svoga materijalnoga nastanka u obliku knjige, časopisa, zbornika i dr. Čak su i redovita, osvježena tiskana izdanja teško izvediva, poglavito zbog rastućih troškova redovitog izlaženja.

[82] The Electronic encyclopedia, KnowledgeSet Corporation i Grolier Electronic Publishing, Inc., 1986.

[83] U to doba na adresi http://www.eb.com/eb.htm

[84] Britannica CD, BCD, Version 1.0., Chicago: Encyclopædia Britannica 1994

Prva elektronička verzija njemačke enciklopedije Brockhaus u skraćenom obliku[85] izdana je 1998. godine (Der Brockhaus Multimedial[86]). Godine 2005. izdana je elektronička verzija cjelokupnog 21. tiskanog izdanja enciklopedije Brockhaus[87] naziva Brockhaus Enzyklopädie Digital[88] na USB memorijskoj kartici. Prve usporedbe tiskanih i mrežnih enciklopedija isticale su njihove fizičke razlike, organizaciju podataka, (ne)ograničenost opsega i (ne)promjenjivost sadržaja. Također je povećana ažurnost tih izdanja, izlazila su uglavnom jednom godišnje, a uz navedene prednosti pretraživanja, multimedijalnosti i dr., postignuto je znatno smanjenje troškova izdavanja.

Suvremene mrežne enciklopedije kontinuirani su projekti, izdaju se svakodnevno ažuriranjem postojećih i objavljivanjem novih natuknica, što ih čini relevantnim izvorom znanja u trenutku u kojem djeluju. Nisu namijenjene tisku, već je njihov razvoj usmjeren k izradi i dizajnu mrežnih stranica, mogućnostima pretrage, postavljanju metapodataka i dr. Najveći projekt takva tipa je otvoreni projekt Wikipedije temeljen na masovnoj kolaboraciji njenih korisnika. Slobodni pristup nosi i određene rizike, poput ipak smanjene točnosti u odnosu na profesionalno izvedena djela, također i relevantnosti, objektivnosti i nekih drugih sastavnica enciklopedičkoga koncepta, koje su temelj enciklopedijskoga rada. Unatoč tome, ovakav globalni izvor informacija, razmjerno pouzdan, nudi velike mogućnosti u informiranju.

Stanford Encyclopaedia of Philosophy[89] kontinuirani je projekt, prisutna je stalna suradnja između autora članaka i urednika. Uredništvo koje odabire natuknice i autore istih obavlja i recenziju tih radova, budući da, kao i autori, predstavljaju stručnjake za odabranu temu. Ažurnost je važno obilježje projekta, autori redovito osvježavaju svoje članke, koji su predbilježeni za reviziju svakih tri do pet godina. Prethodni zapisi arhiviraju se kako bi se izbjegli problemi oko citiranja članaka enciklopedije, budući da se sadržaj konstantno mijenja. Projekt ovoga tipa predstavlja podlogu za konsolidaciju i diseminaciju ažuriranoga enciklopedičkoga sadržaja profesionalnoga tipa. Ovakav način organizacije i kooperativnosti znanstvenih krugova širom svijeta iz područja filozofije možemo nazvati znanstvenom

[85] Elektroničko izdanje enciklopedije Der Brockhaus in fünfzehn Bänden, Mannheim: F.A. Brockhaus, 1997

[86] Der Brockhaus Multimedial, Mannheim : F.A. Brockhaus, 1998.

[87] Brockhaus Enzyklopädie, 21. izd., Mannheim : F.A. Brockhaus, 2005.

[88] Brockhaus Enzyklopädie Digital, Mannheim : F.A. Brockhaus, 2005.

[89] Vidi bilješku 77.

enciklopedijom [90] , odnosno enciklopedijom koja izravno donosi rezultate znanstvenoistraživačkoga rada ili sadržaj na razini preglednih znanstvenih radova. Znatna epistemološka nadogradnja ovakvih projekata, mrežno postavljenih kao javni servis bez postojanja naknade za korištenje, proizlazi iz njihove dostupnosti svakome s internetskom vezom.

Mrežno postavljeni projekt Hrvatske enciklopedije[91] temeljen na tiskanom izdanju također dijeli prednosti mrežnoga sadržaja. Uredništvo ju kontinuirano ažurira kako bi se spriječilo njeno zastarijevanje, ali ujedno i raste njen korpus, broj redaka svakodnevno se povećava, kao i broj natuknica. Kolaborativna je, odnosno otvorena svim korisnicima putem upita i komentara uredništvu, koji nerijetko dovode do novih saznanja i podizanja kvalitete sadržaja. Primjerice, broj komentara korisnika 2015. godine iznosio je 822, da bi nakon pada u 2016. godini (435 komentara), u 2019. godini iznosio 1782. On također pokazuje mjeru interaktivnosti između korisnika i uredništva, zanimanje korisnika za pojedine teme i što smatraju relevantnim. Tablica 3 donosi prikaz broja intervencija uredništva (ažuriranja članaka) i dodanih, novih natuknica koji ukazuju na kontinuirani tijek rada i nedovršnost (dugovječnost) projekta.

Tablica 3. Broj ažuriranih i novih objavljenih članaka Hrvatske enciklopedije Leksikografskog zavoda Miroslav Krleža

Godina	Intervencije (broj članaka)	
	Ažurirani članci	Novi članci
2014.	14 438	248
2015.	5900	200
2016.	4193	176
2017.	10 406	205
2018.	3400	174
2019.	2468	207

[90] Pojam znanstvena enciklopedija odnosi se na dio strukovnih ili posebnih enciklopedija. Okupljaju i obrađuju građu jedne znanosti, umjetnosti, područja. Leksikografski zavod Miroslav Krleža objavio je niz strukovnih enciklopedija poput Pomorske (1. izdanje 1954–64., 8 sv.), Medicinske enciklopedije (1. izdanje 1957–65., 8 sv.), Enciklopedije likovnih umjetnosti (1959–66., 4 sv.), Muzičke (1. izdanje 1958–63., 2 sv.), Šumarske (1. izdanje 1959–63., 3 sv.), Tehničke (1963–97., 13 sv.), Poljoprivredne (1967–73., 3 sv.) i Filmske (1986–90., 2 sv.) enciklopedije, Enciklopedije hrvatske umjetnosti (1995–96., 2 sv.) te personalne Krležijane (1993–99., 3 sv.), posvećene djelu Miroslava Krleže.

[91] https://www.enciklopedija.hr/ (8. 8. 2019.); vidi poglavlje 4. 1. 2.

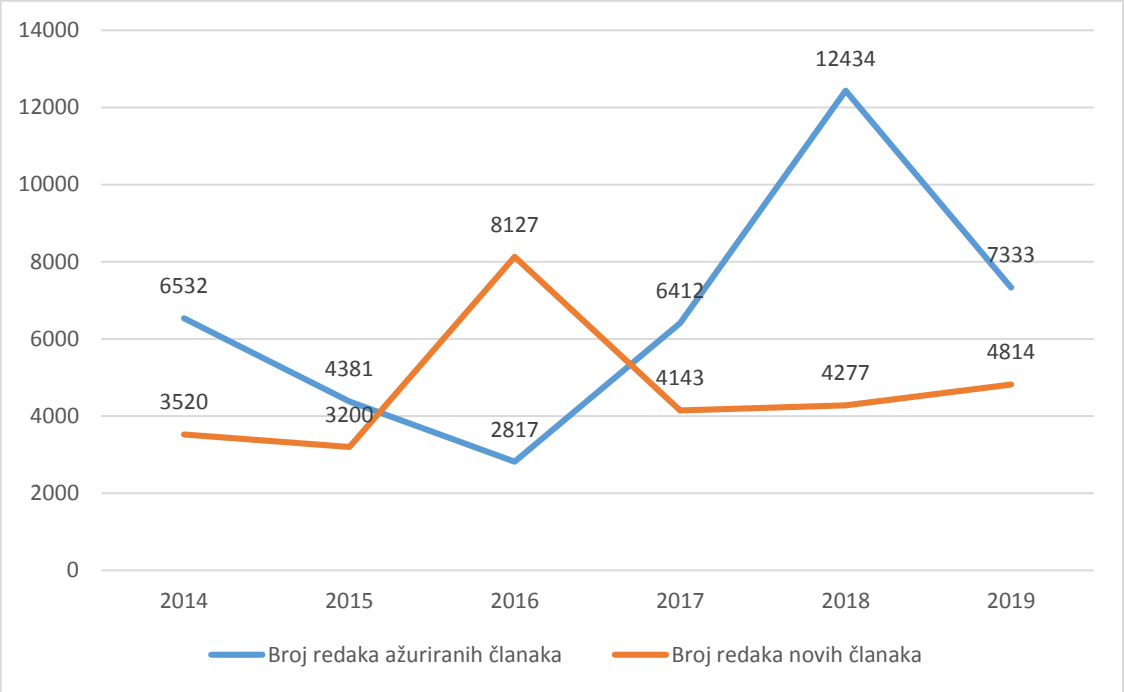
Znatno veći broj intervencija zabilježen je na već postojećim člancima u odnosu na novoobjavljene. To je pokazatelj sve snažnijega trenda održavanja i ažuriranja postojećeg sadržaja uz prisutno dodavanje novih tema projekta Hrvatske enciklopedije.

Hrvatska tehnička enciklopedija[92] također je mrežni projekt Leksikografskog zavoda Miroslav Krleža. Organizirana je kao dinamičan projekt u kojem se rad odvija paralelno, od konstantnih zahvata na popisu natuknica (abecedariju) do pronalaženja autora istih. Ujedno je i prvi projekt hrvatske enciklopedike koji se paralelno objavljuje i kao tiskano i kao mrežno izdanje. Unatoč tome, predstavlja poglavito mrežni enciklopedički projekt koji redovito objavljuje nove natuknice uz ažuriranje postojećih.

2.3.2. NEOGRANIČENOST OPSEGA

Neograničenost opsega mrežnih enciklopedija omogućuje njihovo širenje, dodavanje novih natuknica po potrebi te dopunjavanje postojećih relevantnim informacijama. Tiskana izdanja uglavnom se unaprijed planiraju, odnosno određuje se opseg članaka, pa i pojedinih svezaka. Virtualnom izvedbom opseg postaje stavka koja prestaje biti presudna za organizaciju rada. To ne znači da mrežna enciklopedijska izdanja gube sažetost, odn. visoku informativnost sadržaja, ali podložna su određenom odmaku od stroge leksikografske forme. Kontinuiranost i dinamičnost suvremenih enciklopedija znači i njihov konstantan rast i razvoj, lišen prethodnog ograničavanja opsega. Hrvatska enciklopedija u razdoblju od 2014. do 2019. godine sveukupno je obogaćena za 1210 novih članaka te 67 990 redaka, što u praksi može iznositi i cjelokupni tiskani svezak nekog višesveščanog izdanja (slika 4). Vidljivo je kako se broj novih objavljenih članaka smanjuje, a ukupan broj redaka raste, što upućuje na fleksibilnost mrežnih izdanja, odn. neograničeni prostor i mogućnosti oblikovanja sadržaja koje enciklopedičaru stoje na raspolaganju.

[92] http://tehnika.lzmk.hr/ (8. 8. 2019.); vidi poglavlje 4. 1. 1.

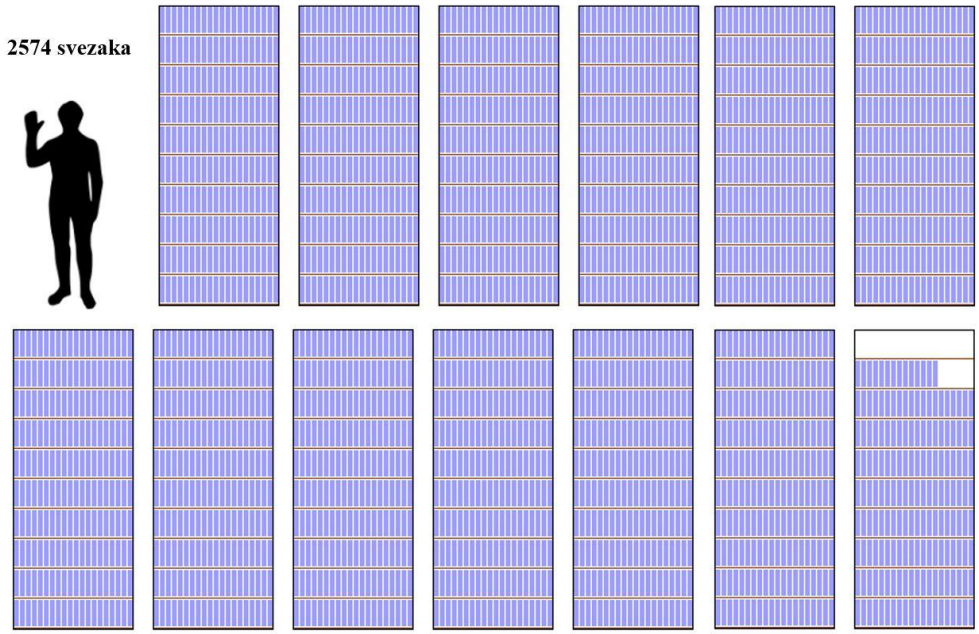


Slika 4. Rast opsega Hrvatske enciklopedije za razdoblje 2014–19.

Primjer neograničene enciklopedije je Wikipedija. Trenutačno su prisutne 294 aktivne inačice Wikipedije na različitim jezicima. Daleko najveća je ona na engleskom jeziku s 5 926 367 članaka, dok hrvatska Wikipedija sudjeluje s 209 319 natuknica.[93] Obujam engleske Wikipedije dnevno se uveća za čak 606 novih članaka. Ako bi se tiskao samo tekst engleske inačice Wikipedije (bez ilustracija i drugog tipa sadržaja) i uvezao po standardima prisutnih tiskanih enciklopedičkih izdanja[94], bilo bi ju potrebno izvesti u čak 2574 sveska. Kako bi se bolje predočio taj podatak, donesen je grafički prikaz slikom 5.

[93] https://en.wikipedia.org/wiki/Wikipedia:Statistics (10. 9. 2019.)

[94] Kao referentno izdanje uzeta je Encyclopaedija Britannica, koje se format približno odnosi na visinu od 25 cm i širinu od 5 cm, sadrži 1 600 000 riječi ili 8 000 000 znakova.



Slika 5. Grafički prikaz opsega engleske inačice Wikipedije izvedenoga po enciklopedijskim svescima (Izvor: https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia)

2.3.3. DOHVAĆANJE (PRETRAŽIVOST) I POVEZANOST SADRŽAJA

Pretraživost enciklopedija u digitalnom obliku znači pristup cjelokupnom sadržaju brzo i učinkovito. Uz pretraživanje naslova članaka (natuknica), pristupa se pretraživanju cjelokupnoga teksta. Međutim, mogućnosti naprednoga pretraživanja i postavljanja upita ovise o daljnjoj strukturiranosti sadržaja mrežno postavljenog izdanja. Sami tekst postavljen na mrežu zapravo je neiskoristiv budući da je računalu neobradiv. Uz to porastom količine informacija na internetu sve je složenije kvalitetno pretraživanje kojim se dolazi do relevantnih rezultata. Sadržaj je potrebno označiti, odnosno postaviti mrežu metapodataka koje će računalo moći obraditi.

Enciklopedičkim djelom nastoji se obuhvatiti što više informacija, ta sinteza znanja predstavlja temeljnu enciklopedičku odliku. Unatoč tome, radi organizacije sadržaja, to znanje dijeli se po natuknicama, kategorijama, strukama, svescima, odnosno dolazi do fragmentacije. Kako bi se održala ukupnost i međusobna povezanost toga znanja, umrežavanje i oblikovanje smislenih cjelina nužnost je enciklopedičkoga rada. Kod tradicionalnih enciklopedičkih djela provodi se uputnicama postavljenima u tekstu, na kraju članka ili kao zasebni članak, koje usmjeruju korisnika na pregled relevantnih i dopunjujućih članaka vezanih uz područje interesa. Umreženost se ostvaruje i uz pomoć indeksa koji, ako je temeljito izveden, može poslužiti kao sredstvo dohvata informacija koje nisu vidljive iz naslova pojedinih članaka.

Digitalizacijom sadržaj enciklopedija izveden je u obliku hiperteksta, a poveznicama je omogućen izravan dohvat povezanoga sadržaja. Uz povezivanje vlastita sadržaja, omogućeno je povezivanje s drugim mrežnim stranicama, koje svojim sadržajem nude dodatne sadržaje vezane uz promatranu natuknicu. Označavanjem sadržaja postavljanjem metapodataka, odnosno strukturiranjem, omogućeno je njegovo povezivanje na više razina. Primjer takva sustava je DBpedia[95], ontološka baza podataka, semantička slika Wikipedije. Objavljivanjem takvoga strojno obradivoga sadržaja na internetu omogućuje se povezivanje s drugim bazama podataka na semantičkoj razini.[96]

2.3.4. ADAPTIVNOST

Adaptivnost ili prilagodljivost korisniku mrežnih enciklopedičkih projekata rezultat je cjelokupne paradigme prisutne pri njihovu razvoju. I u organizaciji tradicionalnih i mrežnih enciklopedičkih izdanja u cilju prilagodbe potencijalnim korisnicima određuje se područje djelovanja tih projekata[97] te je uvjetovan jezik i stil pisanja te odabir pojmova u cilju razumljivosti.

Adaptivnost mrežnih enciklopedičkih projekata odnosi se na radnje potrebne za nadvladavanje statičnosti sustava u cilju pružanja što kvalitetnije informacije. One podrazumijevaju razvijanje korisničkih sučelja, uz prilagodbu sadržaja i strukture koja je u funkciji tog istog sadržaja.[98]

Korisnička sučelja s izraženom grafikom ključna su interakcija korisnika sa sadržajem. Mrežne stranice modernih mrežnih enciklopedičkih izdanja izraženog su dizajna s prisutnim multimedijskim sadržajem i strukturom koja omogućuje funkcionalnosti poput korištenja izbornika, prečaca do traženoga sadržaja, te drugih elemenata korisničkoga sučelja prilagodljivih uvriježenom ljudskom načinu dohvaćanja sadržaja. Adaptivnost mrežnih

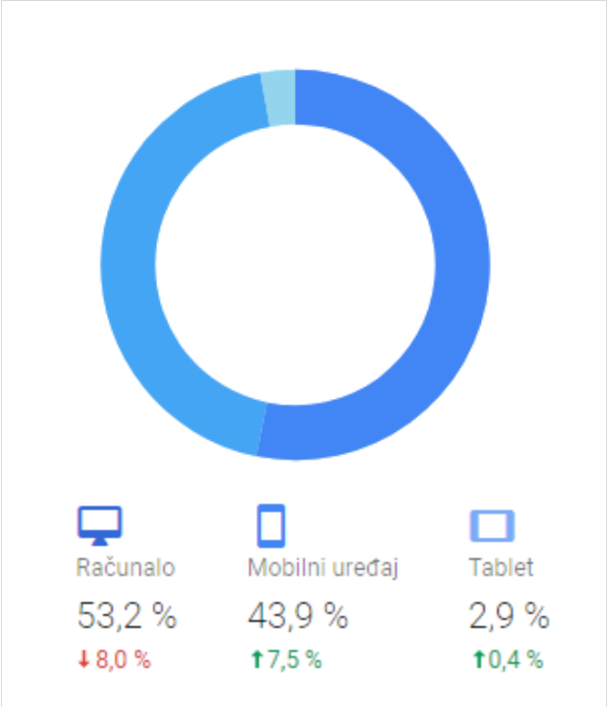
[95] DBpedia crpi metapodatke, odnosno strukturirani sadržaj iz Wikipedijinih članaka u obliku infookvira. Postavljanjem odnosa među objekatima s pripadajućim atributima stvara se isprepletena mreža pojmova koja predstavlja Wikipedijin sadržaj. Pretraživanjem takve baze podataka dobiva se dubinski prikaz cjelokupnoga sadržaja. Više u poglavlju 3. 4. 5. 1.

[96] Takvu vrstu povezanosti predstavlja Linking Open Data project, dostupno na: http://lod-cloud.net/ (8. 3. 2017.)

[97] Npr. opće, specijalizirane ili dječje enciklopedije.

[98] O adaptivnosti enciklopedičkih izdanja i mogućnostima razvoja adaptivnosti više u: Kolbitsch, J., Safran, C., Maurer, H.: Dynamic Adaptation of Content and Structure in Electronic Encyclopaedias. Journal of Digital Information, 8(2007) 3.

enciklopedičkih aplikacija očituje se u njihovoj prilagodbi korisničkim uređajima. Sve češće korištenje enciklopedičkoga sadržaja putem mobilnih uređaja (slika 6) zahtijeva prilagodbu korisničkih sučelja manjim ekranima uz zadržavanje funkcionalnosti i preglednosti, uzimajući u obzir tehničke mogućnosti tih uređaja poput slabijega pristupa mreži i performansi uređaja.



Slika 6. Omjer prisutnosti pojedine vrste uređaja prigodom pristupa mrežnom sadržaju Hrvatske tehničke enciklopedije za razdoblje od 31. 5. 2019. do 28. 8. 2019. s pokazateljima tendencija rasta ili pada u odnosu na istovjetno prethodno razdoblje prikazan aplikacijom Google Analytics

Strukturiranost sadržaja ključna je za njegovu adaptivnost korisniku putem hiperteksta. Uz poboljšanu navigaciju kroz sadržaj i organiziranje jedinica sadržaja u obliku taksonomija, tezaurusa ili ontologija, strukturiranost enciklopedičkoga sadržaja proporcionalna je mogućnostima naprednog dohvaćanja sadržaja postavljanjem izravnih upita nad bazama podataka i postizanju interoperabilnosti s drugim relevantnim izvorima informacija.

3. STRUKTURIRANOST I INTEROPERABILNOST MREŽNIH ENCIKLOPEDIČKIH SADRŽAJA

Suvremena mrežna enciklopedička djela egzistiraju kao informatički sustavi i u tom smislu dijele mogućnosti organizacije sadržaja i razvoja strukture mrežnih stranica te postizanja interoperabilnosti s drugim sustavima koji enciklopedički sadržaj i prezentaciju znanja pronalaze relevantnom.

Ovo poglavlje uz definiranje koncepta strukturiranosti i interoperabilnosti u mrežnome prostoru upućuje na mrežne enciklopedičke projekte koji se ističu u tim poljima i mogu poslužiti kao primjeri u planiranju i razvoju drugih mrežnih enciklopedičkih projekata. Funkcionalnosti sadržaja mrežnih enciklopedičkih djela koje proizlaze iz njihove strukture i razmjena tog sadržaja s drugim sustavima u smislu interoperabilnosti posljedica su kodiranoga sadržaja, odnosno implementacije metapodataka, strojno obradivih podataka koji zastupaju sadržaj namijenjen korisnicima.

3.1. METAPODATCI

Najjednostavnija definicija metapodataka jest da su to podatci o podatcima. Opisuju karakteristike nekog informacijskog sustava u digitalnom obliku, odnosno kao strukturirani podatci informacijskog sustava opisuju, definiraju, lociraju i na druge načine omogućavaju automatizaciju upravljanja resursima. Sukladno tomu, ključni su u stvaranju funkcionalne strukture mrežnih stranica u smislu upravljanja i održavanja takva tipa sadržaja, ali su važni i s gledišta korisnika koji te sustave koriste kao konačne proizvode. Temelj su interoperabilnosti korištenjem definiranih shema (standarda) i komunikacijskih protokola.[99]

Postavljanje metapodataka u informacijskim sustavima naziva se kodiranjem (engl. *encoding*). Postoji više tipova metapodataka koji se koriste u razne svrhe. Dijele se na deskriptivne koji služe za pronalaženje i identificiranje sadržaja, administrativne koji pružaju informacije o uporabi i upravljanju sadržajem, strukturne koji opisuju složenost sadržaja i

[99] O ulozi metapodataka u digitalnim sustavima u: Riley, J.: Understanding metadata: What is metadata, and what is it for?. Baltimore, 2017.

međusobne odnose elemenata strukture, tehničke koji opisuju način podnošenja sadržaja, te metapodatke utkane u sadržaj, npr. označavanje naziva (entiteta) u tekstu ili strukturnih dijelova (odlomci). Tablica 4 donosi pregled tipova metapodataka i njihove uporabe.

Tablica 4. Svojstva i uporaba tipova metapodataka (Riley, 2017)

Tip metapodataka	Svojstva	Uporaba
Deskriptivni	elementi sadržaja poput naslova, autora, spola, datuma objave i sl.	dohvat sadržaja, prikazivanje, interoperabilnost
Administrativni	elementi sadržaja poput veličine i tipa datoteke, autorskih prava i licence	upravljanje sadržajem, očuvanje sadržaja, interoperabilnost
Strukturni	mjesto u hijerarhiji, međusobni odnos među elementima strukture	navigacija kroz sadržaj
Utkani u sadržaju	označavaju kontekst unutar sadržaja	navigacija kroz sadržaj, interoperabilnost

Stvaranje metapodataka predstavlja kodiranje sadržaja i na taj način uspostavljanje strukture razumljive sustavu (računalu). Kodiranje ili označavanje izvodi se izravno s pomoću programskih jezika (engl. *Markup Language*) kao što su XML (engl. *Extensible Markup Language*) ili RDF ili s pomoću razvijenih sučelja za tu namjenu kao dijelova sustava za upravljanje sadržajem (engl. *Content Managment System*, CMS). U cilju pospješavanja procesa stvaranja metapodataka u informacijskim sustavima, razvijeni su algoritmi ili aplikacije namijenjeni što većoj automatizaciji tih procesa. Postoje sustavi za automatsku transkripciju govora, audio i videozapisa, prepoznavanja lica za video i fotografske zapise, za prepoznavanje predmeta na fotografijama i dr. Za tekstualne izvore koristi se semantička analiza kao dio djelomično nadzirajućeg ili poluautomatskog procesa stvaranja metapodataka, gdje se koriste tehnologije označavanja vrsta riječi (engl. *Part of Speech Tagging*, POS Tagging) i automatskog prepoznavanja nazivlja.

Općenito, informacijski sustavi prožeti su metapodatcima kojima se opisuje neki predmet ili sadržaj, te omogućuje interakcija s tim sadržajem. Ugrađeni u mrežne stranice, čine osnovu njihove funkcionalnosti, ključni su u obavljanju digitalnih poslovnih transakcija, na njima počivaju društvene mreže itd. Metapodatcima suvremene knjižnice, arhivi, muzeji i druge ustanove označavaju svoju građu, a razmjenom i dijeljenjem metapodataka omogućuju jednostavniji i učinkovitiji dohvat njihove zajedničke građe, odnosno pretraživanje sadržaja.

Metapodatci se pohranjuju različitim zapisima. Čest način pohranjivanja većine vrsta metapodataka jest u obliku polja u tablicama[100] kao jedinicama relacijskih baza podataka.[101] Učinkovit dizajn u ovom modelu temelji se na odgovarajućoj normalizaciji tablica baze podataka kako bi se povećala učinkovitost pohrane, uravnoteženo s optimizacijom za performanse upita. U ovom se slučaju metapodatci uobičajeno učitavaju u seriju prilagođenim procesima ili se unose ručno uz pomoć namjenski izgrađenih korisničkih sučelja. U 2000-ima pojavio se XML kao uobičajeni mehanizam za kodiranje sadržaja, prijenos i povremeni unutarnji sustav za pohranu metapodataka. Metapodatci u XML-u postoje kao skupovi datoteka, nazvani XML dokumenti. XML definira elemente, odnosno oznake koje označavaju značenje vrijednosti unutar dokumenta. Elementi mogu imati i druge elemente u sebi, a upravo zbog te značajke XML dokumenti dobivaju svoju strukturu. XML dokument je stablo koje započinje s jednim korijenskim elementom.

3.2. STRUKTURIRANOST

Struktura je način na koji je neka cjelina sastavljena od svojih elemenata (dijelova, članova, pojedinosti).[102] Stvaranje strukture nekog sadržaja spada u domenu informacijske arhitekture (engl. *Information Arhitecture*, IA) kao znanstvene discipline. S gledišta informacijskih sustava koji znanje prenose kao proizvod ili uslugu korisnika, obuhvaća niz djelatnosti koje se odnose na usavršavanje tih sustava u cilju što manjih gubitaka i povećanja učinkovitosti prenošenja informacija u interakciji s korisnicima.

Pojam strukture više odgovara suvremenim, mrežnim informacijskim projektima, koja u njihovom slučaju mijenja način planiranja, izrade, te pristupa pri radu i održavanju. Budući da postoje različita shvaćanja pojma organizacije unutar znanstvene zajednice, složen je odnos pojmova organizacije i organizacijske strukture (strukture). Struktura se doživljava kao

[100] Tablica je u informatičkom smislu organizirani skup podataka (engl. *Dataset*) raspoređen u retke (entitete) i stupce (varijable). Pri tome broj redaka je neograničen, a broj stupaca ograničen.

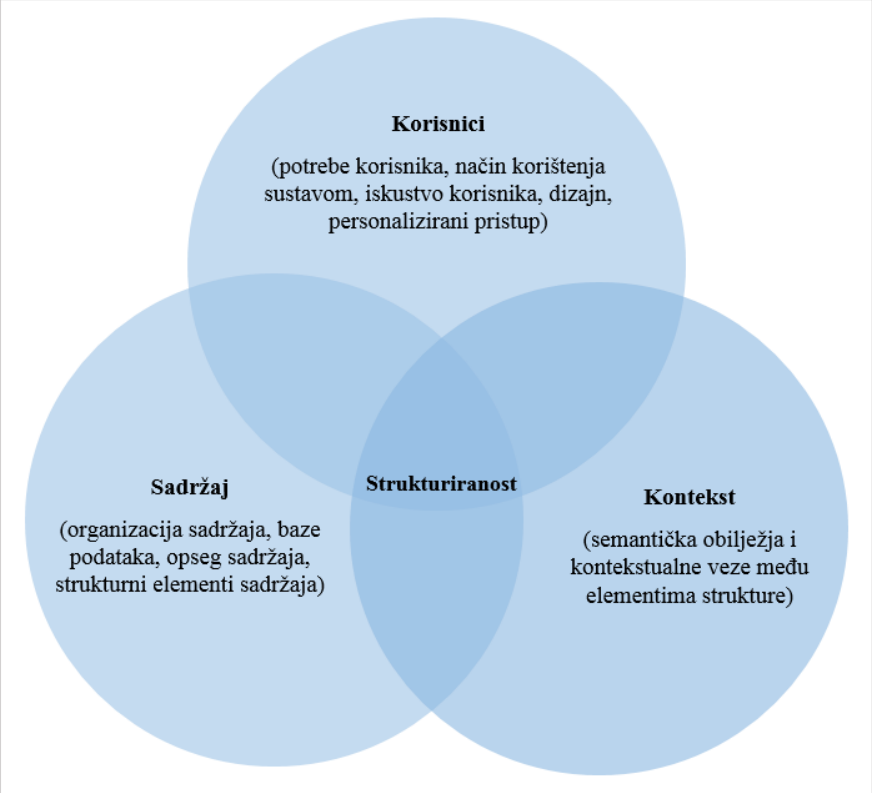
[101] Sustav koji omogućuje funkcionalnosti nad više organiziranih skupova podataka naziva se bazom podataka (engl. *Database*). Relacijske baze podataka koriste se gotovo isključivo u informacijskim sustavima, a sastoje se od niza povezanih tablica ili relacija.

[102] Prema natuknici: struktura. *Hrvatska enciklopedija, mrežno izdanje*. Leksikografski zavod Miroslav Krleža, 2020. http://www.enciklopedija.hr/Natuknica.aspx?ID=58477 (1. 5. 2020.)

obilježje organizacije[103], koje joj daje oblik i omogućuje način njenog funkcioniranja, odnosno uspostavlja organizacijske veze između organizacijskih elemenata, dok je organizacija ciljano uređeno djelovanje.

Podatci egzistiraju u informacijskim sustavima kao strogo strukturirani sadržaj, npr. sustavi koji se koriste bazama podataka, ali i kao nestrukturirani sadržaj, primjerice u obliku objavljenoga prirodnog teksta u mrežnom prostoru. Nestrukturirani sadržaj otežano je održavati, njime upravljati, ali i s korisničke strane dohvaćati sadržaj područja interesa. Strukturiranje sustava odnosi se na izgradnju informacijskih modela koji nadvladavaju razlike između informacijskih sustava i olakšavaju upravljanje informacijama. To se odnosi na organizaciju strukturnih elemenata sustava (metapodataka) i njihovih međusobnih veza te praćenje procesa vezanih uz njihovo nastajanje, upravljanje i korištenje. S korisničke strane važan dio informacijskih sustava predstavlja i dizajn koji omogućuje ugodnije korištenje i utječe na učinkovitost u smislu informiranja korisnika. Kako bi se postigla učinkovita struktura s gledišta korisnika, ali i ljudi koji ju stvaraju i održavaju tako da predstavlja sadržaj, potrebno je informacijski sustav promatrati s gledišta korisnika, sadržaja i konteksta u kojem se taj sadržaj donosi (slika 7).

[103] Vidi: Kapustić, S.: Metodika organizacijskog projektiranja. Samobor, 1984., str. 41−43.



Slika 7. Gledišta postizanja funkcionalne strukturiranosti

Strukturiranost se odnosi na sve aspekte mrežnih stranica, uključujući način na koji je sadržaj organiziran unutar određene stranice, sveobuhvatnost funkcija korisničkog iskustva, organizaciju baza podataka i dr. Strukturu sustava također čine navigacijski elementi, što uključuje izbornike i podizbornike te kontekstualne hiperveze na drugi sadržaj. Na taj način omogućuje se učinkovit i raznolik dohvat sadržaja bez obzira na točku ulaska korisnika u sustav. Strukturirani sadržaj omogućuje i dohvaćanje sadržaja postavljanjem izravnih, složenih upita nad bazom podataka, te je ključan u postizanju interoperabilnosti mrežnoga sadržaja.

3.3. ENCIKLOPEDIČKA STRUKTURIRANOST

Prezentaciju i učinkovito korištenje opsežnog enciklopedičkog sadržaja u tradicionalnom smislu omogućuje njegova organiziranost. Iako se termin organiziranost može poistovjetiti s terminom strukturiranost, on više upućuje na organizaciju sadržaja tradicionalne enciklopedike (poglavlje 2. 2. 8.). Strukturiranost suvremenih mrežnih projekata zasniva se na hipertekstu i metapodatcima utkanima u sadržaj mrežnih enciklopedičkih izdanja, koji u

mrežnom kontekstu služe kao sredstva dohvaćanja sadržaja i navigacijski mehanizmi, ali i kao osnova postizanja interoperabilnosti.

Strukturne značajke vezane uza sadržaj mrežnih enciklopedičkih projekata su hiperveze (poveznice) među jedinicama sadržaja te prema vanjskim izvorima relevantnima za enciklopedički sadržaj. Implementacija elemenata strukture i definiranje njihova odnosa omogućuje provedbu organizacije znanja poput prisutnih primjera u obliku taksonomija (klasifikacija), tezaurusa ili ontologija. Time se pospješuju mogućnosti pretraživanja, pregledavanja i adaptivnosti sadržaja korisniku.

3.3.1. HRVATSKA TEHNIČKA ENCIKLOPEDIJA

Hrvatska tehnička enciklopedija temelj je Portala hrvatske tehničke baštine [104] Leksikografskog zavoda Miroslav Krleža. Organizacija znanja izvedena je po područjima (tablica 5) i kategorijama članaka. Područja su organizirana u tri hijerarhijski organizirane razine, odnosno šira područja koja označavaju sveske tiskanog izdanja, te područja s pripadajućim užim područjima kao nositeljima strukturiranosti sadržaja. Pri odabiru vrijednosti, odnosno oslovljavanju pojedinih elemenata strukture kao sastavnica hijerarhijske organizacije područja koja označavaju jedinice sadržaja Hrvatske tehničke enciklopedije, vodilo se djelomično prema potrebi projekta koji se pretežno bavi poviješću tehnike u Hrvatskoj, ali i prema dijelu Pravilnika o znanstvenim i umjetničkim područjima, poljima i granama vezanima uz područje tehničkih znanosti. Na taj način ususstavljen je jasan skup pojmova kao dijelova strukture koji predstavljaju sadržaj izdanja.

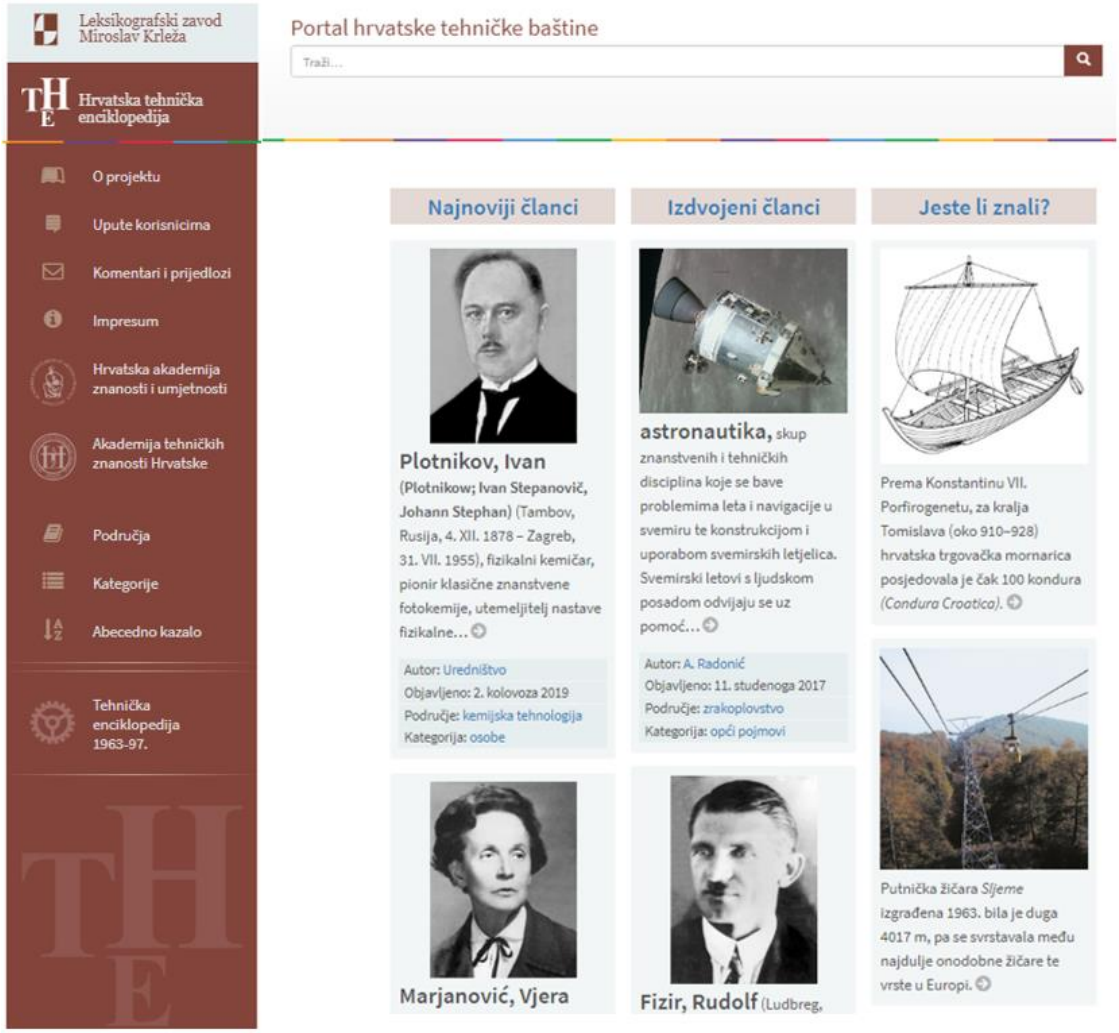
Tablica 5. Raspodjela glavnih područja Hrvatske tehničke enciklopedije po svescima (Jecić, 2018)

Svezak 1	Svezak 2	Svezak 3	Svezak 4
brodogradnja pomorstvo promet strojarstvo zrakoplovstvo	biotehnologija drvna tehnologija grafička tehnologija kemijska tehnologija metalurgija prehrambena tehnologija poljoprivredna tehnika rudarstvo tekstilna tehnologija zaštita okoliša	arhitektura građevinarstvo geodezija	elektrotehnika energetika informatika mjeriteljstvo nuklearna tehnika računarstvo tehnika općenito temeljne tehničke znanosti prirodoslovlje

[104] O projektu Hrvatske tehničke enciklopedije i Portalu hrvatske tehničke baštine više u poglavlju 4. 1. 1.

Natuknicama je pridruženo sedam kategorija koje ih opisuju kao tipove članaka, odnosno kao osobe, poduzeća, udruge, muzeje, časopise, opće pojmove i ustanove, uz pripadajuće potkategorije radi mogućnosti dodatnog strukturiranja. Na taj način metapodatci utkani u sadržaj mrežnog izdanja Hrvatske tehničke enciklopedije, kao i kod primjerice Wikipedijinih kategorija (poglavlje 3. 3. 4.), pospješuju dohvat velike količine sadržaja.

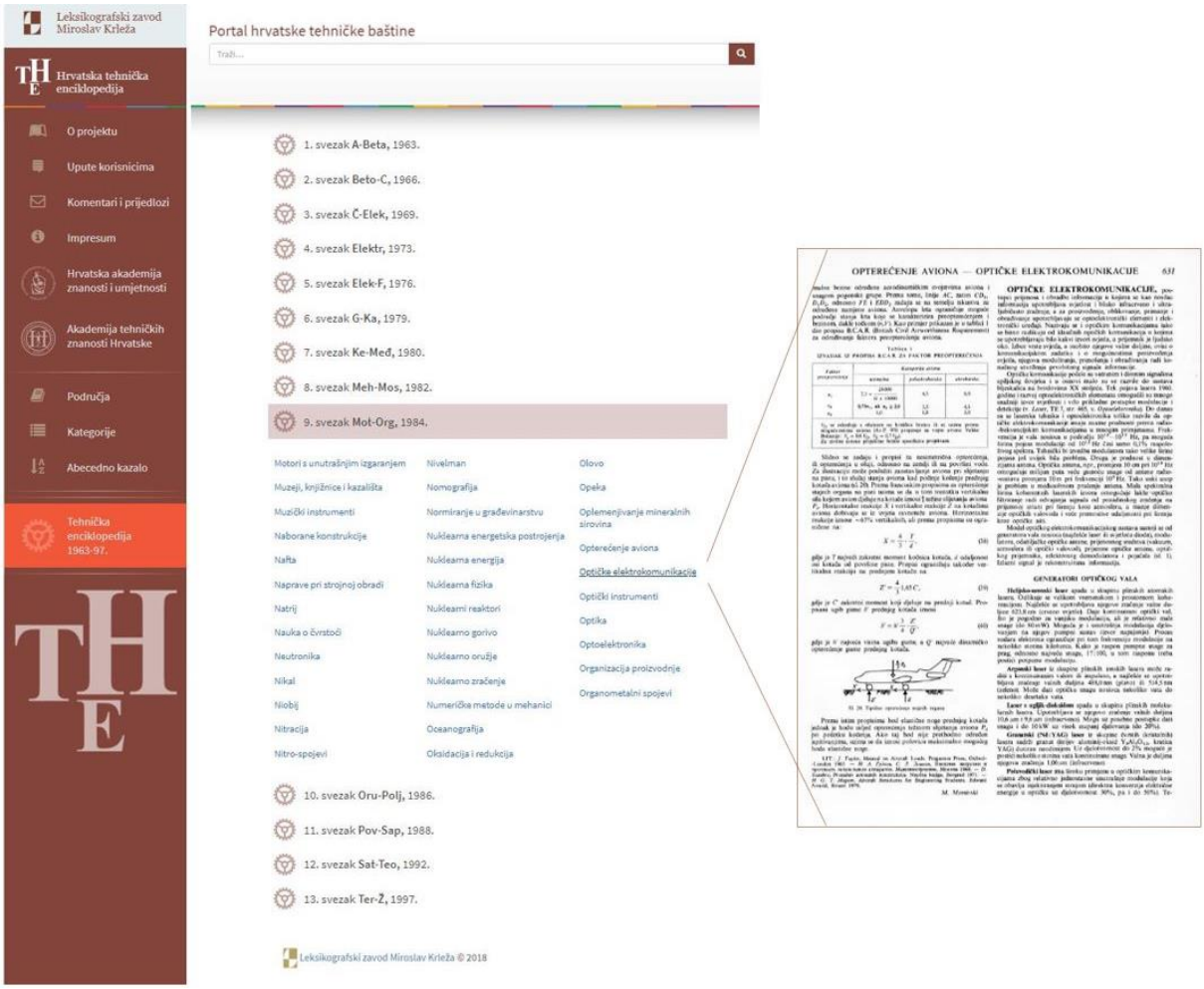
Na temelju tih odrednica sadržaja izvedene su mrežne stranice Portala hrvatske tehničke baštine, a početna stranica donosi prikaz najnovijih objavljenih i istaknutih članaka s posebnom rubrikom *Jeste li znali?* koja svojim zanimljivostima navodi korisnike na temeljni sadržaj Hrvatske tehničke enciklopedije, članke. Početna stranica omogućuje daljnju navigaciju vezanu uz mogućnosti projekta. Slikom 8 prikazana je početna mrežna stranica Hrvatske tehničke enciklopedije kao dijela Portala hrvatske tehničke baštine.



Slika 8. Početna mrežna stranica Hrvatske tehničke enciklopedije i Portala hrvatske tehničke baštine (Izvor: https://tehnika.lzmk.hr/)

Strukturu stranice enciklopedičkoga članka upotpunjuju rubrike koje se odnose na dodatne sadržaje. Hrvatska tehnička enciklopedija omogućuje urednicima dodavanje dodatnih sadržaja vezanih uz članke, kao što je navođenje korisnika na druge članke unutar enciklopedije, a koji su relevantni za promatranu temu (*Vidi još*), popis dodatne literature (*Što pročitati?*), povezivanje s mjestima ili organizacijama koje se može posjetiti, a nude sadržaj vezan uz temu (*Što posjetiti?*), zanimljivosti vezane za temu (*Jeste li znali?*), poveznice na druge sadržaje u mrežnome prostoru (*Mrežne poveznice*) te povezivanje s projektima koje je razvio Leksikografski zavod Miroslav Krleža (*Iz arhive LZMK-a*).

Portal hrvatske tehničke baštine nudi i cjelokupno digitalizirano izdanje 13 svezaka Tehničke enciklopedije, sadržaj koje je raspoređen unutar pdf. datoteka (slika 9).



Slika 9. Prikaz korisničkoga sučelja i digitaliziranoga primjerka Tehničke enciklopedije kao dijela Portala hrvatske tehničke baštine (Izvor: https://tehnika.lzmk.hr/tehnicka-enciklopedija-1966-1997/)

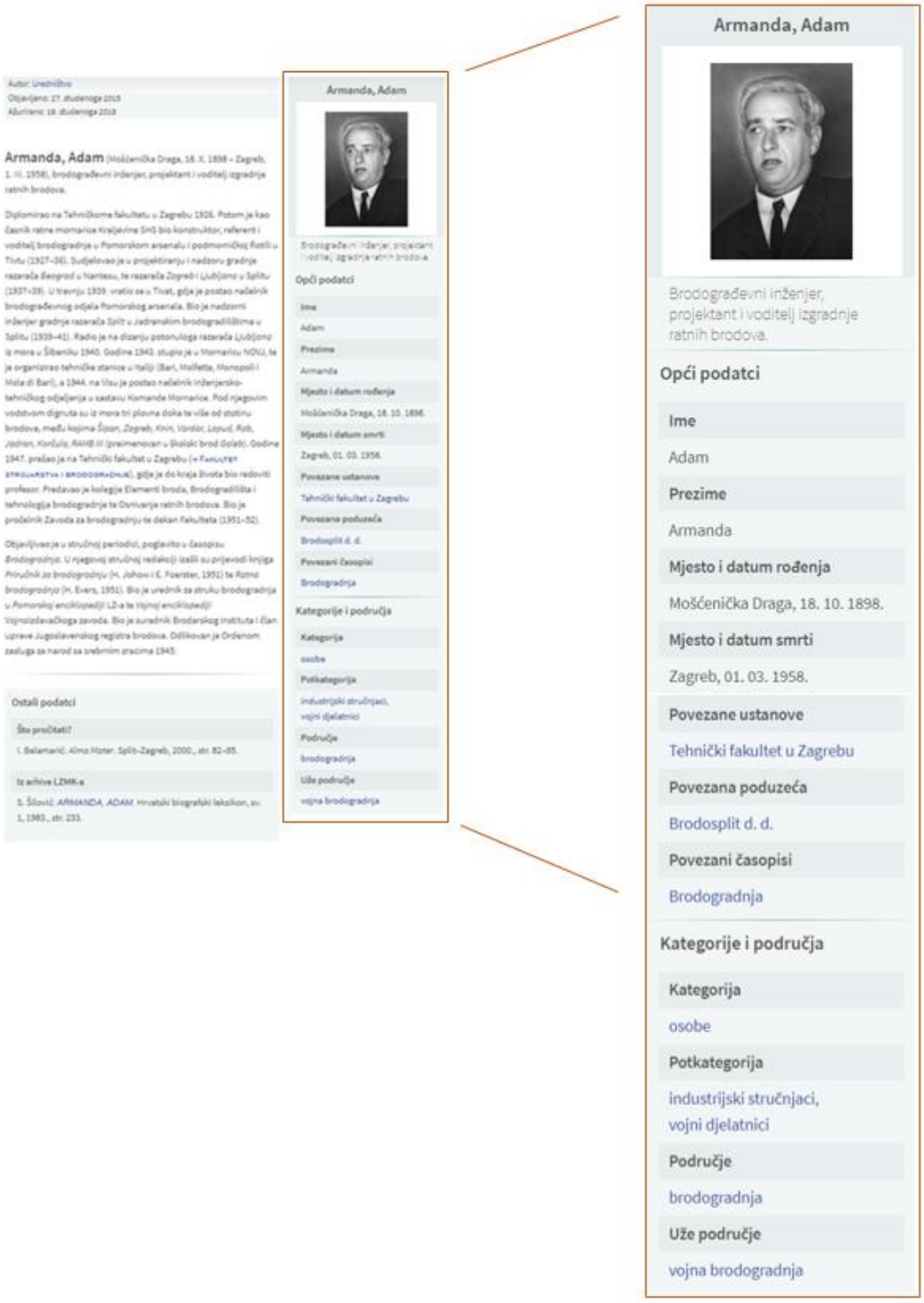
Budući da svaka kategorija članaka donosi određene tipove podataka, sustav za upravljanje sadržajem Hrvatske tehničke enciklopedije omogućuje, s pomoću posebno izvedenoga sučelja, implemenaciju tih podataka u obliku strukture infookvira. Strukturni elementi uključeni u sadržaj izdanja Hrvatske tehničke enciklopedije prilagođeni su potrebama toga izdanja koje se odnose na prezentaciju najvažnijih podataka o obrađivanim temama, međusobno povezivanje sadržaja te mogućnosti naprednog pretraživanja sadržaja na temelju implementirane strukture.

Strukturni elementi koji se pojavljuju u svih sedam kategorija članaka su podatci vezani uz kategoriju i potkategoriju (osim poduzeća) članka, njegovu pripadnost određenom području i užem području kao određenim cjelinama koje obuhvaćaju dijelove cjelokupne građe enciklopedije, te također definiciju ili kvalifikativ vezan uz pojam te noseću ilustraciju koja predstavlja sadržaj članka. To je ujedno i cjelokupna struktura infookvira članaka općih pojmova, odnosno jedinica sadržaja koji se ne mogu pobliže kategorizirati.

Kategorijama članaka osoba i organizacija (poduzeća, udruge, časopisi, ustanove, muzeji) Hrvatske tehničke enciklopedije izvedena su posebna polja strukture (infookvira) pojedinog članka koja ostvaruju međusobnu obostranu vezu s drugim relevantnim jedinicama sadržaja unutar projekta. Ti elementi strukture nazivaju se povezanim osobama, poduzećima i dr. te tvore hipertekstualnu mrežu poveznica relevantnoga sadržaja određene cjeline Hrvatske tehničke enciklopedije, koja nadilazi podatke u tekstu pojedinih članaka i objedinjuje faktografsku vrijednost cjelokupnog tekstualnog (temeljnog) sadržaja određenog područja. Primjerice, članak *Dok-Ing d. o. o.* Hrvatske tehničke enciklopedije donosi podatak o suradnji s *Brodarskim institutom* u Zagrebu, međutim, autor članka *Brodarski institut* taj podatak nije naveo, što je potpuno legitimno s obzirom na opseg članka donošenja najrelevantnijih podataka. Međutim, označavanjem veze između natuknice *Dok-Ing d. o. o.* i *Brodarskog instituta* na temelju jednostranog poznavanja te činjenice ostvarena je obostrana veza između tih pojmova i, u ovom slučaju, nadograđen sadržaj vezan uz problematiku zagrebačkog Brodarskog instituta i njegova djelovanja.

Strukturno su najbogatije izvedeni biografski članci osoba, koji trentačno na raspolaganju imaju 20 funkcionalnih strukturnih elemenata koji se mogu implementirati ako su ti podatci poznati. Kao i kod drugih kategorija članaka, prikazuju se i koriste samo oni strukturni elementi koji su na raspolaganju, dok se ostali ne prikazuju. Od općih podataka moguće je još označiti ime i prezime osobe, inačice imena i prezimena, mjesta i datume rođenja i smrti osoba,

dodijeljene nagrade (poveznice na obrađivane nagrade unutar projekta), te dodati poveznicu na Hrvatsku znanstvenu bibliografiju (Portal CROSBI). Prikaz članka naslova Armanda, Adam Hrvatske tehničke enciklopedije s pripadajućim infookvirom prikazan je slikom 10.



Slika 10. Grafički prikaz članka Armanda, Adam Hrvatske tehničke enciklopedije s pripadajućim infookvirom (Izvor: https://tehnika.lzmk.hr/armanda-adam/)

Također, na slučaju Hrvatske tehničke enciklopedije prikazano je kako struktura izvedena iz bogatoga faktografskog stanja enciklopedičkoga sadržaja može unaprijediti mrežni informacijski sustav poput enciklopedije. Infookviri kao dio Hrvatske tehničke enciklopedije postaju mjesto umrežavanja zasebnih jedinica sadržaja (natuknica) interakcijom istovjetnih elemenata strukture jedinica sadržaja cjelokupnog korpusa Hrvatske tehničke enciklopedije, predmeti formiranja organiziranih skupova podataka, ali i resursi za strojnu obradu podataka. Time je omogućeno obraćanje računalnog sustava bazi podataka, odnosno postavljanja upita nad bazom podataka i dohvaćanje predmeta (u ovom slučaju jedinica sadržaja) iz područja interesa s pomoću sustava za upravljanje bazom podataka. Sve većim strukturiranjem, u ovom slučaju povećanjem broja elemenata strukture infookvira i time baze podataka koja iz njih proizlazi, moguće je postaviti raznovrsnije upite i vrlo precizno dohvaćati sadržaj, bez obzira na enciklopedičku sveobuhvatnost i opsežnost takvih projekata.

Cilj projekta je daljnja digitalizacija drugih izdanja iz tehničkoga područja te rad na sadržaju i razvoju korisničke aplikacije Portala. Portal hrvatske tehničke baštine platforma je za stvaranje, diseminaciju i umrežavanje znanja te interakciju između autora i urednika članaka s jedne strane i korisnika s druge. Trajnim ažuriranjem i dopunjavanjem sadržaja Portal predstavlja suvremeni mrežni enciklopedički projekt i dio je hrvatske znanstvene i stručne infrastrukture.

3.3.2. PROLEKSIS ENCIKLOPEDIJA

Sadržaj Proleksis enciklopedije[105] organiziran je unutar 21 područja, jednorazinski, koja raspoređuju oko 60 000 jedinica sadržaja u smislene cjeline. Ona upućuju na to kako je riječ o općem enciklopedičkom izdanju koje obuhvaća sva područja ljudskog djelovanja. Dijele se na Biološke i medicinske znanosti i područja, Društvene znanosti, Države i narodi, Etnologija, Film, Fizikalno-kemijske i matematičke znanosti, Geografija i srodne znanosti i područja, Geologija, Glazbene umjetnosti i balet, Gospodarske znanosti, Humanističke znanosti i područja, Informacijske i komunikacijske znanosti i područja, Jezične znanosti i područja, Književnost i kazalište, Likovne umjetnosti i arhitektura, Ličnosti, Povijest i povijesne znanosti,

[105] O projektu Proleksis enciklopedije više u poglavlju 4. 1. 3.

Pravo – pravni sustav i državno uređenje, Tehnika i tehničke znanosti, Vojništvo, Šport – igre i rekreacija.

Sadržaj se može dohvaćati i s pomoću ilustracija, a glavna stranica sadrži i rubriku Na današnji dan. Poseban proizvod čini izdvojena cjelina naslova Države svijeta, odnosno zemljovid svijeta koji države povezuje s njihovim natuknicama unutar sadržaja mrežnog izdanja te se uz tekst članka donosi i tablični prikaz s osnovnim informacijama u obliku infookvira (slika 11).

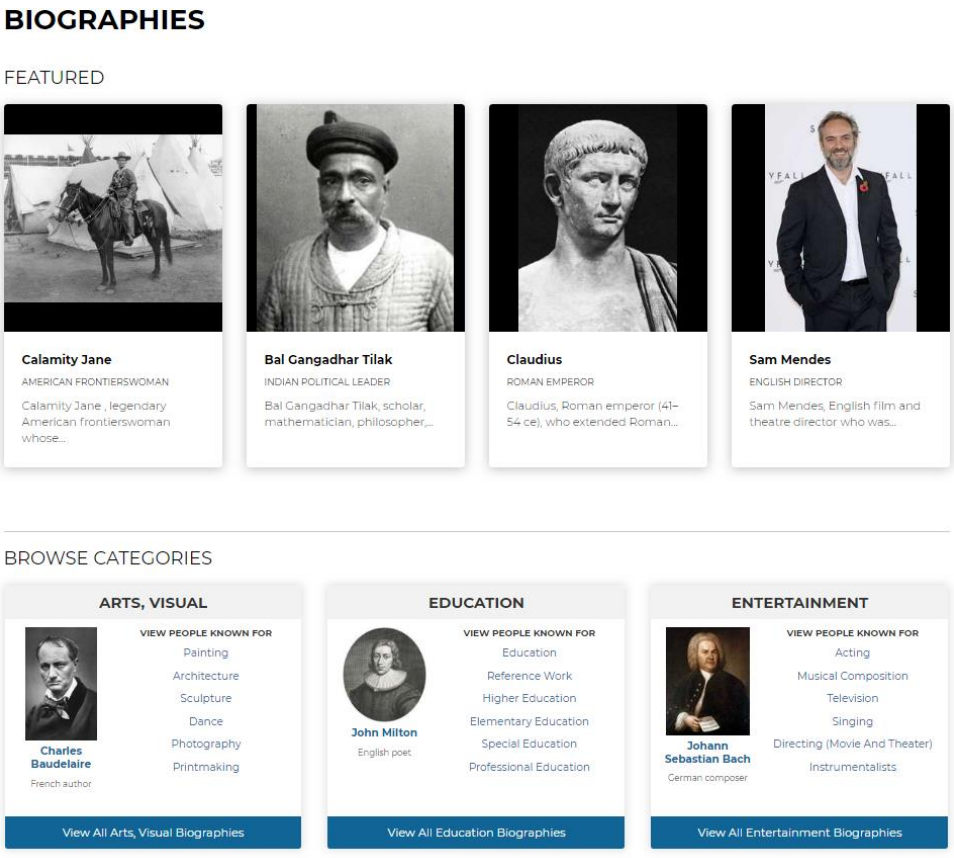


Slika 11. Rubrika Države svijeta Proleksis enciklopedije Leksikografskog zavoda Miroslav Krleža (Izvor: https://proleksis.lzmk.hr/drzave-svijeta/)

Mrežno izdanje Proleksis enciklopedije donosi primjer strukturiranog enciklopedičkog sadržaja koji proizlazi iz tipiziranog enciklopedičkog članka kao što su države. Iako takav strukturirani sadržaj ne spada u područje tehnike, dobar je primjer za to kako iz međusobno relevantnog i unificiranog enciklopedičkoga sadržaja stvoriti strukturu i povećati uporabivost te općenito naglasiti vrijednost sadržaja enciklopedičkih članaka za sva područja enciklopedičkoga rada.

3.3.3. ENCYCLOPAEDIA BRITANNICA

Encyclopaedia Britannica[106] bogato je strukturirano mrežno enciklopedičko izdanje. Organizacija sadržaja je hijerarhijska, odnosno na kategorije i potkategorije koje zastupaju jedinice sadržaja. Encyclopaedia Britannica primjer je profesionalno uređenoga sadržaja s naglašenim dizajnom i vrstama sadržaja. U tom smislu hibridni je proizvod, odnosno enciklopedički sadržaj koristi za izvedbu, npr. kvizova i galerija namijenjenih korisnicima, sadrži posebno strukturiranu stranicu posvećenu osobama (slika 12) i dr. Sadržaj je organiziran u 14 tematskih cjelina, koje čine životinje, umjetnost i kultura, hrana, geografija, zdravlje i medicina, povijest, književnost i jezik, glazba, filozofija i religija, pop kultura, znanost, društvo, sport i rekreacija te tehnologija. Uz spomenute biografske članke, donosi natuknice o organizacijama (ustanove, poduzeća, udruge), državama, gradovima i sl., sve do općih pojmova karaktrističnih za neke teme.



Slika 12. Primjer strukturiranja korisničkoga sučelja namijenjenoga prikazivanju biografskih članaka Encyclopaedije Britannice Online (Izvor: https://www.britannica.com/)

[106] O projektu Encyclopaedije Britannice više u poglavlju 4. 1. 6.

Britannica elemente infookvira koristi kao poveznice na druge jedinice sadržaja unutar svoga korpusa. Ističu se tablični prikazi država, osoba i organizacija kao najbogatijih unificiranim ažuriranim strukturnim elementima, ali i natuknicama općega tipa pridružena je struktura koja ih veže s relevantnim temama unutar projekta Britannice. Pojedine jedinice sadržaja donose elementi strukture infookvira koji ne nadilaze funkcionalnosti tradicionalnih enciklopedičkih izdanja, već djeluju na prezentacijskoj razini. Mrežno izdanje Encyclopaedije Britannice Online slobodnoga pristupa nadilazi granice tradicionalne enciklopedike i upućuje na mogućnosti i potrebu redefiniraja uloge mrežnih enciklopedičkih projekata u globalnom informacijskom sustavu.

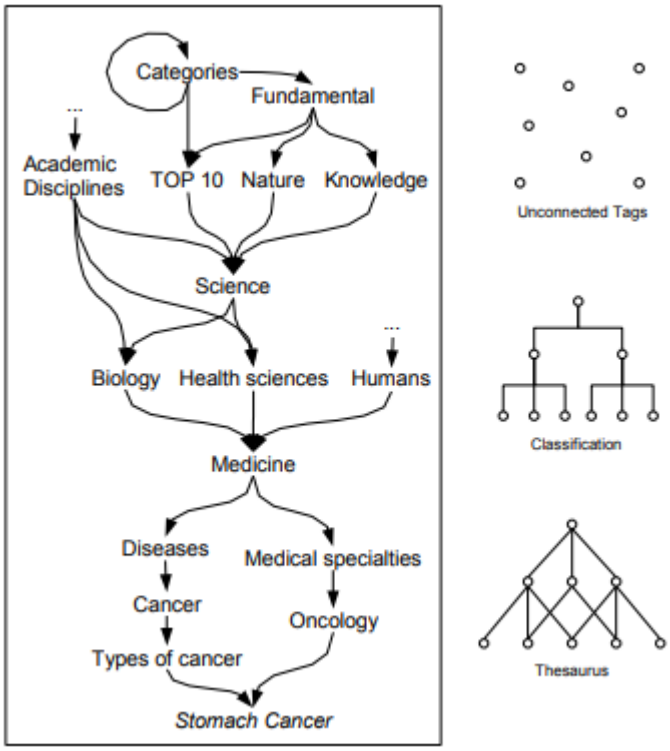
3.3.4. WIKIPEDIJA

Wikipedija[107], otvorena, najveća i enciklopedija koja se najviše koristi primjer je bogato strukturiranog i brzorastućeg mrežnog enciklopedičkog projekta. Strukturno predstavlja višedimenzionalno mrežno enciklopedičko izdanje, budući da kolaborativno označavanje njena sadržaja rezultira organizacijom znanja u obliku tezaurusa.[108] Temelj takve organizacije znanja ili sadržaja čine kategorije kao strukturni elementi i njihovi međusobni odnosi. Pojedinim kategorijama mogu se dodijeliti druge kategorije, što čini hijerarhijski sustav s nadkategorijama i potkategorijama. Većina kategorija povezana je s glavnom nadkategorijom koja je nadređena svima ostalima. Kategorije se prikazuju kao zasebne stranice unutar projekta. Sastoje se od hipertekstualnih prikaza svojih potkategorija i nadkategorija kojima pripadaju, te popisa članaka koji su joj pridruženi. Budući da svaka kategorija može biti dodijeljena mnogim drugim kategorijama, uz postojanje zajedničke nadređene kategorije, organizacija sadržaja Wikipedije nije klasifikacijski sustav, već tezaurus, gdje su kategorije prema definiciji tezaurusa deskriptori koji opisuju sadržaj, odnosno članke enciklopedije.

Sukladno tomu svakoj kategoriji Wikipedije moguće je odrediti šire izraze (nadređene kategorije), uže izraze (potkategorije) te istovjetne izraze (kategorije istorazinskoga značaja), čime skup kategorija Wikipedije čini rječnik ključnih riječi za indeksiranje i dohvaćanje temeljnog enciklopedičkog sadržaja (slika 13).

[107] O projektu Wikipedije više u poglavlju 4. 1. 8.

[108] Više u: Voss, J.: Collaborative thesaurus tagging the Wikipedia way, 2006. https://arxiv.org/abs/cs.IR/0604036 (26. 7. 2018.).



Slika 13. Organizacijska struktura sadržaja Wikipedije na temelju implementiranoga sustava kategorija (Voss, 2006)

Uz sustav kategorija koje navode korisnika do sadržaja Wikipedije, važan su strukturni element Wikipedije infookviri, tablični prikazi najvažnijih podataka vezanih uz svaki članak. Infookviri su izvedeni u formulaciji atribut-vrijednost prikazani u gornjem desnom kutu članaka Wikipedije. Osim za poboljšanje kvalitete i čitljivosti članaka, podatci infookvira koriste se u nekoliko aplikacija izvan Wikipedije, kao što su baza podataka DBpedia ili Googleov grafikon znanja[109] koji izravno prikazuje podatke iz infoofkvira u rezultatima Google pretraživanja. Strukturiranost u obliku infookvira korisna je na više načina. Olakšavaju autorima pisanje članaka, a korisnicima pružaju mogućnosti navigacije i istraživanja drugih članaka korištenjem hiperveza postavljenih u infookvirima do drugih jedinica sadržaja Wikipedije ili vanjskih izvora.

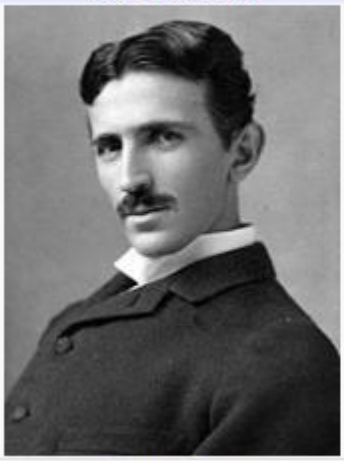
[109] Grafikon znanja (engl. *Knowledge Graph*) predstavlja bazu znanja kojom se sustav Google koristi kako bi poboljšao rezultate pretraživanja s podatcima prikupljenima iz raznih izvora. Podatci se korisnicima prikazuju u infookviru uz rezultate pretraživanja s pomoću aplikacije Google.

Mnoštvu predložaka infookvira koji su razrađeni za velik broj područja i tipova natuknica cilj je strukturiranje što većega broja elemenata strukture iz tekstualnog sadržaja enciklopedije, od kojih neki ne predstavljaju specifične podatke vezane uz određeni tip natuknice. Wikipedijin pristup izradi infookvira predstavlja dobar model u planiranju strukturiranja drugih mrežnih enciklopedičkih projekata na temelju faktografskih podataka, budući da naglašava potrebu razrade predložaka infookvira[110], odnosno odabira strukturnih elemenata za svako obrađivano područje[111] u enciklopediji te za svaki tip enciklopedičkoga članka (osobe, opći pojmovi, organizacije i dr.). Slikom 14 prikazani su primjeri infookvira natuknica Nikola Tesla i Andrija Mohorovičić hrvatske inačice Wikipedije.

[110] Više o planiranju i razradi predložaka infookvira u: Sultana, A. et al.: Infobox suggestion for Wikipedia entities. U: Proceedings of the 21st ACM international conference on Information and knowledge management. Maui, 2012., str. 2307–2310.

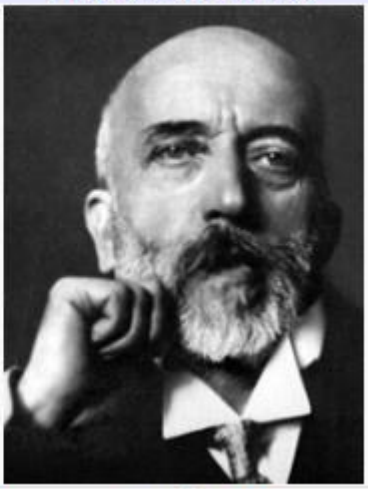
[111] Veća područja za koja su razrađeni i ususta vljeni strukturni elementi infookvira Wikipedije su umjetnost i kultura, geografija, zdravlje i fitness, povijest i događaji, matematika, osobe, religija, znanost i priroda, društvo i društvene znanosti te tehnologija i primijenjene znanosti, koja su dalje hijerarhijski razrađena na uža područja.

Nikola Tesla



Rođenje	10. srpnja 1856.
Mjesto rođenja	Smiljan
Smrt	7. siječnja 1943. New York
Prebivalište	Austro-Ugarska, Francuska, SAD
Državljanstvo	Austro-Ugarska, SAD
Etnicitet	Srbin
Polje	Fizika, elektrotehnika
Institucija	Središnji telegrafski ured u Budimpešti Edisonova telefonska podružnica u Parizu *Tesla Electric Company* u New Yorku
Alma mater	Visoka tehnička škola u Grazu
Poznat po	Izmjenična električna struja Izmjenični električni generator, Indukcijski ili asinkroni elektromotor, Okretno magnetsko polje, Teslin transformator, Tesline struje, Bežični prijenos energije, Radiokomunikacije Teslina turbina
Istaknute nagrade	Medalja Edison (1916.), Zlatna medalja Elliott Cresson (1893.), Medalja John Scott (1934.)

Andrija Mohorovičić



Rođenje	23. siječnja 1857.
Mjesto rođenja	Volosko, Hrvatska
Smrt	18. prosinca 1936. Zagreb, Hrvatska
Državljanstvo	Hrvat
Polje	Seizmologija, geofizika, Meteorologija
Institucija	Pomorska škola Bakar Sveučilište u Zagrebu
Alma mater	Sveučilište u Pragu
Poznat po	Mohorovičićev diskontinuitet Mohorovičićeve epicentrale Mohorovičićev zakon

Slika 14. Primjeri infookvira natuknica Nikola Tesla i Andrija Mohorovičić hrvatske Wikipedije (Izvor: https://hr.wikipedia.org/)

Zbog opsežnog pristupa njihovoj izradi, infookviri se ponekad razlikuju u količini i vrsti uključenih strukturnih elemenata. Međutim, u velikoj mjeri predstavljaju unificirane faktografske podatke pogodne za stvaranje opsežnih organiziranih skupova podataka odnosno baza podataka. Sukladno tomu, okosnica su projekta DBpedije, koji se temelji na ekstrakciji strukturiranoga sadržaja Wikipedije, te ga na taj način čini vidljivim u okviru semantičkoga weba. To znači da ekstrakcijom već strukturiranoga računalno obradivoga sadržaja mrežne enciklopedije nastaje bogata baza metapodataka.

Strukturu Wikipedijinoga sadržaja čine i razne rubrike koje donose podatke o zanimljivostima vezanima uz promatrani pojam, zatim rubrike poput *Vidi još* s poveznicama na relevantan Wikipedijin sadržaj, galerija slika, vanjskih poveznica na mrežna mjesta izvan projekta, popisa bilješki (fusnota) koje se odnose na tekst članaka, popisa literature vezane uz pojam i dr.

3.4. INTEROPERABILNOST

Općenito, interoperabilnost obilježava sposobnost sustava, postrojbi, institucija ili snaga za pružanje i primanje usluga od drugih sustava, postrojbi ili snaga, te uporaba tako razmijenjenih usluga za učinkovito međusobno djelovanje.[112] To upućuje na to kako je taj pojam svojstven za više područja djelatnosti. S tehničkoga gledišta, interoperabilnost je svojstvo koje omogućava neograničeno dijeljenje resursa među više sustava uz minimalni gubitak sadržaja i funkcionalnosti. To se može odnositi na sposobnost dijeljenja podataka među različitim sastavnicama ili uređajima putem softvera i hardvera, među računalima putem lokalnih mreža (engl. *Local Area Network*, LAN) ili širokopojasnih mreža (engl. *Wide Area Network*, WAN).

U računarskom smislu, podatci predstavljaju metapodatke, odnosno strojno obradive podatke koji opisuju sadržaj sustava. Kako bi sustavi postigli interoperabilnost, u smislu precizne razmjene metapodataka u svrhu njihova korištenja kao predstavnika sadržaja, potrebno je zadovoljiti uvjete na sintaktičkoj i semantičkoj razini.

[112] Prema natuknici: interoperabilnost. *Hrvatska enciklopedija, mrežno izdanje.* Leksikografski zavod Miroslav Krleža, 2020. http://www.enciklopedija.hr/Natuknica.aspx?ID=27658 (1. 5. 2020.)

3.4.1. UVJETI POSTIZANJA INTEROPERABILNOSTI

Sintaktička interoperabilnost odnosi se na sposobnost sustava u smislu komunikacije i razmjene podataka. Omogućuje suradnju sustava koji se razlikuju na tehničkoj razini te predstavlja jednostavniji problem. Semantička interoperabilnost odnosi se na razumijevanje podataka među sustavima. Razmijenjene informacije moraju biti kontekstualno smislene kako bi bilo moguća razmjena podataka koje definiraju korisnici sustava.

Preciznije, potrebno je omogućiti interoperabilnost metapodataka na razinama transporta i razmjene (usvajanje komunikacijskog protokola), izlaganja (definiranje strukture vezanja podataka na sintaktičkoj razini), te definiranja metapodataka kao atributa ili varijabli (sheme ili standardi metapodataka) informacijskoga sustava s pridruženim vrijednostima (kontrolirani rječnici metapodataka).

Komunikacijski protokol je skup jednoznačno određenih pravila za razmjenu informacija između dva sustava (programa). Uz pravila, definira sintaktička, semantička i sinkronizacijska obilježja komunikacije (interoperabilnosti). Sintaksa metapodataka odnosi se na tehnički aspekt strukturiranja, odnosno omogućava stvaranje strukturnih polja (metapodataka) prema točno određenim zahtjevima. Kako se jedan metapodatkovni standard može izraziti putem više načina označavanja programskih jezika, od kojih svaki zahtjeva drugačiju sintaksu, potrebno je nadvladati te neusklađenosti u cilju postizanja interoperabilnosti. Semantička interoperabilnost postiže se usvajanjem jedinstvene sheme ili standarda metapodataka te mapiranjem strukturnih polja među sustavima, uz korištenje kontroliranih rječnika koji preciziraju vrijednosti strukturnih elemenata.[113]

Za mrežne projekte poput enciklopedičkih, koji se temelje na strukturi koja predstavlja sadržaj u obliku prirodnoga jezika, izrazito je važno postizanje semantičke interoperabilnosti, odnosno dosljedne i kontekstualno ispravne interpretacije razmijenjenih strukturnih vrijednosti ili podataka. Prije toga potrebno je ostvariti sintaktičku interoperabilnost koja se u kontekstu

[113] Više o svim aspektima interoperabilnosti metapodataka u: Haslhofer, B., Klas, W.: A Survey of Techniques for Achieving Metadata Interoperability. ACM Computing Surveys (CSUR), 42(2007) 2.

jezičnih resursa svodi na sposobnost obrađivanja razmijenjenih strukturnih vrijednosti izravno ili putem pretvorbe.[114]

3.4.2. SHEME (STANDARDI) METAPODATAKA

Kako je potencijalno broj i tip metapodataka (elemenata strukture) nekog sustava neograničen, shema metapodataka na semantičkoj razini definira listu elemenata (atributa, svojstava) koji su dostupni prigodom kodiranja sadržaja uz odgovarajuću sintaksu. Usvajanjem određenoga metapodatkovnoga standarda osigurava se zastupljenost istih strukturnih elemenata u više sustava, što povećava njihovu međusobnu interoperabilnost.[115]

Primjerice, korištenjem XML programskoga jezika u više sustava i zadovoljavanjem tehničkih uvjeta za postizanje interoperabilnosti, svojstvo (atribut) koje pri kodiranju sadržaja označava datum rođenja osobe može biti izvedeno različitim parametrima. Primjeri takvih zapisa mogu biti *<birth_date>2019-09-11</birth_date>*, ali i *<date_of_birth>2019-09-11</date_of_birth>*.[116] Iako je riječ o istom tipu podatka, s gledišta sustava (računala) radi se o različitom značenju sadržaja. Ako sustavi ne koriste iste sheme metapodataka ili ista semantička obilježja elemenata strukture koji ne moraju nužno potpadati pod neki uvriježeni metapodatkovni standard, interoperabilnost se neće moći postići, ili njeni učinci u semantičkom smislu neće biti ispravni. Najzastupljeniji standardi metapodataka su Dublin Core (DC) namijenjen za više domena, Text Encoding Initiative (TEI) namijenjen društvenom i humanističkom području, Sharable Content Object Referebce Model (SCORM) namijenjen području elektroničkog obrazovanja i dr. Pojedine sheme metapodataka također određuju sintaktička obilježja koja uvjetuju način kodiranja elemenata. Većina trenutačnih metapodatkovnih standarda koristi XML jezik za označavanje podataka. Kako bi se omogućio razvoj i održavanje pojedinih shema metapodataka, brigu o njima preuzimaju pojedine organizacije namijenjene standardizaciji metapodataka. Različite sheme metapodataka razvijaju se kao standardi za mnoga područja, poput knjižničarstva, obrazovanja, arhiviranja,

[114] O interoperabilnosti s aspekta jezičnih projekata u: Ide, N., Pustejovsky, J.: What Does Interoperability Mean, Anyway? Toward an Operational Definition of Interoperability for Language Technology. In Proceedings of the Second International Conference on Global Interoperability for Language Resources. Hong Kong, 2010.

[115] Elementi strukture kao sredstva kodiranja sadržaja grupirani su u skupove dizajnirane za određenu namjenu, što rezultira određenom shemom metapodataka ili standardom. Za svaki element navedeni su naziv i semantička obilježja.

[116] Također, sheme metapodataka mogu definirati i formate zapisa. Vidi poglavlje 3. 4. 3.

trgovine, umjetnosti i dr. Tablica 6 donosi pregled odabranih shema metapodataka u području informacijskih znanosti.

Tablica 6. Odabrane sheme metapodataka u informacijskim znanostima

Naziv	Područje	Namjena	Označavanje sadržaja	Standardizacijsko tijelo
Dublin Core (DC)	općenito	označavanje širokog raspona elemenata	XML, RDF	ISO[117] NISO[118]
Data Documentation Initiative (DDI)	arhiviranje, društvene znanosti	označavanje elemenata za opisivanje elemenata u ekonomiji, zdravstvu i sl.	XML	ICPSR[119]
Visual Resources Association Core (VRA Core)	kulturna baština	označavanje radova vezanih uz vizualnu kulturu	XML	VRA[120]
Text Encoding Initiative (TEI)	humanističke i društvene znanosti, lingvistika	označavanje digitalnoga teksta	XML	TEI[121]
Machine-Readable Cataloging (MARC)	knjižničarstvo	označavanje i razmjena bibliografskih podataka	XML	Library of Congress
Sharable Content Object Reference Model (SCORM)	elektroničko učenje	označavanje, sakupljanje i uređivanje obrazovnih objekata	XML	ADL[122]
Resource Description Framework (RDF)	mrežni resursi	modeliranje podataka implementiranih u mrežnim sustavima	XML, RDF	W3C[123]

3.4.3. KONTROLIRANI RJEČNICI

Važni čimbenici u postizanju pune interoperabilnosti metapodataka uz usvajanje određene sheme su i kontrolirani rječnici koji služe kao pravilnici u kodiranju vrijednosti određenih elemenata. Ovaj način usklađivanja vrijednosti metapodataka ključan je za

[117] International Organization for Standardization

[118] National Information Standard Organization

[119] Inter-university Consortium for Political and Social Research

[120] Visual Resources Association

[121] Text Encoding Initiative

[122] Advanced Distributed Learning Initiative

[123] World Wide Web Consortium

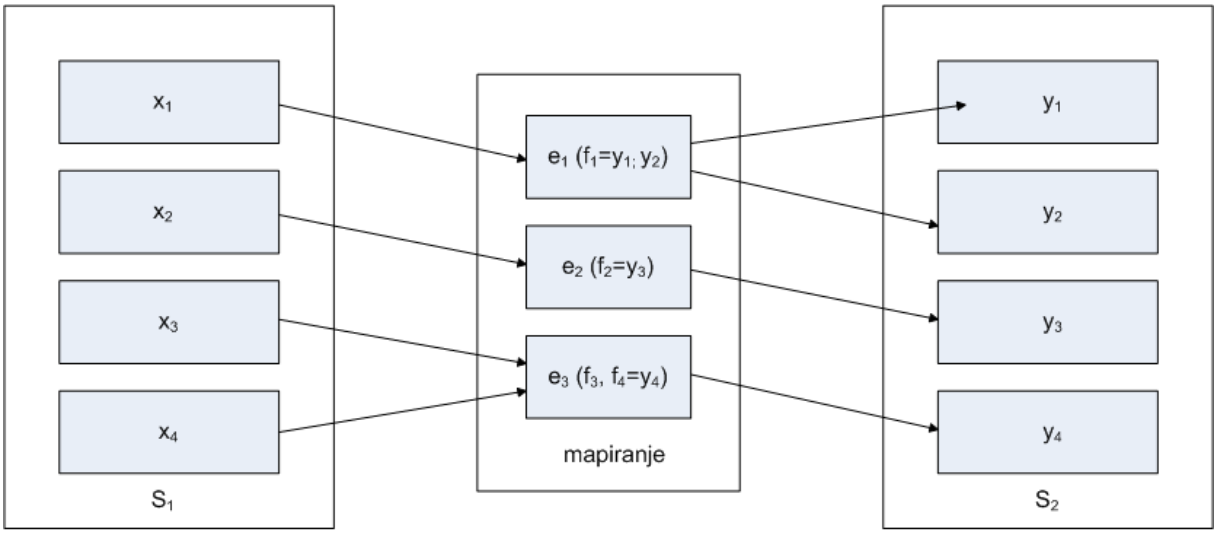
ostvarivanje interoperabilnosti pospješivanjem dohvaćanja i pretraživanja sadržaja, budući da neusklađene vrijednosti atributa rezultiraju smanjenjem odziva pri obavljanju tih radnji i time interoperabilnosti među sustavima. Ovim se pristupom otklanjaju problemi homonimije, sinonimije i polisemije te smanjuju višeznačnosti vrijednosti elemenata strukture.

Ovaj problem može umanjiti i dohvat informacija i zasebnog sustava ako se kodirani sadržaj istoga tipa razlikuje u vrijednostima po kojima ga se dohvaća. Npr. pri isticanju vremenskih obilježja vrijednosti istog elementa strukture kao što je datum rođenja može glasiti *5. 2. 2019., 5. II. 2019., 5. II. 2019*, 5. veljače 2019. ili *February 5, 2019* što sustav koji taj podatak obrađuje onemogućava u prepoznavanju istog tipa podatka. Složenost usvajanja jedinstvenih vrijednosti metapodataka može varirati ovisno o organizaciji strukturnih elemenata, od jednostavnih popisa pojmova, dogovora na hijerarhijskoj razini vezanoj uz taksonomije do sustava koji definiraju pojmove i međusobne semantičke veze (tezaurus, ontologija).

Uvriježeni pravilnici vezani uz definiranje vrijednosti elemenata strukture koji su se također razvijali za određena područja su Library of Congress Subject Headings (LCSH), Dewey Decimal Classification System (DDC), Medical Subject Headings (MeSH) i dr.

3.4.4. MAPIRANJE ELEMENATA STRUKTURE

Interoperabilnost sustava može se postići i mapiranjem metapodataka, odnosno strukturnih polja među sustavima. Mapiranje podrazumijeva definiranje strukturnih i semantičkih veza među elementima strukture više sustava na razini sheme metapodataka, ali i među vrijednostima elemenata na razini instanci (pojedinih podataka). Kako bi se postigla interoperabilnost na ovakav način, potrebno je poznavati sve elemente sustava namijenjenih mapiranju s jasno izraženim sintaktičkim i semantičkim obilježjima. Slikom 15 prikazana je shema mapiranja elemenata strukture dvaju sustava.



Slika 15. Shema mapiranja strukturnih elemenata dvaju sustava

Prikazana su dva sustava S_1 i S_2 s pripadajućim elementima strukture koji su opisani određenim metapodatkovnim standardom (x_n i y_n). Pri tome S_1 predstavlja izvorne podatke, odnosno shemu metapodataka, a S_2 ciljani metapodatkovni standard. Mapiranje elemenata dvaju sustava opisuju svojstva (e) koja su obilježena vezama (f) među povezanim shemama metapodataka, odnosno među zasebnim elementima strukture, koja se odnose na usklađivanje odnosa atributa i njihovih vrijednosti.

Prigodom premošćivanja razlika među shemama metapodataka koriste se razrađeni zapisi koji usklađuju elemene struktura i njihove vrijednosti više shema metapodataka (engl. *Crosswalks*). Izazov pri usklađivanju shema metapodataka predstavlja nejednakost elemenata strukture koji se javljaju u različitim domenama koje ti standardi predstavljaju. Jedna shema može sadržavati strukturno polje koje ne postoji u drugoj shemi ili može imati polje koje je u drugoj shemi podijeljeno na dva ili više različita polja. To uzrokuje gubitke podataka prigodom mapiranja elemenata iz složene sheme u jednostavniju.

3.4.5. ENCIKLOPEDIČKA INTEROPERABILNOST

3.4.5.1. DBpedia

Projekt DBpedie temelji se na ekstrakciji metapodataka iz Wikipedijine građe, odnosno na strukturiranom Wikipedijinu sadržaju. Temelj sadržaja enciklopedičkih članaka objavljenih u mrežnom prostoru je tekst, međutim, u slučaju Wikipedije bogati su strukturiranim sadržajem u obliku infookvira, podataka o kategorijama kojima pripadaju, slikama, geokordinatama,

poveznicama na vanjske sadržaje, poveznicama među Wikipedijinim člancima, pa čak i na međujezičnoj razini, odnosno među različitim inačicama Wikipedije.

Najznačajniji strukturirani oblici vezani uz izgradnju projekta DBpedije su infookviri koji donose najznačajnije podatke vezane uz članke koje predstavljaju, a velik broj predložaka infookvira koji je razvijen omogućuje ekstrakciju velikog broja raznolikih metapodataka.

Domenu znanja DBpedije čine instance (objekti), klase (zbirke ili tipovi objekata), atributi (pripadajuća svojstva, pojave, karakteristike ili parametri koje objekt može imati ili distribuirati) te odnosi (način na koji se objekti odnose jedni prema drugima). Na taj način znanje je organizirano kao ontologija, odnosno formalan prikaz pojmova s dobro definiranim međusobnim odnosima. Ekstrakcijom metapodataka opisuje velik broj entiteta (osobe, mjesta, glazbeni radovi, filmovi, organizacije) u obliku RDF trojki (tripleta; vidi poglavlje 3. 4. 5. 1. 1.) ekstrahiranih iz različitih višejezičnih izdanja Wikipedije. DBpedia pokriva mnoge domene, ažurira se usporedno s Wikipedijom, višejezična je i dostupna na internetu.[124]

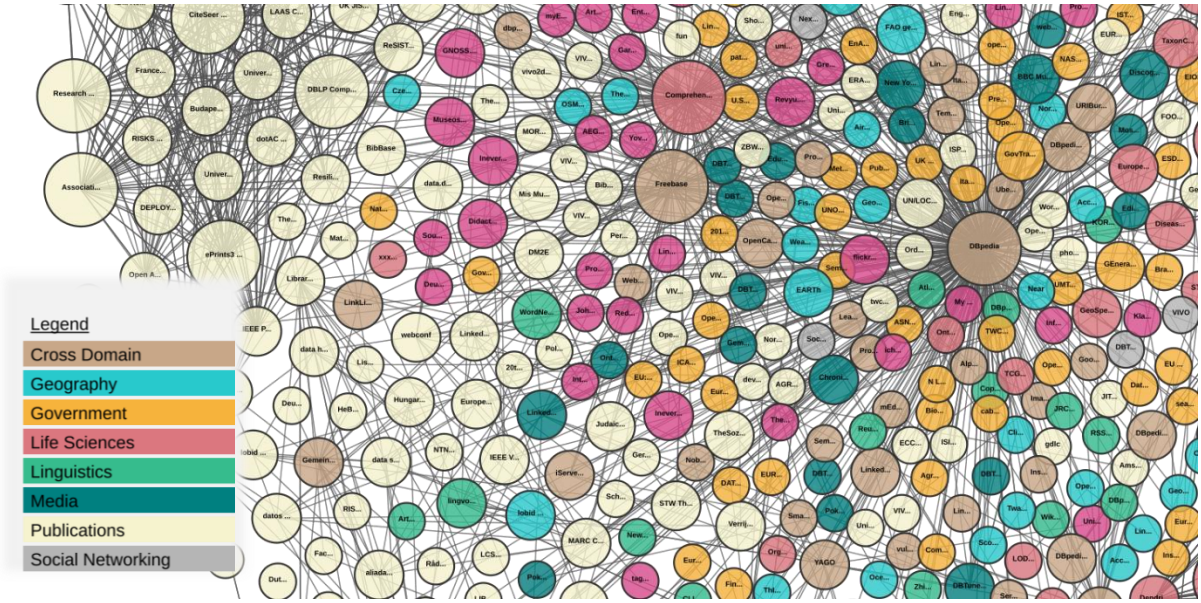
Kako bi izgradila svoj sadržaj, DBpedia koristi ekstraktore koji dohvaćaju specifične dijelove podataka iz Wikipedijine građe kao što su sažetci, slike, vanjske poveznice, geokordinate, osobni podatci osoba, kategorije i dr. Jedan je od izazova pri izvlačenju informacija iz Wikipedije izražavanje istih pojmova s pomoću različitih parametara, odnosno neusklađenosti u označavanju istih elemenata strukture (nepostojanje jedinstvenoga metapodatkovnog standarda). Kao rezultat toga, razvijen je jezik za mapiranje DBpedije koji pomaže u mapiranju tih svojstava u ontologiju uz smanjenje broja sinonima. [125]

Engleska verzija DBpedije opisuje oko 4,58 milijuna entiteta, od čega su 4,22 milijuna klasificirana u konzistentnu ontologiju, uključujući 1 445 000 osoba, 735 000 mjesta (uključujući 478 000 naseljenih mjesta), 411 000 umjetničkih djela (uključujući 123 000 glazbenih albuma, 87 000 filmova i 19 000 videoigara), 241 000 organizacija (uključujući 58 000 tvrtki i 49 000 obrazovnih ustanova), 251 000 vrsta organizama i 6000 bolesti. DBpedia je povezana s ostalim povezanim skupovima podataka s približno 50 milijuna RDF veza. Sveukupno izdanje DBpedije sastoji se od tri milijarde informacija (RDF trojki) od kojih je 580 milijuna ekstrahirano iz engleskog izdanja Wikipedije, 2,46 milijardi iz drugih jezičnih izdanja

[124] https://wiki.dbpedia.org/ (28. 8. 2019.)

[125] O projektu DBprdije u: Lehmann, J., Isele, R., Jakob, M.: DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. Semantic Web, 6(2015) 2.

uključujući i hrvatsko[126], te su objavljeni i interoperabilni u okviru tehnologija semantičkoga weba. Slika 16 prikazuje DBpediju s drugim bazama podataka objavljenima kao dio semantičkoga weba i njihovim međusobnim vezama.



Slika 16. Skupovi podataka objavljeni u okviru semantičkoga weba s međusobnim vezama (Izvor: https://en.wikipedia.org/wiki/Linked_data)

Uz interoperabilnost s metapodatcima drugih skupina podataka zahvaljujući postojanoj semantici, sadržaj objavljen isključivo u obliku metapodataka kakav predstavlja projekt DBpedije moguće je pretraživati postavljanjem složenih semantičkih upita nad bazom podataka koja predstavlja cjelokupnu Wikipedijinu građu[127], a ne samo natuknice i dio teksta koji znakovno odgovaraju upitu, što je moguće u slučaju Wikipedije kao mrežne enciklopedije.

Kako DBpedia objedinjuje podatke iz različitih domena, njezni skupovi podataka mogu se koristiti ne samo za poboljšano pretraživanje Wikipedijinog sadržaja već mogu i poslužiti kao izvor podataka za razvoj drugih mrežnih aplikacija. Njihov razvoj i mogućnosti mogu se temeljiti na pretraživanju DBpedije kao ontologijske baze podataka i korištenju rezultata tih pretraživanja, što je dostupno pod licencijama Creative Commons Imenovanje-Dijeli pod istim

[126] https://wiki.dbpedia.org/about (28. 8. 2019.)

[127] Npr. za upit Zagreb, koji predstavlja objekt, kao rezultat pretraživanja niže se niz atributa, poput imena gradonačelnika, poštanskoga broja, osoba rođenih, umrlih i aktivnih u gradu Zagrebu, tvrtki koje su smještene u Zagrebu i dr. s pripadajućim vrijednostima. Dostupno na: http://dbpedia.org/page/Zagreb (11. 9. 2019.).

uvjetima (engl. *Creative Commons Attribution-Share Alike*) [128] i GNU licenca slobodne dokumentacije (engl. *GNU Free Documentation Licence*).[129]

Kako je korpus DBpedije zapisan u obliku RDF trojki, za postavljanje upita i dohvaćanje podataka potrebno je koristiti SPARQL upitni jezik (engl. *Simple Protocol and RDF Query Language)*. Budući da je on većini korisnika nerazumljiv, razvijen je niz mrežnih aplikacija koje se naslanjaju na DBpedijinu bazu podataka i omogućuju njeno korištenje bez potrebnih znanja o SPARQL-u. Primjer toga su fasetni sustavi za pregledavanje LODmilla[130], LodLive[131], Graphity Browser at Semantic Report[132], gFacet[133], Openlink Virtuoso[134], SpaceTime[135] i dr. Primjeri sustava koji predstavljaju korisnička sučelja za dohvaćanje ili pretraživanje DBpedijinih RDF zapisa (engl. *Query Builder*) su Sparklis[136] i RelFinder[137].

DBpedia svojom izvedbom podržava procese obrade prirodnog jezika (poglavlje 5. 1.). Za tu namjenu DBpedia uključuje specijalizirane skupove podataka koji se koriste za procjenu dvosmislenosti izraza, prigodom odabira nedvosmislenih identifikatora dvosmislenim izazima ili za pružanje alternativnih naziva pojmovima u tekstu. Time unapređuje postavljanje semantički ispravih anotacija, budući da predstavlja strukturiranu sliku Wikipedije s mnoštvom podataka vezanih uz pojmove u stvarnome svijetu koji se kao identifikatori (URI) koriste za označavanje naziva tekstova prirodnog jezika. Značajan potencijal DBpedija pruža automatskom prepoznavanju naziva. Primjer slobodno dostupnoga alata koji omogućuje

[128] https://creativecommons.org/licenses/by-sa/4.0/deed.hr (5. 5. 2020.)

[129] https://www.gnu.org/licenses/fdl-1.3.html (5. 5. 2020.)

[130] https://wiki.dbpedia.org/projects/lodmilla (24. 1. 2020.)

[131] https://wiki.dbpedia.org/projects/lodlive (24. 1. 2020.)

[132] https://wiki.dbpedia.org/projects/graphity-browser-semantic-reports (24. 1. 2020.)

[133] https://wiki.dbpedia.org/projects/gfacet (24. 1. 2020.)

[134] https://wiki.dbpedia.org/projects/openlink-virtuoso-built-faceted-browser-and-search-find-service (24.1.2020.)

[135] https://wiki.dbpedia.org/projects/spacetime (24. 1. 2020.)

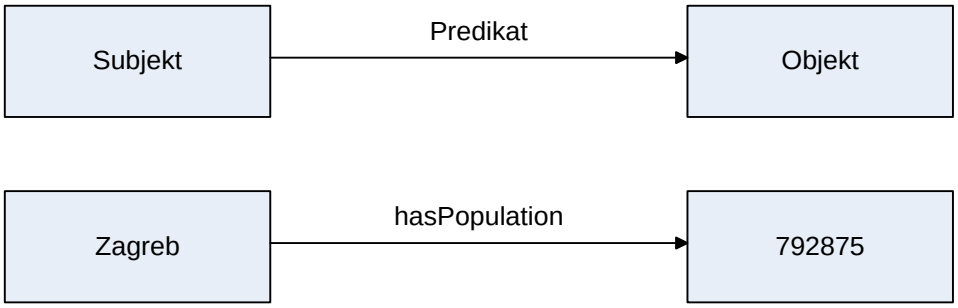
[136] https://wiki.dbpedia.org/projects/sparklis (24. 1. 2020.)

[137] https://wiki.dbpedia.org/projects/relfinder (24. 1. 2020.)

korištenje DBpedije u tu svrhu jest DBpedia Spotlight[138] koji kao kontekstualni model omogućava povezivanje izraza u tekstu s resursima DBpedije na semantičkoj razini.

3.4.5.1.1. Semantički web

DBpedijina građa objavljena je u okviru Semantičkoga weba, koji kao nadogradnju postojećeg mrežnog sustava na kojem se objavljuje sadržaj objavljuje strukturirane podatke koji su podložni strojnoj obradi i ljudima u međusobnoj suradnji.[139] Na najosnovnijoj razini semantički web počiva na sloju koji služi za definiranje pojmova mreži (www), a čine ga standard za kodiranje niza znakova uporabom kojega svi ljudski jezici mogu biti napisani i čitani na mreži (Unicode) te niz znakova koji osiguravaju standardni oblik za identifikaciju izvora (engl. *Uniform Resource Identifier*, URI). Drugi sloj arhitekture semanitčkoga weba je XML koji predstavlja temeljnu strukturu zapisa podataka i omogućuje korištenje zajedničke sintakse na semantičkome webu. Treći sloj predstavlja okvir za opisivanje informacija i resursa na webu RDF. Temelji se na trojkama (tripletima) koje sadrže subjekt, predikat i objekt (slika 17).



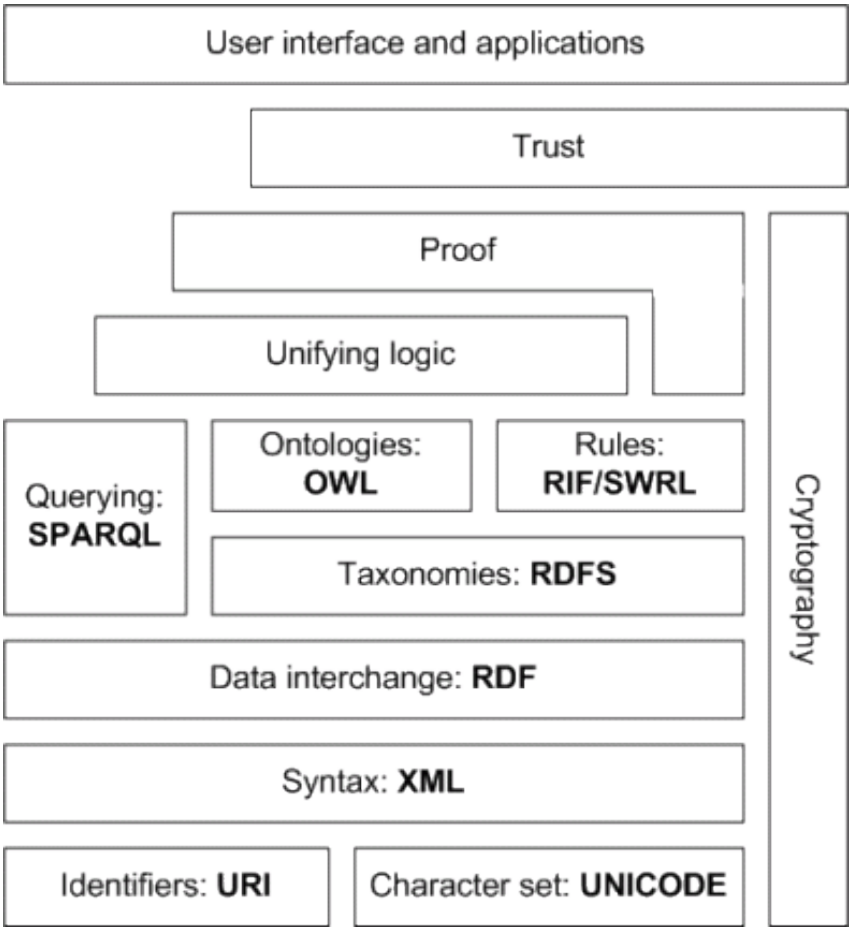
Slika 17. Shematski prikaz RDF trojke (tripleta)

Trojke čine grafove s dva čvorišta povezane predikatom. Subjekt predstavlja resurs i definiran je URI-jem, predikat je kao odnos također definiran URI-jem, a objekt se može definirati kao URI ili literalna vrijednost. Ovakav način sastavljanja informacija kao nadsloja koji predstavlja mrežne resurse na internetu omogućuje sustavima postavljanje složenih upita, pretraživanje i analiziranje sadržaja. RDF Schema (RDFS) predstavlja rječnik, kako bi se

[138] Više o projektu DBpedia Spotlight u: Mendes, P. N. et al.: DBpedia Spotlight: Shedding Light on the Web of Documents. U: Proceedings of the 7th International Conference on Semantic Systems (I-Semantics 2011). Graz, 2011., str. 1–8. Također i na adresi: https://www.dbpedia-spotlight.org/ (24. 1. 2020.)

[139] Izgradnja semantičkoga weba ovisi o dostupnoj količini metapodataka trenutačnih mrežnih sustava. Semantički web dodaje formalne strukture i semantička obilježja na mrežni sadržaj u svrhu boljeg upravljanja i pristupa.

omogućio opis klasa i svojstava kao glavnih elemenata ontologija. Sukladno tomu, RDF je model koji povezuje podatke, a RDFS daje osnovni smisao i značenje vezama te omogućuje organizaciju podataka. Krovni sloj RDF-a predstavlja ontologijski jezik OWL (engl. *Web Ontology Language*) koji pruža dodatnu standardiziranost podataka i definiranu sintaksu te pojačanu interoperabilnost. Predstavlja svojevrsni gramatički sustav semantičkoga weba. Gornji slojevi semantičkoga weba služe prikazivanju pouzdanih rezultata pri radu aplikacija.[140] Slika 18 donosi prikaz arhitekture semantičkoga weba.



Slika 18. Arhitektura semantičkoga weba (Izvor: https://www.obitko.com/tutorials/ontologies-semantic-web/semantic-web-architecture.html)

Budući da DBpedia kao semantičko ogledalo Wikipedije sadrži izrazito velik broj RDF zapisa (trojki), predstavlja mjesto automatskog povezivanja na semantičkoj razini i interoperabilnosti s mnogim podatcima objedinjenima u okviru semantičkoga weba iz mnogih

[140] O semantičkom webu i objavljivanju i povezivanju strukturiranih podataka (engl. *Linked Dana*) u: Bizer, C., Heath, T., Berners-Lee, T.: Linked Data – The Story So Far. (2009)

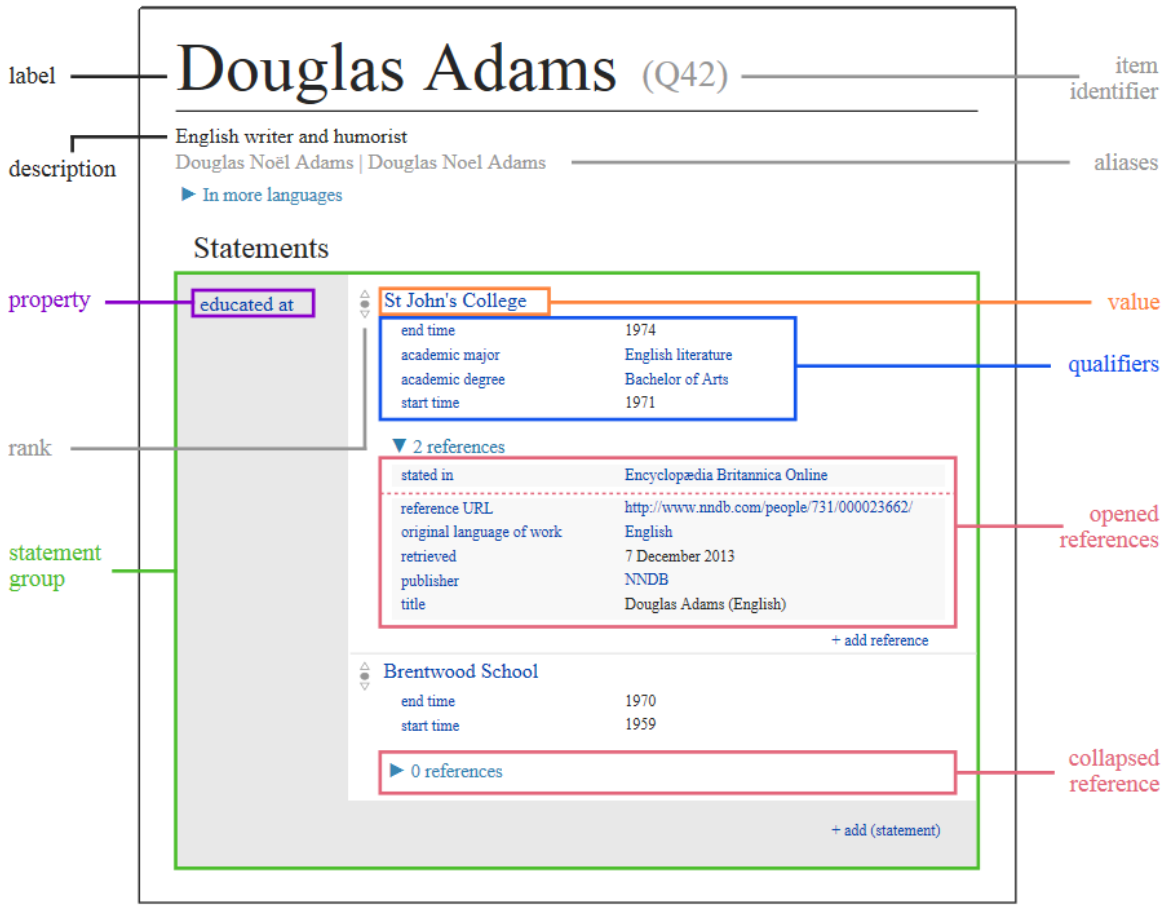
skupina podataka, predstavljajući svojevrsni čvor u naporima izgradnje povezanih, interoperabilnih podataka u mrežnome prostoru.

3.4.5.2. Wikidata

Wikidata kao baza znanja temelji se isključivo na strukturiranome sadržaju, odnosno opsežnoj kolekciji pojmova koje opisuju zasebni entiteti s pridodanim vrijednostima. Kao i projekt DBpedije, na strukturiran i strojno obradiv način predstavlja sadržaj Wikipedije uz znatnu razliku. Wikidata, kao i Wikipedija, temelji se na kolektivnom naporu zajednice koja uređuje njezin sadržaj, koji je tako neovisan o sadržaju Wikipedije. Neovisno stvaranje i održavanje sadržaja prednost je u odnosu na automatsku ekstrakciju sadržaja koja ovisi o stanju izvornika, kao što je u slučaju DBpedije.

Strukturirani sadržaj Wikidate bilježi se u obliku izjava (engl. *statements*) koje mogu sadržavati više formulacija atribut-vrijednost (engl. *property-value pairs*) kao predstavnike podataka. Npr. atribut bi predstavljao tip podatka (autor), a vrijednost na što se odnosi (Miroslav Krleža). Atribut može poprimiti više vrijednosti izraženih kao tekst, brojeve, medijske datoteke i dr. S pomoću kvalifikatora (engl. *qualifiers*) moguće je povećati kontekstualnu informativnost izjava. Izjave donose i poveznice koje vežu s relevantnim izvorima (engl. *references*). Svaka izjava dio je stavke (engl. *item, label*) kao jedinice sadržaja Wikidate, što je jednako Wikipedijinim člancima. Svaka je stavka označena brojčano s prefiksom Q, što omogućuje prepoznavanje stavki (tema) neovisno o Wikipedijinoj višejezičnosti. Sukladno tomu, Wikidata se sastoji od niza infookvira kao strukturiranoga sadržaja neovisnog o Wikipediji (slika 19).[141]

[141] Više o projektu Wikidata i vezi s Wikipedijom u: Piscopo, A. et al.: What do Wikidata and Wikipedia Have in Common? An Analysis of their Use of External References. U: Proceedings of the 13th International Symposium on Open Collaboration. Galway, 2017.



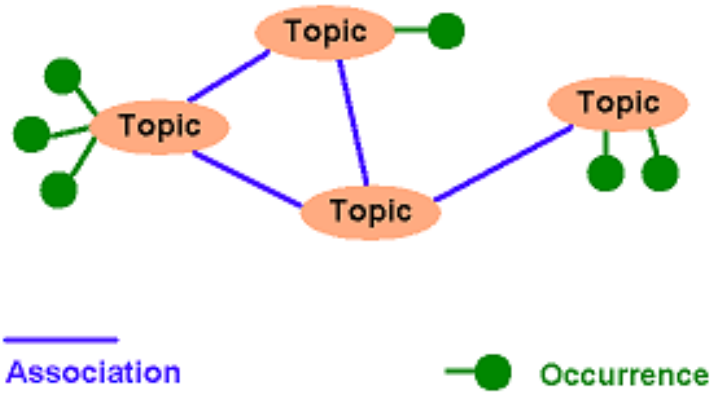
Slika 19. Sastavnice strukturiranoga sadržaja Wikidate (Izvor: https://en.wikipedia.org/wiki/Wikidata)

Strukturirani sadržaj Wikidate izveden je i u RDF metapodatkovnome standardu i na taj način uključen u okviru semantičkoga weba gdje svoj sadržaj čini interoperabilnim s drugim izvorima podataka.[142] Kao i kod DBpedije, koja ponajprije proizlazi iz metapodataka Wikipedijinih infookvira, interoperabilnost Wikidate moguća je zahvaljujući visokostrukturiranom sadržaju s izraženim semantičkim obilježjima. Wikidata zapravo predstavlja mrežno enciklopedičko izdanje bez prisutnosti prirodnog jezika, strogo strukturirano i održavano uz pomoć urednika, koje je razumljivo i ljudima i računalima, odnosno predstavlja javno dostupnu enciklopedičku bazu podataka.

[142] O uključivanju Wikidate u prostor sematičkoga weba i postizanju interoperabilnosti u: Erxleben, F. et al.: Introducing Wikidata to the Linked Dana Web. U: Proceedings of the 13th International Semantic Web Conference - Part I. Riva del Garda, 2014., str. 50–65.

3.4.5.3. Gran Enciclopedia Aragonesa

Označavanje digitaliziranoga teksta Gran Enciclopedie Aragonese[143] izvedeno je prema standardu tematskih mapa (engl. *topic map*) s glavnim ciljem organizacije, korištenja i razmjene sustava za organizaciju znanja (slika 20). Glavna zamisao tematskih mapa jest koncept temeljnoga modela koji se sastoji od naziva, izvora i vrsta odnosa.



Slika 20. Prikaz temeljnih koncepata tematskih mapa (Izvor: https://en.wikipedia.org/wiki/Topic_map)

Mrežna verzija standarda poznata je pod nazivom XTM (engl. *XML Topic Maps*) zasnovana na XML-u i po tome komplementarna s drugim mrežnim sustavima. Elementi strukture omogućavaju formulaciju tri ključna koncepta tematskih mapa: teme (engl. *topic*), veze (engl. *associations*) i pojave (engl. *occurrences*). Teme predstavljaju bilo koji koncept, entitet i dr., asocijacije veze među temama, a pojave informacijske resurse relevantne za određenu temu. Dodatna preciznost pri uspostavljanju veza između tema i vanjskih izvora moguća je dodavanjem uloga (engl. *role*) koje ih dodatno određuju. Npr. ako kao teme egzistiraju jedinice sadržaja poput *Miroslav Krleža* i *Povratak Filipa Latinovicza* koje imaju određenu međusobnu vezu, prva će biti okarakterizirana kao autor, a potonja kao publikacija. Pojave koje se također vežu uz teme mogu biti objekti u mrežnome prostoru poput standardnih Word, Excel ili Power Point datoteka, mrežnih stranica, XML dokumetata, baza podataka, slika i dr., ali i izvora poput bibliografskih referenci u konvencionalnom smislu.[144]

[143] http://www.enciclopedia-aragonesa.com/ (9. 10. 2019.)

[144] Više o problematici vezanoj uz tematske mape na https://ontopia.net/topicmaps/materials/tao.html i http://www.topicmaps.org/ (11. 9. 2019.)

Tematske mape slične su konceptualnim i mentalnim mapama kao tehnikama za vizualiziranje veza među različitim pojmovima, a tehničkom izvedbom RDF shemi metapodataka koja je izvedena u obliku tripleta. Razlika je u tome što je tehnologija tematskih mapa više usredotočena na entitete (teme) i njhove međusobne veze, što rezultira sustavom više namijenjenim ljudima, a ne računalima.

Na primjeru Gran Enciclopedie Aragonese prikazana je izrazita pogodnost pristupa tematskih mapa za stvaranje informacijskoga sustava na temelju enciklopedičkih tekstova, budući da jedinice sadržaja takvih izdanja predstavljaju entitete (teme) koje su međusobno relevantne i moguće je ostvariti veze među njima, a također su u velikoj mjeri vezane uz vanjske izvore (pojave) koji se također mogu vezati uz njih. Kako bi stvorili semantički nadsloj koji predstavlja i veže enciklopedički sadržaj, razvili su algoritme namijenjene analizi teksta u cilju definiranja koncepata za izgradnju sustava tematskih mapa.[145] Na taj način omogućili su jednostavnije korištenje sustava, povećali dohvat sadržaja vezan uz pojedinu temu te ga povezali s drugim izvorima informacija zahvaljujući tehnologiji tematskih mapa.

[145] Više u: Garrido P., Tramullas, J., Martinez, F. J.: Application of Semantic Tagging to Generate Superimposed Information on a Digital Encyclopedia. U: Metadata and Semantic Research, 4th International Conference. Alcalà de Henares, 2010., str. 84–94.

4. ANALIZA SADRŽAJA MREŽNOGA ENCIKLOPEDIČKOG SADRŽAJA IZ TEHNIČKOGA PODRUČJA

Analiza (raščlamba, raščimba) predstavlja raščlanjivanje neke složene cjeline (materijalne, procesualne, pojmovne) na sastavne elemente, osobito radi proučavanja.[146] Krippendorff[147] donosi definiciju analize sadržaja u širem smislu te ju navodi kao istraživačku tehniku namijenjenu donošenju replikabilnih i valjanih zaključaka iz tekstova (ili drugih materijala) u kontekste njihove uporabe.

Osim tekstova, sadržaji komunikacije mogu egzistirati kao slike ili video i zvučni zapisi. Analiza sadržaja zasigurno predstavlja interdisciplinarno područje, a temelji se na sustavnom razmatranju sadržaja kojemu su dodijeljene oznake u cilju upućivanja na prisutnost njegovih smislenih dijelova. Sustavnim označavanjem sadržaja kvantitativno se anliziraju obrasci sadržaja koristeći statističke metode, a kvalitativnim metodama određuju se značenja dijelova sadržaja.

Kvantitativna analiza sadržaja temelji se na deduktivnom pristupu te joj prethodi postavljanje hipoteza vezanih uz odnose razmatranih varijabli. Omogućava identifikaciju određenih obrazaca podataka te testiranje njihovih odnosa i uspostavljanje uzročno-posljedičnih veza. Rezultati dobiveni na određenom uzorku primjenjuju se na cjelokupnu populaciju. Kvalitativna analiza sadržaja induktivna je i nasuprot hipotezama započinje istraživačkim pitanjima. Nastoji obuhvatiti značenja (kontekste) i razjasniti organizaciju i proces prezentacije dijelova sadržaja.

Analiza sadržaja nerijetko obuhvaća kombinaciju kvantitativnog i kvalitatvnog pristupa. Kvantitativna metoda pruža vjerodostojnost i dosljednost podataka (uzorka) te mogućnost generaliziranja rezultata i donošenje objektivnih zaključaka, dok kvalitativna analiza sadržaja koja nije ograničena varijablama uključenima u istraživanje može poslužiti kao dopuna

[146] Prema natuknici: analiza. *Hrvatska enciklopedija, mrežno izdanje.* Leksikografski zavod Miroslav Krleža, 2020. http://www.enciklopedija.hr/Natuknica.aspx?ID=2464 (1. 5. 2020.)

[147] Krippendorff, K.: Content analysis: An introduction to its methodology. Thousands Oaks, 2004., str. 18.

kvantitativnoj analizi, pružiti potpuno razumijevanje sadržaja te omogućiti stvaranje novih spoznaja i teorija.

4.1. OBUHVAĆENA DJELA

Ovim istraživanjem sadržaja mrežnih enciklopedičkih izdanja bilo je uključeno šest javno dostupnih projekata, a pridodano je i izdanje Der Brockhaus multimedial 2008 premium na kompaktnom disku (CD-u), kao njemačko izdanje koje baštini dugu tradiciju Brockhaus enciklopedije. Kako su uz to ispitane i tri inačice Wikipedije (hrvatska, engleska i njemačka) koje po svemu predstavljaju zasebne enciklopedije, broj izdanja uključenih u ispitivanje iznosio je deset. Osim spomenutih projekata, glavninu uzorka činila su temeljna moderna izdanja hrvatske enciklopedike, odnosno Hrvatska tehnička enciklopedija, Hrvatska enciklopedija, Proleksis enciklopedija, Istarska enciklopedija i Hrvatski biografski leksikon, te mrežno izdanje Encyclopaedije Britannice.

4.1.1. HRVATSKA TEHNIČKA ENCIKLOPEDIJA

Hrvatska tehnička enciklopedija (HTE)[148] središnje je izdanje ovog istraživanja i ključno za provođenje strukturiranja i izgradnje modela interoperabilnosti. Projekt je Leksikografskog zavoda Miroslav Krleža pokrenut 2014. godine, kojim se prvi put, po enciklopedičkim principima, nastoji kontinuirano prikupiti i ustanoviti znanje o povijesti i današnjim dosezima tehnike u Hrvatskoj, odnosno pridonijeti stvaranju spoznaje o bogatoj hrvatskoj tehničkoj baštini. Ujedno je i okosnica Portala hrvatske tehničke baštine Leksikografskog zavoda Miroslav Krleža, pokrenutog potkraj 2018. godine te čini platformu za istraživanje povijesti i dosega tehnike u Hrvatskoj. Pokretač i glavni urednik projekta je Zdenko Jecić.

Projekt Hrvatske tehničke enciklopedije usporedno se izvodi kao tiskano i mrežno izdanje, te po tome predstavlja prvijenac u hrvatskoj enciklopedici.[149] Službeno je predstavljena 10. prosinca 2018. godine, na dan 100. obljetnice donošenja odluke o osnivanju Tehničke

[148] Hrvatska tehnička enciklopedija. Leksikografski zavod Miroslav Krleža, Zagreb, 2018–, 1. svezak

[149] Više o pothvatu Hrvatske tehničke enciklopedije i Portalu hrvatske tehničke baštine u: Jecić, Z.: Od Tehničke enciklopedije do Hrvatske tehničke enciklopedije. Studia lexicographica, 12(2018) 23, str. 53–79.

visoke škole u Zagrebu, prethodnice današnjih tehničkih fakulteta. Tiskano izdanje organizirano je tematski u četiri sveska, od kojih svaki obrađuje zaokruženu cjelinu proizašlu iz tradicionalnog ustroja tehničkoga visokog školstva u Hrvatskoj, odnosno organizacije Tehničkoga fakulteta u Zagrebu.[150]

Mrežno izdanje Hrvatske tehničke enciklopedije javno je dostupan servis, kontinuirano se razvija u smislu nadogradnje korisničke aplikacije u cilju unapređenja funkcionalnosti i učinkovitosti korištenja, te nadogradnje sadržaja, odnosno objavljivanja novih članaka u trenutku kada zadovolje stroge leksikografske standarde. Svi su članci potpisani uz donesen datum njihova objavljivanja i ažuriranja. Uz to, svakom članku moguće je uređivanje dodatnih rubrika koje omogućuju povezivanje s vanjskim izvorima poput digitalnih zbirki, arhiva, videozapisa, knjižnica, muzeja ili drugih mrežnih stranica posvećenih temi koju članak obrađuje.[151] Sadržaj prvoga sveska izveden je u 992 članka.

Portal hrvatske tehničke baštine nudi i cjelokupno digitalizirano izdanje Tehničke enciklopedije[152], koja se u 13 svezaka objavljivala u razdoblju od 1963. do 1997. godine kao najopsežnije enciklopedičko djelo na hrvatskome jeziku ikad priređeno. Obuhvaća 645 članaka na 9500 stranica s 22 000 ilustracija i tehničkih crteža.[153]

4.1.2. HRVATSKA ENCIKLOPEDIJA

Hrvatska enciklopedija (HE)[154] izdavala se kao višesveščano enciklopedičko djelo Leksikografskog zavoda Miroslav Krleža u razdoblju od 1999. do 2009. godine. Objedinila je 1070 autora i vanjskih suradnika, sadrži ukupno 9272 stranice, 67 077 članaka s više od 17 000 ilustracija i priloga. Objavljeno je ukupno 11 svezaka, a glavni urednici bili su Dalibor Brozović (1–3. svezak), August Kovačec (4–7. svezak), te Slaven Ravlić (8–11. svezak) koji je i dalje

[150] Zagrebački Tehnički fakultet osnovan 1926. sljednik je Tehničke visoke škole. Godine 1956. iz njegova ustroja su proizašli Arhitektonsko-građevinsko-geodetski, Elektrotehnički, Strojarsko-brodograđevni i Kemijsko-prehrambeno-rudarski fakultet. Upravo ta podjela čini i organizaciju znanja Hrvatske tehničke enciklopedije po svescima.

[151] Više o strukturi i izvedbi mrežnoga izdanja Hrvatske tehničke enciklopedije i Portala hrvatske tehničke baštine u poglavlju 3. 3. 1.

[152] Vidi bilješku 28.

[153] Objavljena je i cjelokupna bibliografija Tehničke enciklopedije, koja uz abecedni popis članaka po svescima sadrži i razrađeno imenski i strukovno kazalo. Više u: Klobučar Srbić, I., Smolčić, I., Tolj, J.: Bibliografija Tehničke enciklopedije (1963–97). Studia lexicographica, 12(2018) 23, str. 199–257.

[154] Hrvatska enciklopedija. Leksikografski Zavod Miroslav Krleža, Zagreb, 1999−2009, 11 svezaka.

voditelj mrežnoga projekta. Mrežno izdanje[155] temelji se na tiskanom, tradicionalnom izdanju, objavljeno je 2013. godine te je sustavno podvrgnuto nadopunjavanju i ažuriranju sadržaja. Članci su organizirani u taksonomiju od pet glavnih područja (Geografija, Jezik i književnost, Povijest i društvo, Priroda i tehnika i Umjetnost) koja obuhvaćaju ukupno 124 struke u koje je raspoređen cjelokupan korpus enciklopedije.[156] Veliki pregledni i makropedijski članci podijeljeni su u poglavlja, koja kao hiperveze omogućuju lakše pregledavanje tako velikog opsega sadržaja unutar jedne jedinice.

Hrvatska enciklopedija je središnje izdanje Leksikografskog zavoda Miroslav Krleža, svojom veličinom i zastupljenošću tema ostvaruje najveći odjek u mrežnome prostoru, te uz hrvatsku Wikipediju ostvaruje najveći utjecaj u pružanju usluga enciklopedičkoga sadržaja na hrvatskom jeziku.

4.1.3. PROLEKSIS ENCIKLOPEDIJA

Mrežno izdanje Proleksis enciklopedije (PE)[157] zasniva se na tiskanom izdanju Opće i nacionalne enciklopedije u 20 knjiga.[158] Knjige (svesci) su izlazile u razdoblju od 2005. do 2007. godine, a 2009. godine izašla je i dopunska, 21. knjiga. Izdavači su bili tvrtke PRO LEKSIS d. o. o. i Večernji list d. d., uz potporu Leksikografskog zavoda Miroslav Krleža. Tiskano izdanje sastoji se od 6000 stranica, 80 000 članaka i 19 000 ilustracija. Članci su kraćega formata, te djelo ne sadrži veće geografske karte ili reprodukcije umjetničkih djela većega formata.

Godine 2009. predstavljeno je i DVD izdanje, a zatim i eksperimentalno mrežno izdanje Proleksis enciklopedije, u suradnji s CARNetom, a za korisnike te Hrvatske akademske i istraživačke mreže. Potkraj 2010. godine daljnju brigu o uređivanju i izdavanju mrežnog izdanja Proleksis enciklopedije preuzima Leksikografski zavod Miroslav Krleža, na koji su prenesena

[155] http://www.enciklopedija.hr/ (13. 9. 2019.)

[156] Više o cjelokupnom pothvatu Hrvatske enciklopedije od tiskanoga do mrežnog izdanja u: Starčević Stančić, I., Kraus, C.: Hrvatska enciklopedija – od tiskanoga do mrežnoga izdanja. Studia lexicographica, 8(2014) 1(14).

[157] https://proleksis.lzmk.hr/ (13. 9. 2019.)

[158] Opća i nacionalna enciklopedija u 20 knjiga. PRO LEKSIS, Večernji list, Zagreb, 2005–2007., 21 knjiga.

i sva izdavačka prava. Sadržaj je organiziran u ukupno 21 veće područje, koja najbolje zaokružuju sadržaj izdanja.[159]

4.1.4. ISTARSKA ENCIKLOPEDIJA

Istarska enciklopedija (IE)[160] objavljena je 2005. godine (glavni urednici Miroslav Bertoša i Robert Matijašić), sadrži 3094 članka s 1410 ilustracija (fotografija, karata, grafikona i tablica). Mrežna inačica[161] u potpunosti se temelji na tiskanom izdanju.

Jednosveščano je leksikografsko djelo užega područja interesa, koje obrađuje teme vezane uz istarski poluotok. Organizirajući sadržaj u članke o općim i zemljopisnim pojmovima, ustanovama i osobama, predstavlja sintezu znanja vezanog uz istarski prostor s gledišta geografije, povijesti, kulture, gospodarstva. Svojim postavkama nastoji prikazati višeznačnost utjecaja koji su oblikovali Istru tijekom povijesti. Istarska enciklopedija prvi je leksikografski projekt u Hrvatskoj koji obuhvaća, s enciklopedičkoga gledišta, problematiku jedne hrvatske regije. Jedno je u nizu sličnih izdanja Leksikografskog zavoda Miroslav Krleža toga tipa izvedeno u suradnji s područnom i lokalnom zajednicom, budući da je objavljeno i djelo Enciklopedija hrvatskoga zagorja, a u izradi je i projekt vezan za geografsko područje Turopolja.

4.1.5. HRVATSKI BIOGRAFSKI LEKSIKON

Hrvatski biografski leksikon (HBL)[162] opsežno je leksikografsko biobibliografsko djelo te jedan od temeljnih projekata Leksikografskog zavoda Miroslav Krleža. Po enciklopedijskim postavkama donosi biografske članke Hrvata, ali i pripadnika drugih naroda čije je djelovanje bilo relevantno za hrvatsku komponentu. Razmatra osobe rođene do 1945. godine, uz moguće iznimke. Prvi svezak objavljen je 1983. godine, a do danas je objavljeno osam svezaka tiskanoga izdanja s 11 423 članka. Glavni urednici bili su Nikica Kolumbić (1. svezak), Aleksandar Stipčević (2. svezak) i Trpimir Macan (3–8. svezak). Danas projekt vodi Nikša

[159] Vidi poglavlje 3. 3. 2.

[160] Istarska enciklopedija. Leksikografski zavod Miroslav Krleža, Zagreb, 2005.

[161] http://istra.lzmk.hr/ (13. 9. 2019.)

[162] Hrvatski biografski leksikon. Leksikografski zavod Miroslav Krleža, Zagreb, 1983–, 8. svezaka

Lučić. Od 2015. godine mrežno izdanje Hrvatskoga biografskog leksikona[163] donosi sve članke objavljene u tiskanom izdanju, u izvornom obliku. Također, objavljen je dio sadržaja koji još slijedi, a predstavlja opsežne životopise hrvatskih velikana.

Ovaj projekt Leksikografskog zavoda Miroslav Krleža naglašene je temeljitosti, sustavnosti i pouzdanosti, rad na člancima nerijetko se temelji na znanstvenim principima[164], a područje tehnike obrađeno je vrlo detaljno i često predstavlja jedini izvor informiranja o pojedinim temama.

4.1.6. ENCYCLOPAEDIA BRITANNICA

Encyclopaedia Britannica (EB)[165] najstarija je opća enciklopedija na engleskom jeziku. Prvi je put objavljena 1768. godine u Edinburghu u Škotskoj, a do danas je ostvarena u 15 izdanja. Posljednje, prvi put objavljeno 1974. godine, sastoji se od 32 sveska raspoređena u tri dijela, odnosno organizirana u Micropaediju (1–12. svezak)[166], Macropaediju (13–29. svezak)[167] te dva sveska Propaedije[168]. Posljednja inačica objavljena 2010. godine sadrži 32 sveska, više od 40 000 članaka na 18 251 stranici teksta s 8500 fotografija, karata i ilustracija. Sadržaj nije izveden kroz Micro i Macropaediju, već kao jedinstvena cjelina karakteristična za prvih 14 izdanja.

Godine 1985. sastavnicama Encyclopaedije Britannice pridodan je i indeks, a 1994. godine cjelokupan korpus tiskanog izdanja postao je dostupan i u mrežnome prostoru[169] zahvaljujući razvoju projekta Britannica Online, koji se od svojih početaka zasnivao na

[163] http://hbl.lzmk.hr/ (13. 9. 2019.)

[164] U razdoblju od 2001–13. godine bio je ključni dio znanstvenog projekta Hrvatska bibliografska baština u potpori Ministarstva znanosti Republike Hrvatske.

[165] Encyclopaedia Britannica, Encyclopaedia Britannica Inc., Chicago, 1768–2010. (15. izdanje od 1974–1984. s 30 svezaka te od 1985. s 32 sveska).

[166] Micropaedija se sastoji od 12 svezaka koji donose niz kratkih natuknica, opsegom uglavnom manje od 750 riječi. Uz rijetke iznimke, uz sadržaj približno 65 000 članaka nisu priloženi podatci o bibliografskim referencama ni o autorima, te je cjelokupna građa ponajprije namijenjena brzoj provjeri činjenica i kao vodič za 700 opsežnih članaka Macropaedije.

[167] Macropaedija je u konačnici organizirana u 17 svezaka i približno 700 članaka u rasponu od 2 do 310 stranica teksta. Za razliku od Micropaedije, članci su uglavnom autorizirani s popisanim referencama.

[168] Propaedija predstavlja organizaciju znanja i tematski vodič kroz cjelokupan sadržaj Micropaedije i Macropaedije.

[169] https://www.britannica.com/ (14. 9. 2019.); o strukturiranosti mrežnog projekta Encyclopaedije Britannice više u poglavlju 3. 3. 3.

hipervezama i mrežnome pretraživanju sadržaja. Iste godine izdano je i prvo CD-ROM enciklopedijsko izdanje Encyclopaedije Britannice.

Također, svoje znanje i iskustvo Encyclopaedija Britannica koristi za razvoj niza drugih proizvoda i usluga koji proizlaze iz enciklopedičkoga djelovanja. Tako su nastali portali Britannica Kids[170], prilagođen djeci i mladima, te Britannica School[171], sadržaji ustrojeni kao servis učenicima osnovnih i srednjih škola. Također, Britannica je pokrenula i niz tiskanih publikacija, enciklopedičkih djela namijenjenih mlađoj publici poput Compton's by Britannica, My First Britannica, Discover America i Britannica Discovery Library, ali i niza drugih publikacija dostupnih i u digitalnom obliku.[172] Kao i u svojim početcima prije više stoljeća, Britannica je i danas važan dio sustava naobrazbe i informiranja velikog broja ljudi zahvaljujući neprestanom razvoju usmjerenom prilagodbi korisniku.[173]

4.1.7. DER BROCKHAUS MULTIMEDIAL 2008 PREMIUM

Brockhaus Enzyklopädie (BRO)[174] najveće je suvremeno enciklopedičko izdanje na njemačkom jeziku. Prvo izdanje izlazilo je u razdoblju od 1796. do 1808. godine, a do danas objavljeno je ukupno 21 izdanje. Godine 2014., nakon više od 200 godina, distribucija tiskanog izdanja u potpunosti je obustavljena. Od 2015. godine tvrtka Brockhaus SE GbmH iz Münchena razvija i omogućuje korištenje digitalnog portala Brockhaus Wisensservice, kao sljednika Brockhaus Enzyklopädije. Od svojih početaka 1796. godine pa nadalje, obrađuje širok raspon tema i predstavlja pravo sveobuhvatno enciklopedičko izdanje. Posljednje, 21. izdanje, sadrži oko 300 000 članaka, na 24 000 stranica teksta s približno 35 000 karata, grafikona i tablica.

Na njemu se temelje digitalne inačice, a u ovo istraživanje je uključeno izdanje Der Brockhaus multimedial 2008 premium distribuirano na DVD nosaču, odnosno elektronička

[170] https://kids.britannica.com/ (14. 9. 2019.)

[171] https://school.eb.com/ (14. 9. 2019.)

[172] Sve publikacije s pripadajućim bibliografskim podatcima moguće je dohvatiti na https://store.britannica.com/ (14. 9. 2019.)

[173] O razvoju i povijesti Enciklopedije Britannice, više u opsežnoj mrežnoj enciklopedičkoj natuknici Encyclopaedia Britannica. https://www.britannica.com/topic/Encyclopaedia-Britannica-English-language-reference-work (14. 9. 2019.)

[174] Brockhaus Enzyklopädie. F. A. Brockhaus AG, Leipzig, 1796–2006. (21. izdanje od 2005., 30 svezaka).

inačica enciklopedije Der Brockhaus in funfzehn Bänden.[175] Predstavlja nekoliko opsegom manjih izdanja[176] temeljenih na glavnom projektu Brockhaus Enzyklopädije. Ovo izdanje opremljeno je mnogim funkcionalnostima i dodatnim sadržajima, poput interaktivnog modula za putovanje kroz povijest arhitekture, atlasa s dva milijuna geografskih jedinica, planetarija s više od 120 000 objekata, videozapisa i animacija, kviza s više od 10 000 pitanja i dr. Glavninu sadržaja čine enciklopedički članci, njih 260 000 s više od 20 000 fotografija i ilustracija.

4.1.8. WIKIPEDIJA

Wikipedija[177] je višejezična mrežna enciklopedija otvorenoga tipa dostupna svima. Predstavlja kolaborativni pothvat svih koji žele sudjelovati, odnosno stvarati i uređivati sadržaj. Zahvaljujući kolaborativnom pothvatu svih njenih korisnika, podložna je brzom rastu sadržaja. Wikipedija je dio širega projekta Wikimedije[178], podržanog softverom naziva MediaWiki[179], koji je posebno razvijen kako bi omogućio njihov rad. Tu spadaju Wikidata[180], baza podataka (metapodataka) koja na visokostrukturirani način predstavlja građu drugih Wikimedia projekata, višejezični rječnik i tezaurus pojmova Wiktionary[181] (Wikirječnik), repozitorij slobodno dostupnih elektroničkih knjiga Wikibooks[182] (Wikiknjige), zbirka citata znamenitih osoba, književnih djela i poslovica Wikiquote[183] (Wikicitat), zatim repozitorij fotografija, crteža, videozapisa i glazbe Wikimedia Commons[184] (Zajednički poslužitelj), digitalna knjižnica Wikisource[185] (Wikizvor), zbirka tečajeva i projekt podržavanja obrazovnih i

[175] Der Brockhaus in fünfzehn Bänden, Mannheim : F.A. Brockhaus, 1997–

[176] Primjerice desetosveščano izdanje Brockhaus Enzyklopädie (Der Brockhaus in zehn Bänden, F. A. Brockhaus, Mannheim, 2005), šestsveščano (Der Brockhaus in sechs Bänden, F. A. Brockhaus, Mannheim, 2009) i dr.

[177] https://www.wikipedia.org/ (14. 1. 2020.); o strukturiranosti sadržaja Wikipedije vidi poglavlje 3. 3. 4.

[178] https://www.wikimedia.org/ (14. 1. 2020.)

[179] https://www.mediawiki.org/wiki/MediaWiki (14. 1. 2020.)

[180] https://www.wikidata.org/wiki/Wikidata:Main_Page (14. 1. 2020.)

[181] https://www.wiktionary.org/ (14. 1. 2020.)

[182] https://en.wikibooks.org/wiki/Main_Page (14. 1. 2020.)

[183] https://www.wikiquote.org/ (14. 1. 2020.)

[184] https://commons.wikimedia.org/wiki/Main_Page (14. 1. 2020.)

[185] https://wikisource.org/wiki/Main_Page (14. 1. 2020.)

znanstvenih zajednica Wikiversity[186] (Wikiučilište), taksonomski katalog vrsta Wikispecies[187] (Wikivrste), višejezični projekt pružanja turističkih informacija Wikivoyage[188], središnje mjesto koordinacije svih projekata Wikimedije Meta-Wiki.[189]

Unatoč odstupanjima od odrednica enciklopedičkoga koncepta, poput nedosljednosti prigodom stvaranja sadržaja, smanjene točnosti u odnosu na profesionalne enciklopedije zatvorenoga tipa[190], pristranosti i dr., slobodan pristup i mogućnosti sudjelovanja i stvaranja sadržaja učinili su Wikipediju najvećim i najkorištenijim mrežnim enciklopedičkim projektom, te u tom smislu predstavlja važan dio globalnog informacijskog sustava. Trenutačno je aktivno 294 inačice Wikipedije, odnosno mrežnih enciklopedija na 294 jezika.[191]

Ovim istraživanjem uključene su tri inačice Wikipedije kao zasebna enciklopedička djela, i to ona na hrvatskom (W_{hr}), engleskom (W_{eng}) i njemačkom (W_{de}) jeziku. Najopsežnija Wikipedija na engleskom jeziku prvi je put objavljena 10. siječnja 2001. godine što predstavlja početak cjelokupnoga višejezičnog projekta. Trenutačno sadrži 5 903 051 članak[192], međutim broj članaka znatno se mijenja svakoga dana. Njemačka inačica Wikipedija objavljena je kratko nakon engleske, u ožujku 2001. godine. S 2 428 738 članaka[193] trenutačno je četvrto najveće izdanje Wikipedije. Hrvatska Wikipedija dostupna je od 16. veljače 2003. godine. Trenutačno sadrži 208 240 članaka.[194]

4.2. METODOLOGIJA

Uzorkovanje (odabir natuknica u uzorku) provedeno je na deset enciklopedičkih izdanja navedenih u poglavlju 4. 1. Iz svakog navedenog izdanja izdvojeno je 50 natuknica, kojih odabir predstavlja cjelokupan sadržaj (popis natuknica ili članaka, odnosno abecedarij), pazeći

[186] https://www.wikiversity.org/ (14. 1. 2020.)

[187] https://species.wikimedia.org/wiki/Main_Page (14. 1. 2020.)

[188] https://www.wikivoyage.org/ (14. 1. 2020.)

[189] https://meta.wikimedia.org/wiki/Main_Page (14. 1. 2020.)

[190] Vidi poglavlje 2. 2. 1.

[191] Pristupljeno 1. kolovoza 2019.

[192] Pristupljeno 1. kolovoza 2019.

[193] Pristupljeno 1. kolovoza 2019.

[194] Pristupljeno 1. kolovoza 2019.

pritom na potencijalne kategorije članaka u cilju sadržavanja (dohvaćanja) unificiranih podataka, kao i na to da pripadaju tehničkome području ili su usko vezane uz područje tehnike. Projekt Istarske enciklopedije predstavlja 35 natuknica budući da je riječ o enciklopedičkom izdanju koje zastupa jednu regiju te je po tome u manjem opsegu u odnosu na navedena nacionalna izdanja. Također, projekt Hrvatskoga biografskog leksikona predstavlja 20 natuknica, budući da je riječ o djelu koje donosi samo jedan tip članaka, biografije, ali je vrijedan za usporedbu jer predstavlja temeljni biografski projekt u Hrvatskoj. Kao temelj odabira uzorka odabran je projekt Hrvatske tehničke enciklopedije, koji je poslužio kao referentno djelo vezano uz pojmove iz područja tehnike.

Kako bi se odredio uzorak svakoga izdanja koji bi ga što bolje predstavljao, učinjeno je preliminarno ispitivanje struke (područja) metalurgija Hrvatske enciklopedije kao dijela sadržaja tog općeg enciklopedičkog izdanja koje pripada području tehnike. Ispitano je 99 natuknica ukupnog opsega 2219 redaka teksta. Uvidom u faktografsko stanje članaka, svrstani su u tri kategorije, koje ih opisuju kao osobe, organizacije i opće pojmove. Tablica 7 donosi prikaz primjera natuknica po kategorijama i vrste podataka koje donose.

Tablica 7. Prikaz kategorija članaka s pripadajućim faktografskim podatcima struke metalurgija Hrvatske enciklopedije

Kategorija enciklopedičkoga članka		
osobe	organizacije	opći pojmovi
Primjeri natuknica		
Bessemer, Henry; Chloupek, Lujo; Héroult, Paul Louis Toussaint; Logomerac, Vladimir; Mamuzić, Ilija; Markotić, Ante; Martin, Piere-Émile; Thomas, Sidney Gilchrist	Hrvatsko metalurško društvo; Metalurgija; Željezara Sisak	alitiranje; alpaka; alučel; alumetiranje; aluminotermija; amalgam; amalgamiranje; austenit; bijeli lim; bijeli metal; boriranje; bronca; čelik; duralumin; elektrometalurgija i dr.
Faktografski podatci prisutni u tekstu natuknica		
mjesto rođenja; mjesto smrti; datum rođenja; datum smrti; godina rođenja; godina smrti; djelatan; godina diplomiranja; godina doktoriranja; naslov objavljenih publikacija; relevantni suvremenici; postignuti izumi; ustanove postizanja diplome i doktorata znanosti; članstvo; nagrade; osnovane organizacije; voditelji organizacija; članovi obitelji; znamenitosti	godina osnutka; sastavnice; suradničke organizacije; vlasničke organizacije; sjedište; broj zaposlenika	pioniri/doajeni područja ili struke; znamenitosti; istaknuti stručnjaci; izumitelji; godine izuma; djelatne organizacije; publikacije

Najzastupljeniji su članci općih pojmova, ali i najopširniji i ne toliko bogati unificiranim faktografskim podatcima pogodnima za stvaranje strukturiranoga mrežnog enciklopedičkog sadržaja. Faktografski najzastupljeniji su članci životopisa osoba, a pogodni su i članci organizacija koji sustavno donose određene tipove podataka. Na temelju tih saznanja određeni su udjeli pojedinih kategorija u uzorku za svako enciklopedičko djelo zastupljeno u istraživanju.

Kvantitativnom analizom sadržaja odredit će se ukupan broj naziva (entiteta) [195] u tekstu za svaki enciklopdički članak uzorka svih izdanja, koji će biti klasificirani u četiri vrste: osobe, organizacije, lokacije i vremenska obilježja. Također, označeni nazivi enciklopedičkih tekstova služit će ispitivanju alata namijenjenih automatskom prepoznavanju naziva (poglavlje 5. 2.) korištenih pri provedbi analize sadržaja, kako bi se upozorilo na mogućnosti njihova korištenja u enciklopedici u cilju veće automatizacije procesa analize sadržaja i stvaranja strukture. Rezultati kvalitativne analize sadržaja temelj su komparativne analize kojom će se pojedini nazivi opisati te klasificirati, ali i usporediti te uputiti na faktografske sličnosti i različitosti izdanja u uzorku s ciljem zaključaka oko mogućnosti donošenja objedinjavanja njihove strukture i potencijalnih učinaka takvih napora. Rezultati kvantitativnih istraživanja bit će prikazani tablično i doneseni kao prilozi (1–10), dok su rezultati kvalitativnih istraživanja prikazani grafički (histogrami).

Međusobnom usporedbom zastupljenosti određenih naziva tipiziranih jedinica više mrežnih enciklopedijskih izdanja cilj je doznati donose li različita izdanja unificirane podatke koji bi se mogli upotrijebiti za izgradnju metapodatkovne mreže unutar pojedinoga izdanja i među njima.

4.3. REZULTATI ISTRAŽIVANJA

4.3.1. ANALIZA SADRŽAJA UZORKA POJEDINIH MREŽNIH ENCIKLOPEDIČKIH DJELA

Ovo ispitivanje enciklopedičkoga sadržaja uključivalo je kvantitativne i kvalitativne tehnike analize sadržaja. Kvantitativnom analizom utvrđen je točan broj naziva po jedinicama sadržaja koji opisuju njihovu informativnost i zastupljenost određenih vrsta podataka, a ti su

[195] Vidi bilješku 80.

rezultati korišteni i pri donošenju zaključaka vezanih uz primjenjivost alata namijenjenih automatskom prepoznavanju nazivlja na enciklopedičkim tekstovima (poglavlje 5.). Kvalitativnom analizom su nazivi zabilježeni kvantitativnom analizom kategorizirani i opisani, kako bi se utvrdilo postojanje unificiranih faktografskih podataka u enciklopedičkim člancima određene kategorije.

Ukupno su utvrđena 63 tipa naziva koji su razvrstani u pet skupina (vrsta), odnosno kao osobe, organizacije, lokacije, vremenska obilježja i razno. Kategoriji natuknica osoba utvrđena su ukupno 33 tipa naziva svih vrsta, što ju čini faktografski najbogatijom s gledišta ovog ispitivanja. Svi nazivi kategorija osoba s navedenim značenjem prikazani su tablicom 8.

Tablica 8. Vrste i tipovi naziva zabilježeni ispitivanjem s pripadajućim značenjima kategorije članaka osoba

Vrste i tipovi naziva		Značenje
osobe	članovi obitelji	osobe kao članovi uže i šire obitelji
	inačice imena	varijante imena obrađivane osobe
	mentor	mentor doktorske disertacije
	relevantni suvremenici	osobe djelovanje kojih je utjecalo na stvaralaštvo obrađivane osobe
	suradnici	suradnici vezani uz područje djelovanja osobe
organizacije	baštinici imena	organizacije koje promiču djelovanje osobe
	član	organizacije kojih su obrađivane osobe bile članovi, poput nacionalnih akademija i raznih udruga
	diplomirao	organizacija postizanja diplome
	djelatan	poslodavci (tvrtke, ustanove i dr.), odnosno pružatelji radnih mjesta
	doktorirao	organizacija postizanja doktorata znanosti
	magistrirao[196]	organizacija postizanja magisterija znanosti
	osnivač	organizacije koje je osnovala obrađivana osoba
	publicirao	organizacije kao pružatelji usluga objavljivanja radova
	voditelj	organizacije koje je vodila obrađivana osoba
lokacije	država rođenja	država u kojoj je osoba rođena
	država smrti	država u kojoj je osoba umrla
	mjesto rođenja	mjesto u kojem je osoba rođena
	mjesto smrti	mjesto u kojem je osoba umrla
vremenska obilježja	datum rođenja	dan i mjesec rođenja osobe
	datum smrti	dan i mjesec smrti osobe
	diplomirao	godina postizanja diplome
	doktorirao	godina postizanja doktorata znanosti
	godina rođenja	godina rođenja osobe
	godina smrti	godina smrti osobe
	magistrirao	godina postizanja magisterija znanosti
	maturirao	godina završetka srednjoškolskog obrazovanja
	umirovljen	godina umirovljenja osobe
razno	disertacija	naslov doktorske disertacije
	izum	naziv/naslov izuma/patenta
	nagrade	naziv dodijeljene nagrade
	počasno zvanje	naziv počasnoga zvanja
	publikacije	naslov objavljenih publikacija
	znamenitosti	svojstva velike vrijednosti vezane uz obrađivanu osobu, po kojoj je ona prepoznatljiva (npr. nagrade, ulice, spomenici i dr.)

[196] Tipovi naziva naslovljeni kao magistrirao, diplomirao i doktorirao u kategoriji natuknica osoba javljaju se unutar vrsta (klasa) naziva organizacija (tekstualna vrijednost) i vremenskih obilježja (numerička vrijednost).

Utvrđena su i ukupno 23 tipa naziva svih vrsta natuknica kategorije organizacija (tablica 9) koji su proizašli iz ukupno 13973 zabilježena naziva u člancima uzorka ove kategorije.

Tablica 9. Vrste i tipovi naziva zabilježeni ispitivanjem s pripadajućim značenjima kategorije članaka organizacija

Vrste i tipovi naziva		Značenje
osobe	osnivači	osnivači obrađivanih organizacija
	voditelji	voditelji obrađivanih organizacija (direktori, predsjednici, predstojnici i dr.)
	zaposlenici	zaposlenici obrađivanih organizacija
organizacije	članstvo	druge organizacije kao pružatelji članstva obrađivanomu pojmu (npr. nacionalni savezi, udruge i dr.)
	inačice naziva	drugi nazivi obrađivanih organizacija
	kratica	prepoznatljive kratice (skraćeni nazivi) obrađivanih organizacija
	naručitelji usluga/proizvoda	organizacije korisnici usluga ili proizvoda obrađivanih pojmova
	osnovane organizacije	organizacije koje su osnovale obrađivane organizacije
	prethodnici	organizacije koje su svojim djelovanjem bile ključne za nastajanje obrađivanih organizacija, a koje su njihovi sljednici
	prijašnji nazivi	nazivi obrađivanih organizacija korišteni u određenim trenutcima njihova postojanja
	puni naziv	puni naziv obrađivanih pojmova (ako je prepoznatljiv po kraticama)
	sastavnice	organizacije u sastavu obrađivanih pojmova (npr. zavodi, podružnice, preuzete organizacije i sl.)
	sljednici	organizacije kao nasljednici pružanja usluge ili proizvodnje obrađivanih organizacija
	suradničke organizacije	organizacije koje suradnjom s obrađivanim organizacijama imaju važnu ulogu u njihovoj djelatnosti
	vlasničke organizacije	organizacije u vlasništvu kojih su obrađivane organizacije
LOK	sjedišta	mjesta obavljanja djelatnosti organizacija, upravna središta
VO	današnji naziv	godina prihvaćanja današnjega naziva organizacija
	godina osnutka	godina osnivanja organizacija
	godina prestanka rada	godina prestanka provođenja djelatnosti obrađivanih organizacija
razno	broj zaposlenika	broj djelatnika obrađivanih organizacija u određenom trenutku njihova postojanja
	nagrade	nazivi dodijeljenih nagrada obrađivanim organizacijama
	proizvodi	nazivi proizvoda obrađivanih pojmova
	publikacije	naslovi publikacija vezanih uz obrađivane pojmove

Faktografski najmanje raznolika pokazala se kategorija članaka općih pojmova. Donosi sedam tipova naziva koji su opisani tablicom 10.

Tablica 10. Vrste i tipovi naziva zabilježeni ispitivanjem s pripadajućim značenjima kategorije članaka organizacija

Vrste i tipovi naziva		Značenje
osobe	doajeni/pioniri struke	osobe ključne za početni razvoj područja ili struke
	istaknuti stručnjaci	osobe koje su ostvarile znatni utjecaj u određenom području
	izumitelji	osobe autori izuma ili patenata
ORG	djelatne organizacije	organizacije nositelji djelatnosti područja ili struke
VO	godina izuma	godina postizanja izuma vezanoga uz pojam
razno	publikacije	važnije publikacije navedene uz pojam
	znamenitosti	svojstva velike vrijednosti vezane uz pojam (npr. spomenici, izlošci, građevine i dr.)

4.3.1.1. Hrvatska tehnička enciklopedija

Ukupno je analizi sadržaja bilo podvrgnuto 50 članaka mrežnog izdanja Hrvatske tehničke enciklopedije, od čega 16 biografskih članaka, 25 članaka koji se odnose na organizacije te devet članaka općih pojmova (prilog 1).

Pojavnost najviše unificiranih faktografskih podataka utvrđena je kod natuknica koje se odnose na životopise osoba, gdje je uočeno čak 30 tipova naziva. U vrstu naziva vezanih uz prepoznavanje osobnih imena i prezimena (koji se odnose na druge osobe), kvalitativnom analizom sadržaja članaka Hrvatske tehničke enciklopedije uočeno je pet tipova podataka. Najzastupljeniji je naziv koji se odnosi na mentore doktorskih disertacija (37,5%)[197], zatim o inačicama imena i prezimena (25%), suradnicima i relevantnim suvremenicima vezanih za područje djelovanja (18,8%) te članovima obitelji (6,3%).

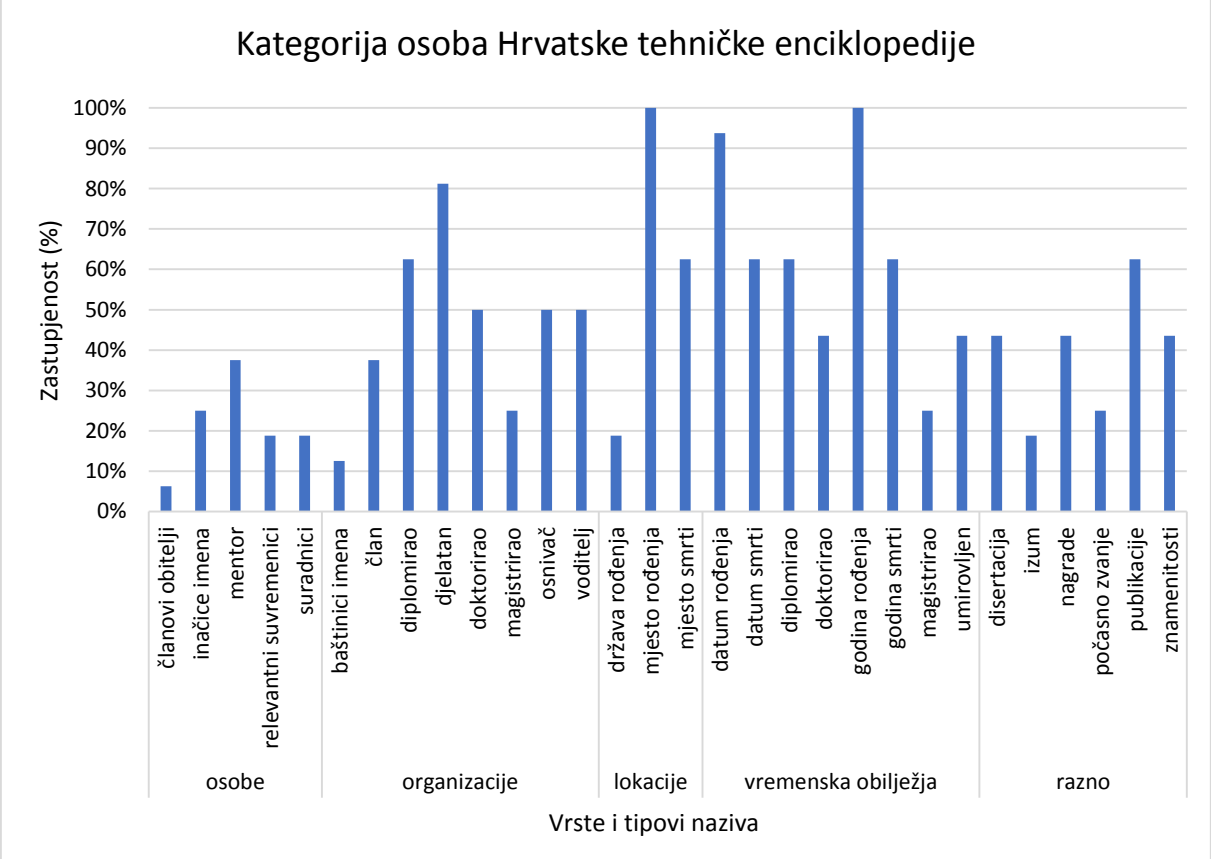
Prisutni su nazivi organizacija u kojima je osoba radila (81,5 %), diplomirala (62,5%), magistrirala (25%) i doktorirala (50%). Sustavno se još donose podatci o organizacijama u

[197] Brojčane vrijednosti navedene uz pojedini naziv ili podatak koji se sustavno javlja u tekstovima članaka označavaju zastupljenost (udio) tog podatka za pojedinu kategoriju određenog enciklopedičkoga djela uključenog u ispitivanje kao postotak.

kojima su osobe bile na voditeljskim dužnostima i koje su osnovale (50%), kojih su bile članovi (37,5%) te koje danas baštine njihovo ime (12,5%). Od vrsta entiteta vezanih uz lokacije, svi ispitivani članci donose podatke o mjestu rođenja, zatim slijede oni o mjestu smrti (62,5%) i državi u kojoj je osoba rođena (18,75%).

Unificirani podatci vezani uz vremenska obilježja također su zastupljeni u formi datuma rođenja (93,8%) i smrti (62,5%), te godinama diplomiranja (62,5%), magistriranja (25%), doktoriranja (43,8%) i umirovljenja kandidata (43,7%), odnosno osobe koju enciklopedička natuknica obrađuje.

Nazivi koji su klasificirani pod kategoriju razno su opus u publicističkome smislu, odnosno najvažnija djela koja autorstvom vežu uz osobu (62,5%), naslov objavljene disertacije (43,5%), dodijeljene nagrade (43,7%), te izumi (18,7%), znamenitosti (43,5%) i počasna zvanja (25%) koja se vežu uz promatrani pojam. Statistički prikaz vrsta i tipova naziva vezanih uz kategoriju članaka osoba Hrvatske tehničke enciklopedije prikazani su grafičkim prikazom (slika 21).

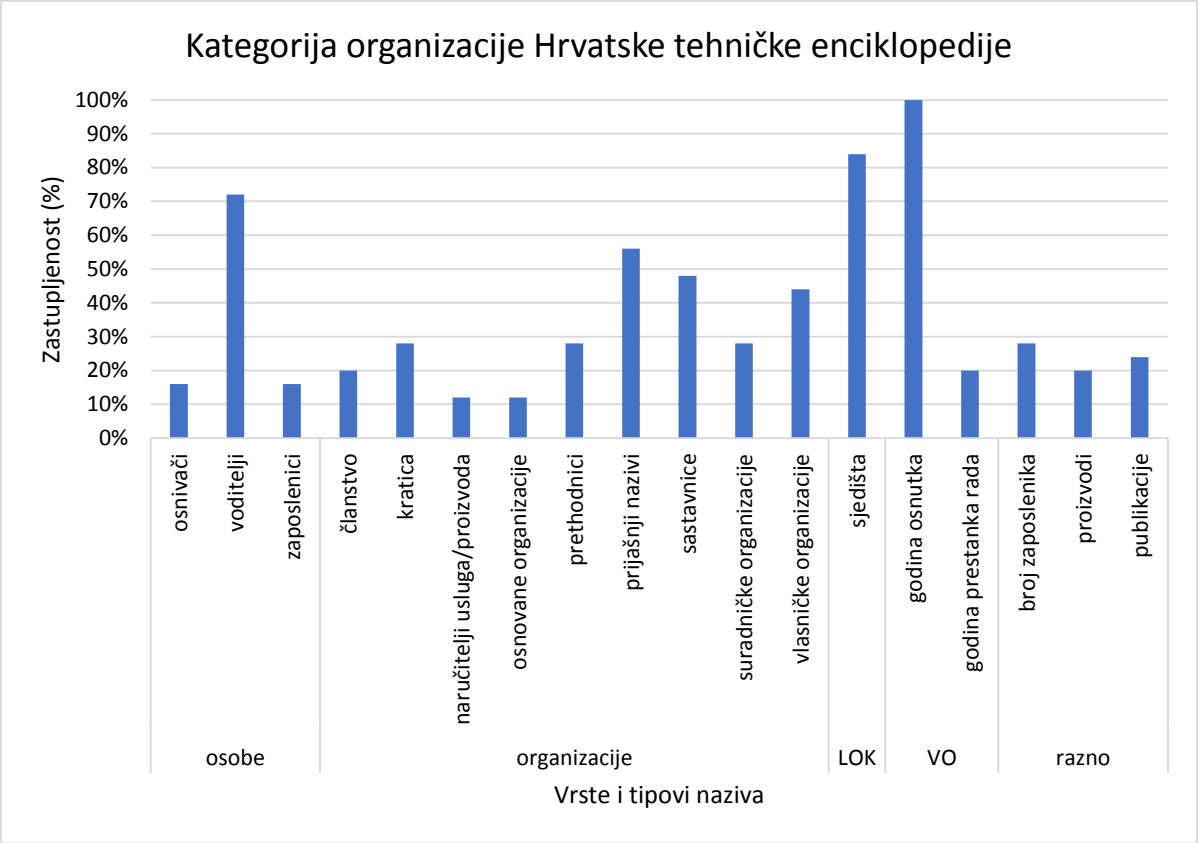


Slika 21. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Hrvatske tehničke enciklopedije

Analizom kategorije natuknica organizacija u uzorku Hrvatske tehničke enciklopedije uočeno je ukupno 18 tipova naziva. Uz prepoznavanje osoba vežu se podatci karakterizirani kao voditelji (72%), zaposlenici i osnivači (16%) organizacija.

Najveća zastupljenost uočena je kod naziva koji zastupaju organizacije kao podatke u tekstu članaka, njih čak devet tipova. Odnose se na prijašnje nazive promatranih pojmova (56%), njihove sastavnice (48%), vlasničke organizacije (44%), organizacije koje su im prethodile i s kojima su ostvarile suradnju (28%), kratice naziva (28%), članstvo u drugim organizacijama (20%), organizacije koje su osnovale i kojima su pružale usluge ili proizvode (12%).

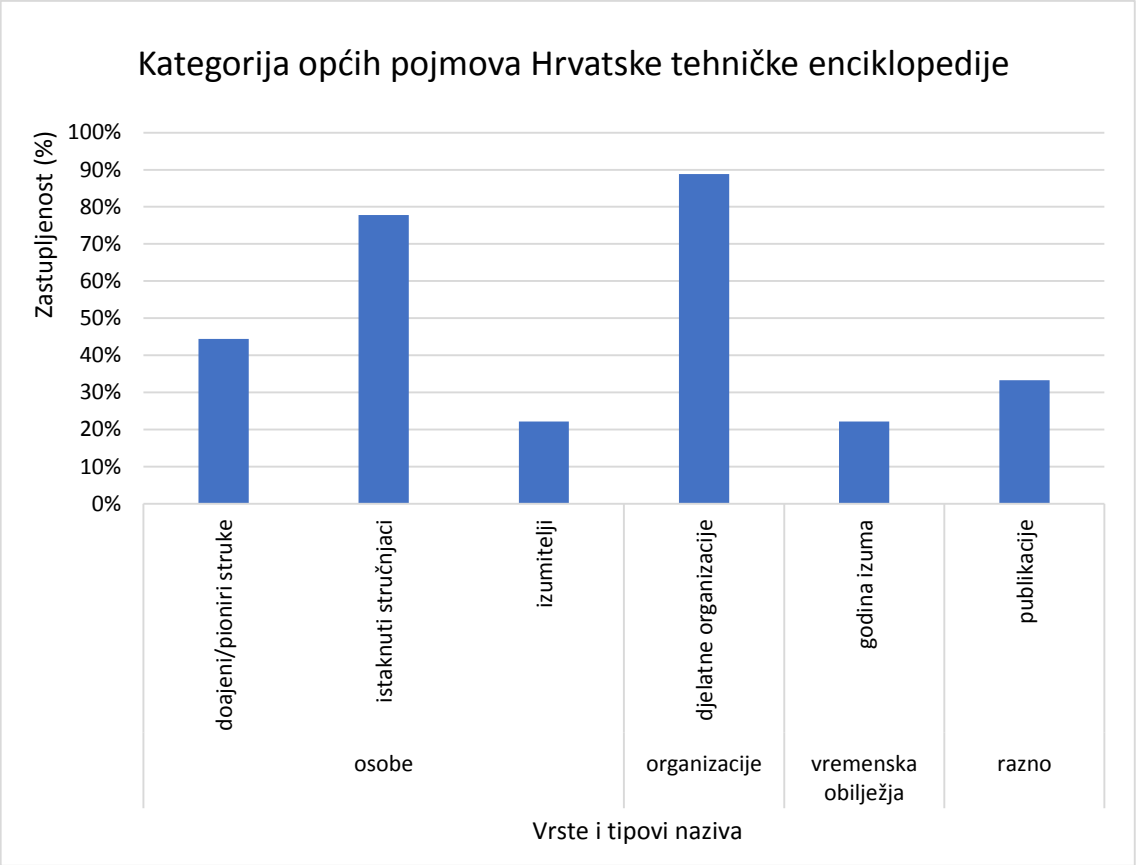
Uz lokacije kao podatke koji se mogu iskoristiti za strukturiranje sadržaja uočeno je sustavno donošenje mjesta kao sjedišta promatranoga pojma (84%). U potpunosti je zastupljen podatak o godinama osnutka organizacija koje su bile uključene u istraživanje (100 %), dok se navode i podatci o prestanku djelovanja (20%). U skupinu razno uključeni su nazivi o broju zaposlenih u nekom razdoblju djelovanja (28%), objavljenim publikacijama (24%) te proizvodima po kojima su organizacije prepoznatljive (20%). Statistički prikaz vrsta i tipova naziva vezanih uz tip članka organizacije Hrvatske tehničke enciklopedije prikazani su histogramom (slika 22).



Slika 22. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija Hrvatske tehničke enciklopedije

Nazivi koji opisuju veći broj natuknica najmanje su zastupljeni kod kategorije članaka općih pojmova. Tako vrstu naziva vezanih uz prepoznavanje imena osoba predstavljaju podatci o istaknutim stručnjacima vezanima uz promatrani pojam (77,8%), doajenima ili pionirima struke (područja) (44,4%) i izumiteljima (22,2%).

Od organizacija može se istaknuti naziv djelatnih organizacija vezanih uz pojam koji je ujedno i najzastupljeniji u ovom tipu članaka (88,9%). Vremenska obilježja zastupljena su u formi podataka o godinama izuma koji se javljaju za pojedine pojmove koji takav podatak mogu sadržavati (22,2%), a ujedno izdanje Hrvatske tehničke enciklopedije u određenoj mjeri navodi najznačajnije publikacije koje su relevantne za određeno područje (33,3%). Statistički prikaz vrsta i tipova naziva vezanih uz kategoriju članaka organizacija Hrvatske tehničke enciklopedije prikazani su grafičkim prikazom (slika 23).



Slika 23. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka općih pojmova Hrvatske tehničke enciklopedije

Kvantitativnom analizom sadržaja (prilog 1) utvrđeno je kako je ukupno ispitano 5345 redaka teksta enciklopedičkih članaka Hrvatske tehničke enciklopedije podijeljenih u tri kategorije (osobe, organizacije i opći pojmovi). Promatrajući uzorak kao cjelinu, zabilježen je ukupno 4561 naziv, od čega ih se 533 odnosi na osobe, 1132 na organizacije, 1073 na lokacije te 1823 na vremenska obilježja.

Informativnost pojedinih tipova članaka ujednačena je (osobe 0,83 naziva po retku, organizacije 0,86 te opći pojmovi 0,85), ali u ovom slučaju taj podatak može zavarati, budući da rezultati kvalitativne analize sadržaja pokazuju kako opći pojmovi nisu raznoliki vrstama naziva i smanjene su faktografske vrijednosti u odnosu na tipove članaka osoba i organizacija koji donose široki raspon vrsta podataka koji mogu biti semantički definirani i poslužiti kao sredstvo strukturiranja enciklopedičkoga sadržaja.

Također, usporedbom kvalitativne i kvantitativne analize sadržaja uzorka Hrvatske tehničke enciklopedije, smanjeni broj naziva osoba u svim trima tipovima članaka očituje se u manjoj zastupljenosti tih naziva za sve kategorije, poglavito za tip članaka osobe. Vidljivo je i

kako povećani broj naziva vezanih uz lokacije i vremenska obilježja ne utječe na raznolikost vrsta istih, što najviše vrijedi za tip članaka organizacije.

4.3.1.2. Hrvatska enciklopedija

Analizom 50 članaka općeg izdanja Hrvatske enciklopedije vidljive su razlike u odnosu na rezultate analize Hrvatske tehničke enciklopedije. Također je bilo uključeno 16 biografskih natuknica, 25 onih koje se odnose na pojmove organizacija te devet članaka općega tipa.

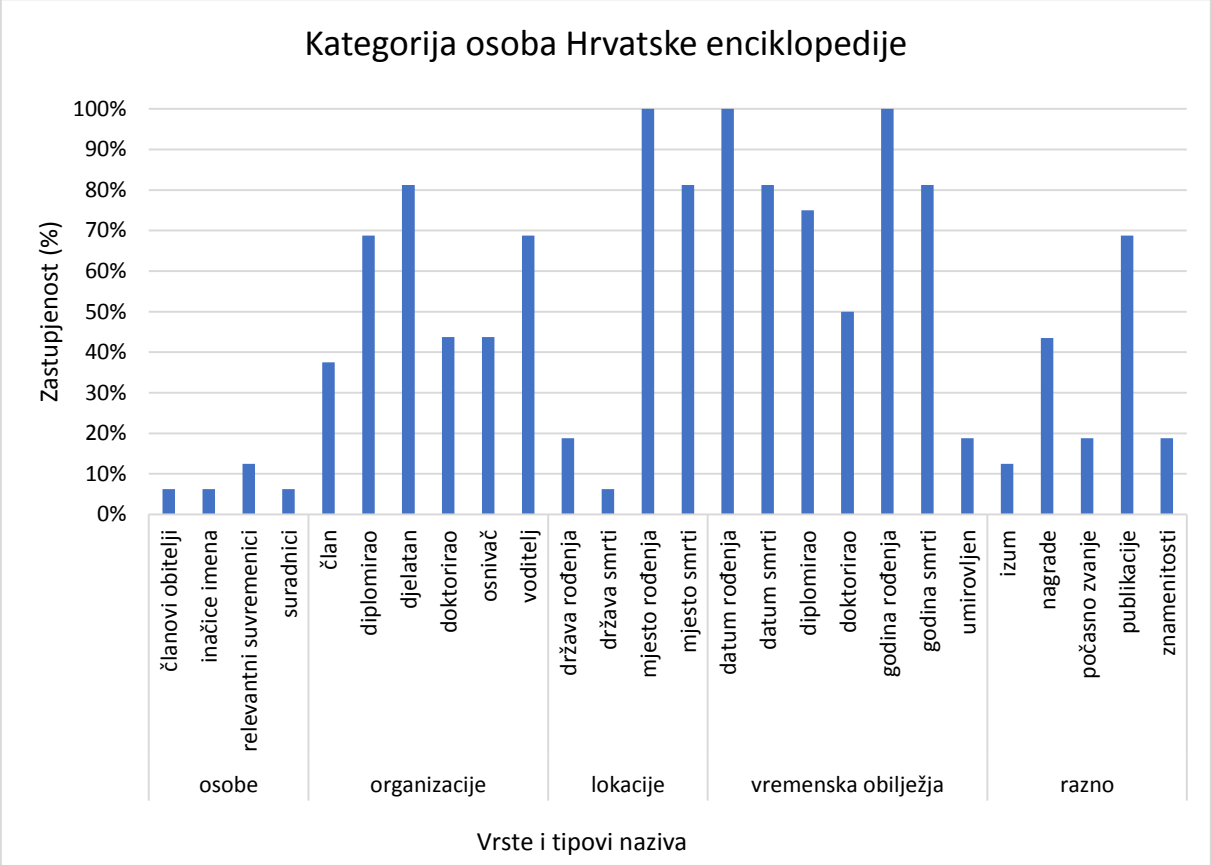
Kako se ne radi o specijaliziranome enciklopedičkom izdanju posvećenome tehnici, ne pojavljuju se nazivi poput mentora i naslova disertacije, a posebno mala zastupljenost uočena je kod pojavljivanja naziva vezanih uz osobe, i to ne samo u biografskim člancima. Životopisi osoba, kojima je analizom dodijeljeno ukupno 26 tipova naziva, rijetko donose nazive koji se odnose na osobe, tako da su zabilježeni oni poput članova obitelji, suradnika, inačica imena koje se odnose na promatrane osobe (6,2%) i relevantnih suvremenika vezanih uz pojam (12,5%).

Zabilježeno je šest naziva koji se pojavljuju u tekstu članaka vezanih uz organizacije, a odnose se na radno aktivno djelovanje osoba (81,3%), vođenje organizacija (68,8%), ustanove na kojima su diplomirale (68,8%) i doktorirale (43,8%), kojih su osnivači (43,8%) i članovi (37,5%). Standardizirani podatci koji se pojavljuju u velikoj mjeri u uzorku Hrvatske enciklopedije vezani su uz mjesta rođenja (100%) i smrti (81,3%) osoba ako je taj podatak potrebno donijeti, a u nekim slučajevima, ako je riječ o mjestima koja su izvan hrvatskoga geografskog korpusa, uz navedene podatke donose se i informacije o državama u kojima su osobe rođene (18,8%) ili umrle (6,3%).

Najveću zastupljenost pokazuju nazivi vezani uz vremenska obilježja, koji su u ovom uzorku zastupljeni sustavno. To se ponajprije odnosi na datume i godine rođenja osoba (100%), zatim na datume i godine smrti osoba (81,3%), te na podatke o godinama postizanja diploma (75%) i doktorata znanosti (50%). U manjem broju ispitanih članaka donosi se podatak o godini umirovljenja (18,8%).

Pod skupinu naziva razno klasificirani su nazivi vezani uz publicistički opus osoba (68,8%), navođenje dodijeljenih nagrada za razne društvene zasluge (43,8%), ostvarenih izuma (12,5%), te počasnih zvanja i znamenitosti koje predstavljaju ostavštinu promatranih osoba i

podsjećaju na njihovo djelovanje (18,8%). Statistički prikaz vrsta i tipova naziva vezanih uz kategoriju članaka osoba Hrvatske enciklopedije prikazani su grafičkim prikazom (slika 24).



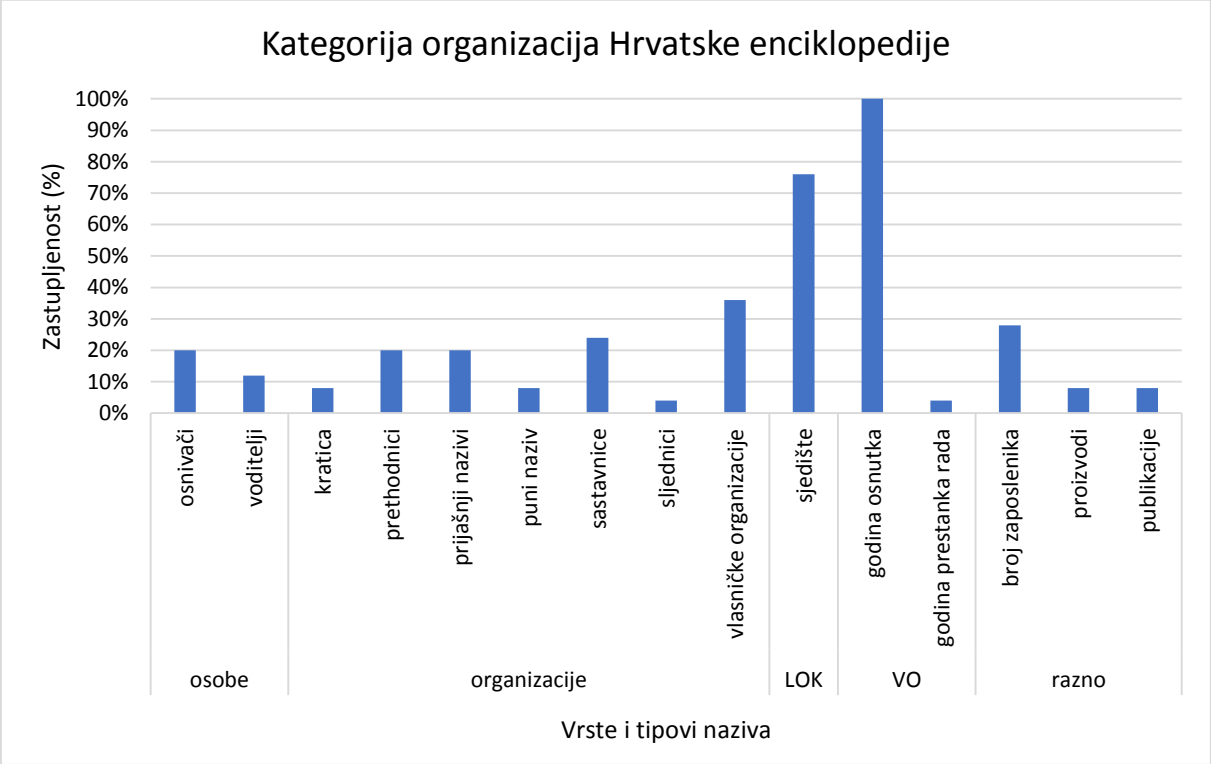
Slika 24. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Hrvatske enciklopedije

Analizom kategorije natuknica organizacija u uzorku Hrvatske enciklopedije uočeno je ukupno 16 tipova naziva. Oni vezani uz prepoznavanje osoba donose podatke vezane uz osnivače (20%) i voditelje (12%).

Sažetiji pristup općeg enciklopedičkog izdanja rezultirao je smanjenom zastupljenošću naziva vezanih uz organizacije, iako je prisutno sedam tipova podataka. Odnose se na vlasničke organizacije (36%), sastavnice promatranih organizacija (24%), prijašnje nazive promatranih organizacija i organizacije koje su im prethodile (20%), zastupljeni su nazivi koji se pojavljuju kao kratice i kao puni nazivi promatranih pojmova (8%) te kao sljednici koji nastavljaju njihovu djelatnost (4%).

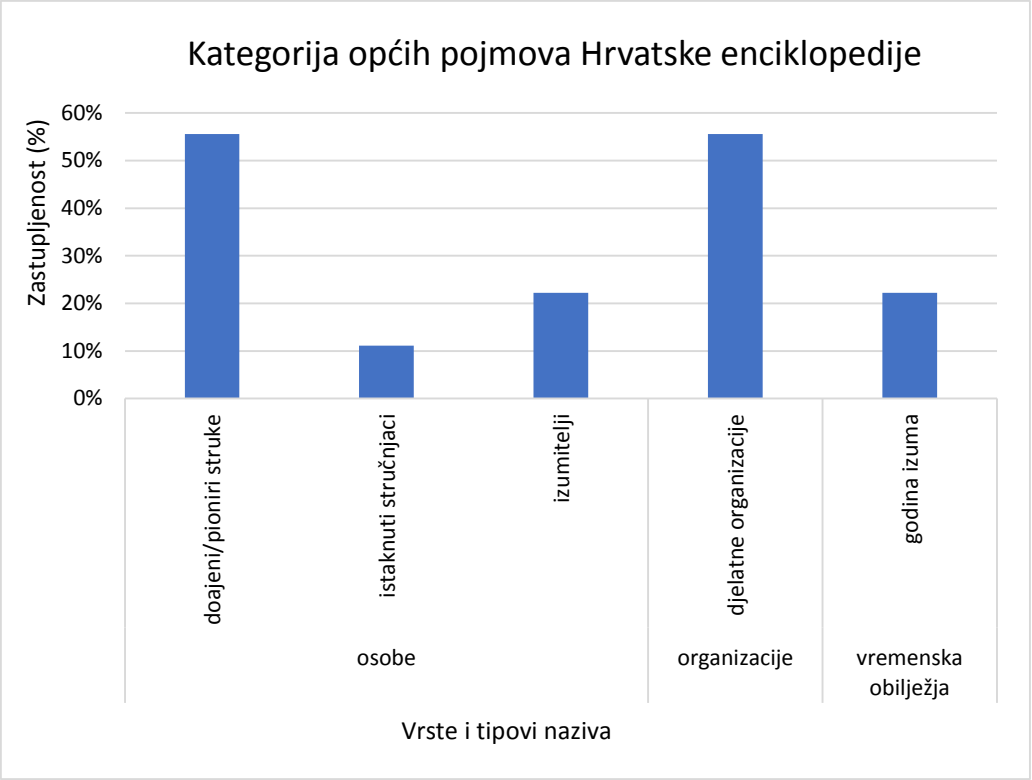
Mjesto koje se odnosi na sjedište zastupljeno je s visokih 76%. U potpunosti je zastupljen podatak o godini osnutka organizacija (100%), uz mogućnost pojavljivanja i godine

prestanka djelovanja (4%). U skupinu razno uključeni nazivi su o broju zaposlenih u nekom razdoblju djelovanja (28%), te objavljene publikacije i proizvodi (8%). Statistički prikaz vrsta i tipova naziva vezanih uz kategoriju članaka organizacija Hrvatske enciklopedije prikazani su grafičkim prikazom (slika 25).



Slika 25. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija Hrvatske enciklopedije

Nazivi vezani uz prepoznavanje imena osoba ispitanih općih pojmova Hrvatske enciklopedije predstavljaju podatke o doajenima ili pionirima struke (područja) (55,5%), izumiteljima (22,2%) i istaknutim stručnjacima u vezi s promatranim pojmom (11,1%). Organizacije se opisuju kao djelatne, odnosno one koje nose djelatnost promatranoga područja (55,5%), dok se od vremenskih obilježja izdvojiti može samo podatak o godinama postignutih izuma ili patenata ako je riječ o pojmu koji obrađuje takvu tematiku. Statistički prikaz vrsta i tipova naziva vezanih uz članke općega tipa Hrvatske enciklopedije prikazani su grafičkim prikazom (slika 26).



Slika 26. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka općih pojmova Hrvatske enciklopedije

Od ispitanih 50 članaka Hrvatske enciklopedije obrađeno je ukupno 2137 redaka enciklopedičkoga teksta i uočeno je 1487 naziva (prilog 2). Unatoč smanjenom opsegu članaka i broju naziva, vrste i tipovi naziva nisu izostali, što upućuje na to da određeni tipovi podatka i njihova pojava predstavljaju svojevrsni standard enciklopedičkih tekstova ili sadržaja. Ipak, njihovom smanjenom zastupljenošću smanjuje se i mogućnost strukturiranja mrežnog izdanja Hrvatske enciklopedije. U ovom je izdanju vidljivo kako se vrsta naziva osoba kod većine članaka osoba i organizacija ne pojavljuje ili jedva pojavljuje, što se odrazilo i na rezultate kvalitativne analize sadržaja.

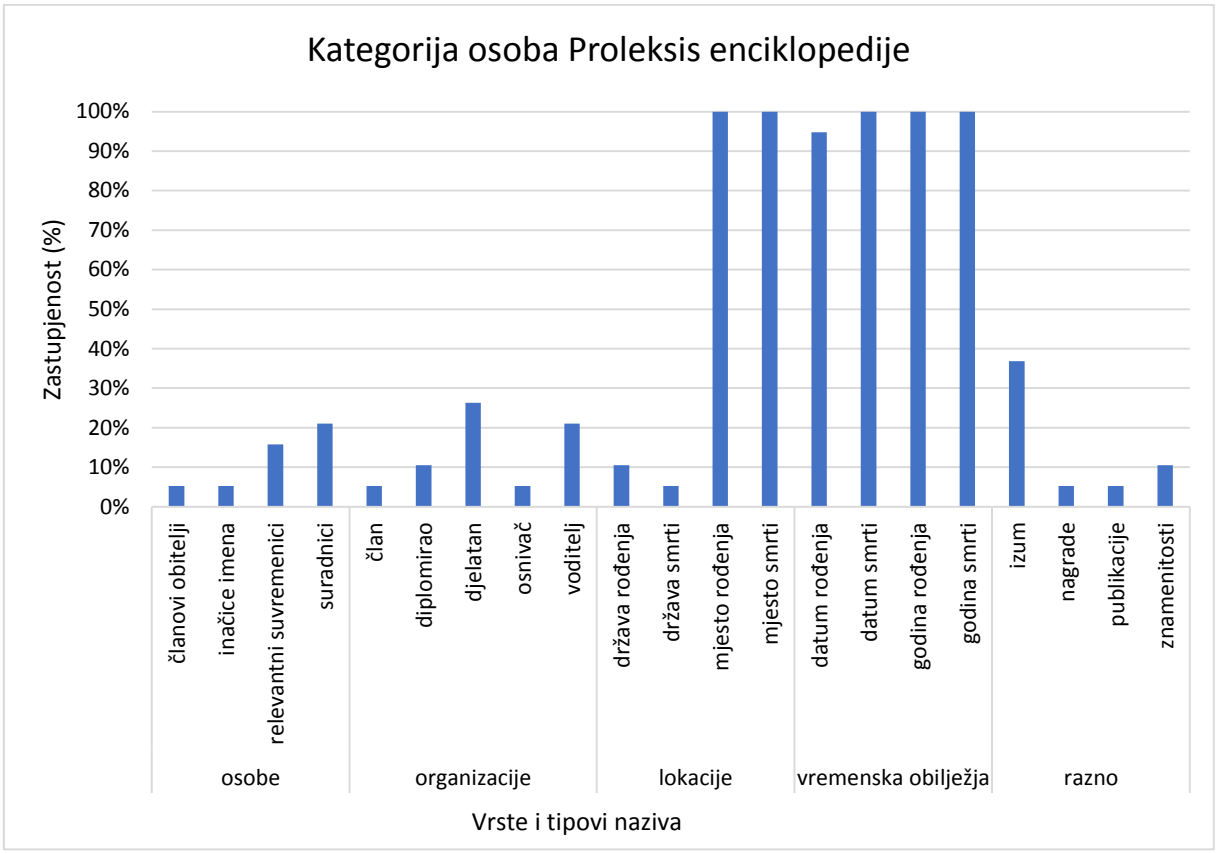
4.3.1.3. Proleksis enciklopedija

Kvalitativnom i kvantitativnom analizom sadržaja ispitano je također 50 članaka mrežnog izdanja Proleksis enciklopedije. Pri sastavljanju uzorka zamijećen je povećani broj općih pojmova u dijelu sadržaja vezanoga uz tehniku i tehničke znanosti, tako da uzorak koji je ispitan sačinjava povećani broj članaka toga tipa. Ispitano je ukupno 20 članaka općih pojmova, 19 biografskih članaka osoba te 11 članaka sadržaj kojih se odnosi na organizacije.

Kvantitativnom analizom sadržaja koja donosi uvid u opseg članaka i broj naziva osoba, organizacija, lokacija i vremenskih obilježja (prilog 3), vidljivo je da je riječ o veoma sažetom i jezgrovitom izdanju, koje se svojom formom može protumačiti kao leksikon, odnosno enciklopedijski rječnik koji pretežno donosi samo najopćenitije informacije u faktografskom smislu, a nerijetko samo definicije pojmova.

Raznolikost tipovima naziva kategorije članaka osoba upućuje na to kako broj tipova naziva ne zaostaje za drugim izdanjima u uzorku, međutim, ako izuzmemo nazive koji se odnose na lokacije i vremenska obilježja i koji se pojavljuju u gotovo svakoj natuknici uzorka Proleksis enciklopedije, poput mjesta rođenja i smrti, te podataka o datumima i godinama rođenja i smrti osoba, zastupljenost ostalih podataka je znatno manja nego u drugim izdanjima i oni se pojavljuju kod istaknutijih biografskih članaka.

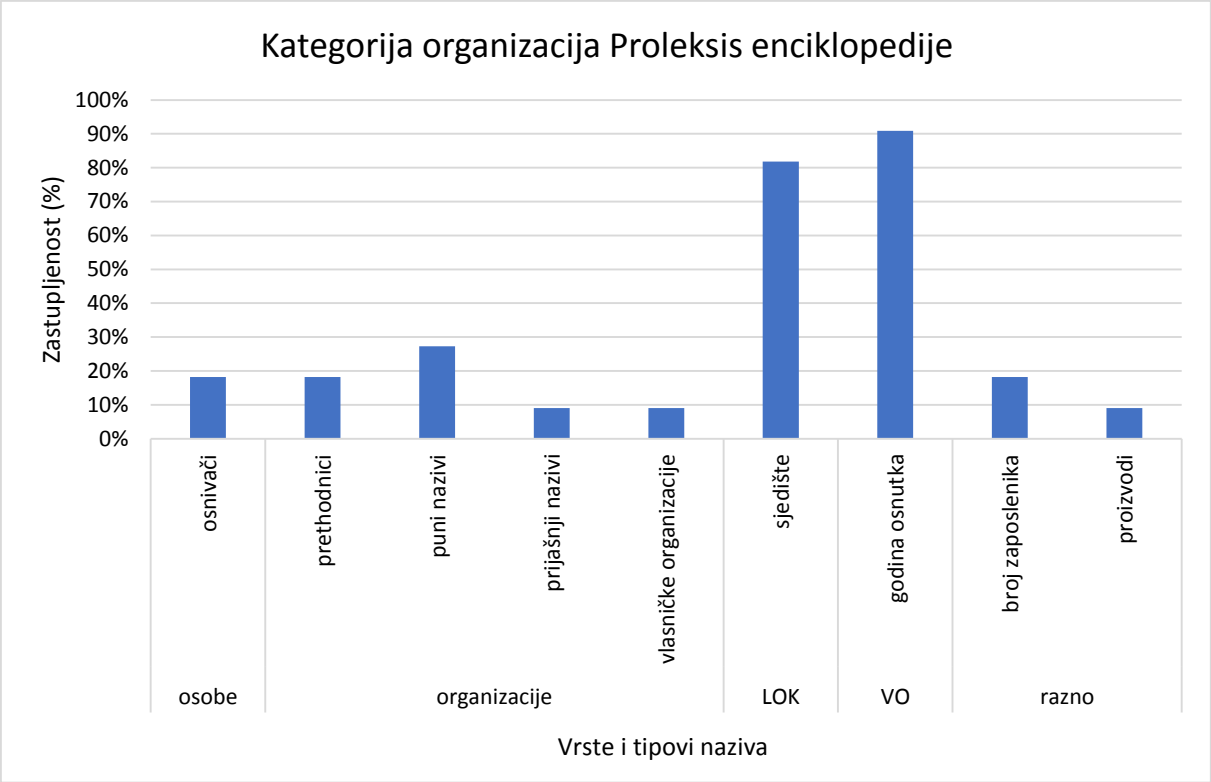
Te nazive čine podatci o članovima obitelji (5,3%), relevantnim suvremenicima (15,8%), suradnicima (21,1%), inačice imena promatranih osoba (5,3%), podatci o radno aktivnom djelovanju unutar raznih organizacija (26,3%), njihovu osnivanju (5,3%) i vođenju (21,1%), članstvu (5,3%) te institucijama na kojima su diplomirale (10,5%). Od lokacija pojavljuju se nazivi kao podatci o državama rođenja (10,5%) i smrti (5,3%) osoba, ali i nazivi dodijeljenih nagrada (5,3%), postignutih izuma (36,8%), objavljenih publikacija (5,3%) te znamenitostima (10,5%) kao podsjetnicima na djelovanje promatranih osoba. Statistički prikaz vrsta i tipova naziva vezanih uz životopise osoba Proleksis enciklopedije prikazani su grafičkim prikazom (slika 27).



Slika 27. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Proleksis enciklopedije

Smanjeni broj naziva uz smanjenu zastupljenost zabilježen je kod kategorije natuknica organizacija. Od naziva koji se odnose na lokacije i vremenska obilježja, zabilježeni su sjedište (81,8%) i godina osnutka (90,9%) promatranih organizacija. Te vrste podataka u ovom tipu enciklopedijskoga članka pokazuju visoki stupanj standardiziranosti. Ostali tipovi naziva svih pet vrsta koje su zastupljene ovim istraživanjem pojavljuju se u manjoj mjeri i sa smanjenom zastupljenošću.

Prisutni su podatci o osobama kao osnivačima promatranih organizacija (18,2%), organizacijama kao prethodnicima (18,2%) i vlasnicima (9,1%) promatranih pojmova, prijašnji nazivi promatranih organizacija (9,1%) te njihovi puni nazivi ako su navedeni kraticom (27,3%), a od vrste naziva koji se ne mogu izričito svrstati u neku od četiriju glavnih nositelja naziva ističu se podatci o broju zaposlenika (18,2%) i o proizvodima (9,1%) obrađivanih organizacija. Statistički prikaz vrsta i tipova naziva vezanih uz tip članka organizacija Proleksis enciklopedije prikazani su grafičkim prikazom (slika 28).



Slika 28. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija Proleksis enciklopedije

Analizom sadržaja kategorije članaka opisane kao opći pojmovi, utvrđeno je kako oskudijevaju nazivima i utvrđene su tri vrste podataka koje su pojavljuju. Najzastupljeniji naziv predstavljaju izumitelji (25%), uz njega dolazi i podatak o godinama postizanja tih izuma (15%), a osobe su također karakterizirane kao doajeni ili pioniri određenih područja ili struka (15%).

Pogledom na rezultate kvantitativne analize sadržaja mrežnoga izdanja Proleksis enciklopedije (prilog 3) vidljivo je kako su članci tog izdanja opsegom najkraći. Ukupno je ispitano 413 redaka enciklopedičkoga teksta, što je manje od opsega nekih članaka drugih izdanja u uzorku. Ukupno je ispitano samo 419 naziva. Znatan broj članaka ne sadrži ni jedan naziv od svih vrsta koje su bile ispitivane, a opsegom rijetko prelaze deset redaka teksta. Po tome Proleksis enciklopedija, što se tiče članaka koji svojom tematikom spadaju u područje tehnike i tehničkih znanosti, u dobroj mjeri predstavlja rječnik pojmova s pripadajućim definicijama.

Vrijednost ovako opsežnog projekta je u njegovu opsegu kad je riječ o broju članaka, odnosno bogatom abecedariju. Iako je strukturiranje na temelju naziva koji su ispitivani veoma ograničeno i velik bi broj članaka bio oskudne strukture, problem kod ovako koncipiranog

izdanja rješiv je postizanjem interoperabilnosti s drugim izdanjima koja te podatke posjeduju. Na taj način mrežno izdanje Proleksis enciklopedije, kao i Hrvatske enciklopedije kod koje je taj problem manje izražen, postali bi enciklopedički projekti bogati strukturom i velikim brojem informacija u računalno čitljivom obliku (metapodataka) zbog svoga širokog područja interesa.

4.3.1.4. Istarska enciklopedija

Analizom sadržaja ispitano je 35 članaka Istarske enciklopedije, od čega 15 članaka koji obrađuju organizacije, 11 životopisa osoba te devet natuknica općih pojmova. Smanjeni udio ovoga izdanja u cjelokupnom uzorku rezultat je suženoga područja interesa koje obrađuje ovo enciklopedičko djelo, koje samim time u konačnici sadrži manji broj natuknica. Unatoč tomu, sadržava sve tipove enciklopedičkih članaka koji su ispitivani te se u tom smislu može promatrati i kao bilo koje drugo opće ili specijalizirano izdanje koje je ovim istraživanjem zahvaćeno, odnosno njezina vrijednost kao potencijalno strukturiranoga i interoperabilnoga mrežnog enciklopedičkog projekta nije smanjena. Po broju naziva kreće se u granicama rezultata ostalih izdanja uključenih u ovo istraživanje, izuzevši Hrvatsku tehničku enciklopediju kao specijalizirano izdanje za povijest i dosege tehnike te Hrvatskoga biografskog leksikona koji donosi najviše podataka vezanih uza svoju specijalnost, biografije osoba.

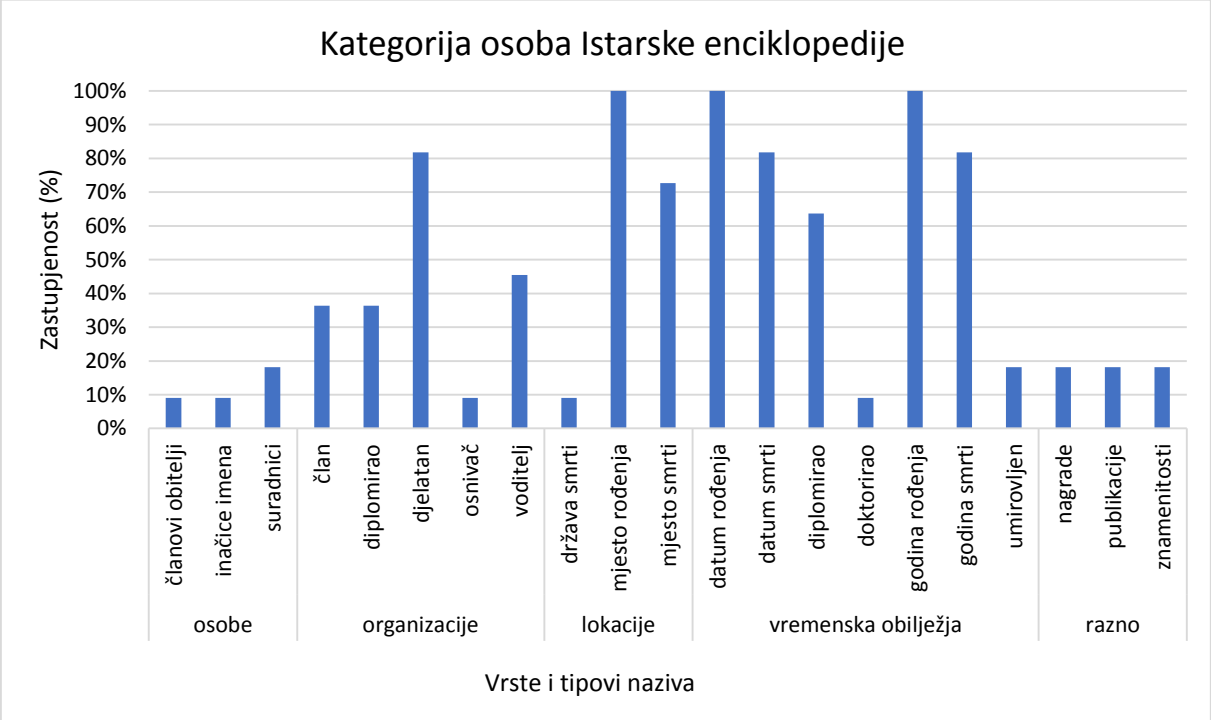
Kad je riječ o kategoriji natuknica osoba, zabilježen je ukupno 21 naziv. Nazivi koji se pojavljuju kao osobe koje su relevantne uz promatrani pojam smanjene su zastupljenosti. Zabilježeni su članovi obitelji (9,1%) i suradnici (18,2%) osoba te inačice njihova imena (9,1%).

Veza između obrađivanih osoba i organizacija kojih se znatan broj također nalazi u sadržaju izdanja Istarske enciklopedije izražena je, što se najbolje vidi u navođenju djelatnih organizacija vezanih uz obrađivane osobe (81,8%), ali i u navođenju organizacija koje su vodili (45,5%) te institucija kojih su bili članovi i na kojim su diplomirali (36,4%). U manjoj je mjeri zabilježen podatak o organizacijama koje su osnovale promatrane osobe (9,1%).

Jedna od konstanti koja se pojavljuje u rezultatima svih ispitanih izdanja, pa tako i Istarske enciklopedije, a u vezi s kategorijom članaka osoba jest visoka zabilježenost naziva koji spadaju u skupine lokacija i vremenskih obilježja. Kad je riječ o lokacijama to su mjesta rođenja (100%) i mjesta smrti ako se ne radi o živućim osobama (72,7%), a kod vremenskih obilježja datumi rođenja (100%) i smrti (81,8%) te godine rođenja (100%) i smrti (81,2%).

Povećana zastupljenost zabilježena je kod naziva koji donosi podatak o godini diplomiranja (63,6%).

U manjoj mjeri pojavljuju se još podatci o nazivu države u kojoj je osoba umrla (9,1%), godini postizanja doktorata znanosti (9,1%) i umirovljenja (18,2%). Od skupine naziva koji se klasificiraju pod skupinu razno zabilježena je prisutnost podataka o dodijeljenim nagradama, objavljenim publikacijama te znamenitostima vezanima uz djelovanje obrađivanih osoba (18,2%). Statistički prikaz vrsta i tipova naziva vezanih uz životopise osoba Istarske enciklopedije prikazani su grafičkim prikazom (slika 29).

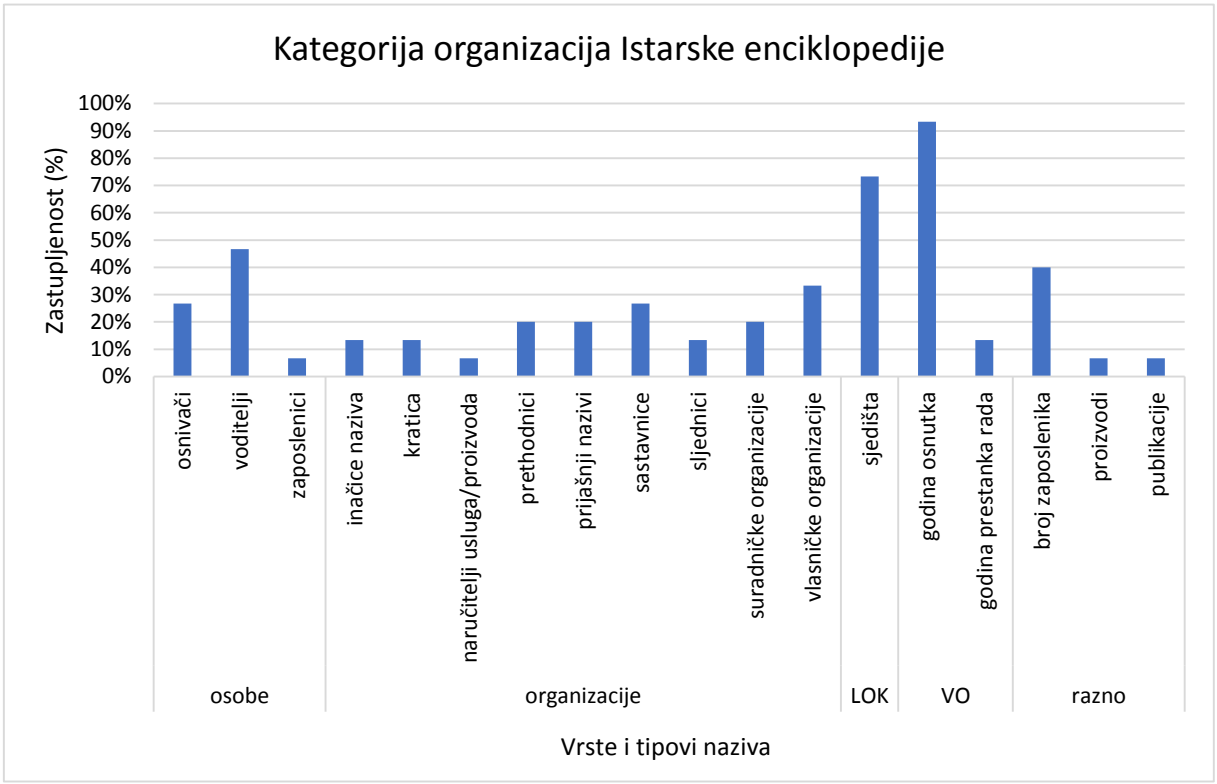


Slika 29. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Istarske enciklopedije

Kvalitativnom analizom sadržaja tekstova članaka kategorije organizacija zabilježeno je ukupno 18 tipova naziva koji se pojavljuju u određenoj mjeri i pogodni su za stvaranje strukture. Povećane su zastupljenost podataka o voditeljima (46,7%) i osnivačima (26,7%) obrađivanih organizacija, te je zabilježen i podatak o njihovim zaposlenicima (6,7%). Vrsta naziva vezana uz druge organizacije koje se javljaju u tekstovima članaka ovoga tipa smanjene je zastupljenosti, iako ih je zabilježeno devet. Uzorak ovoga izdanja jedini donosi podatak o inačicama naziva organizacija (13,3%), a uz to navode se još kratice obrađivanih organizacija (13,3%), organizacije kao njihovi prethodnici (20%) i sljednici (13,3%), naručitelji usluga ili

proizvoda (6,7%), prijašnji nazivi (20%), sastavnice (26,7%) te suradničke (20%) i vlasničke organizacije (33,3%).

Najveća zastupljenost podataka zabilježena je kod entiteta vezanih uz lokacije i vremenska obilježja. Iako ova kategorija enciklopedičkoga članka Istarske enciklopedije donosi uglavnom dva tipa podataka, oni se redovito u velikoj mjeri javljaju u uzorcima svih ispitivanih izdanja. To se odnosi na sjedišta organizacija (73,3%) i godine osnutka (93,3%). Ako poduzeće više nije aktivno, neki od članaka u uzorku donose podatak o godini prestanka rada (13,3%), pojavljuje se podatak o broju zaposlenika u različitim razdobljima (40%), a ako djelatnost organizacije to omogućuje, prisutni su podatci o proizvodima i objavljenim publikacijama (6,7%) koje organizacije proizvode, odnosno objavljuju. Statistički prikaz vrsta i tipova naziva vezanih uz kategoriju članka organizacija Istarske enciklopedije prikazani su grafičkim prikazom (slika 30).



Slika 30. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija Istarske enciklopedije

Utvrđena su tri tipa podataka koji se pojavljuju u kategoriji članaka općih pojmova Istarske enciklopedije. Oni se odnose na istaknute stručnjake vezane uz područje ili struku

(33,3%), organizacije koje su nosile djelatnost područja ili su još uvijek aktivne (55,5%) te znamenitosti kao dijelove materijalne ili nematerijalne baštine velike vrijednosti (22,2%).

Ukupno je ispitano 1517 redaka enciklopedičkoga teksta mrežnog izdanja Istarske enciklopedije (prilog 4) i utvrđeno je postojanje 1648 naziva. Važno je napomenuti kako je riječ o smanjenom uzorku, odnosno manjem broju ispitanih članaka. Međutim, većina naziva raspoređena je u dvije skupine (vrste), odnosno karakterizirana kao lokacije ili vremenska obilježja, njih čak 1303. To je razlog smanjenog udjela pojedinih tipova podataka ostalih dviju vrsta naziva, pogotovo u kategorijama članaka osoba i organizacija.

Unatoč tomu, Istarska enciklopedija raznolika je tipovima podataka, a ovo istraživanje pokazalo je kako znatan broj naziva (podataka) donosi jednoliko ili s većom zastupljenošću u odnosu na druga izdanja uključena u istraživanje.

4.3.1.5. Hrvatski biografski leksikon

Hrvatski biografski leksikon nacionalni je biobibliografski projekt koji donosi životopise Hrvata i pripadnika drugih naroda koji su sudjelovali u razvoju hrvatskih zemalja. Budući da donosi jednu kategoriju enciklopedičkih članaka, uzorak namijenjen analizi sadržaja obuhvaćao je 20 članaka. Ukupno je zabilježeno 29 tipova naziva svih četiriju vrsta. Osim raznolikosti tipova unificiranih podataka, povišena je i razina njihove zastupljenosti, što zapravo projekt Hrvatskog biografskog leksikona čini bazom biografskih podataka osoba vezanih uz hrvatski korpus.

Podatci koji se odnose na donošenje spoznaja o drugim osobama i njihovu vezu s promatranim pojmovima zastupljeni su kao i u drugim izdanjima u uzorku, a njihova zastupljenost također je na razini drugih izdanja. Sukladno tomu, pojavljuju se nazivi u tekstu članaka koji označavaju članove obitelji obrađivanih osoba (40%), njihove suradnike (15%) i relevantne suvremenike (10%) te inačice imena po kojima su također prepoznatljivi (25%).

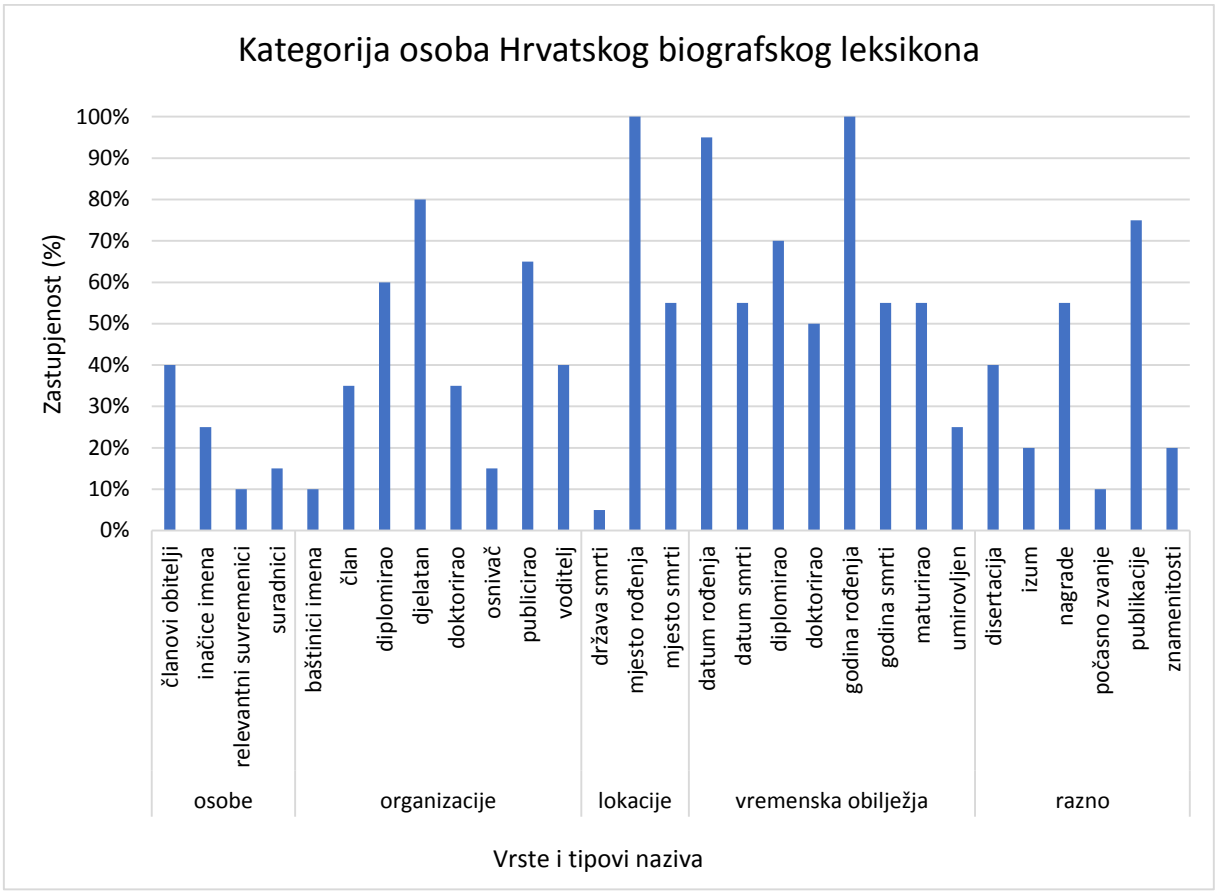
Veći broj tipova podataka s povećanom zastupljenošću zabilježen je kod ostalih triju vrsta naziva koji se ispituju ovim istraživanjem. Naziva koji se odnose na pojavnost podataka o organizacijama zabilježeno je osam. Znatan broj članaka Hrvatskoga biografskog leksikona sadrži podatke o organizacijama kao poslodavcima (80%), institucijama na kojima su osobe diplomirale (60%) i doktorirale (35%), kojih su bili voditelji (40%), članovi (35%) i osnivači (15%). U manjoj mjeri donosi se podatak o institucijama kao baštinicima imena promatranih

osoba (10%), iako je takvu vrstu informacije moguće donijeti samo za određen broj osoba, poglavito velikana, te se podatak o zastupljenosti te informacije treba uzeti s odmakom. Ovo izdanje jedino sustavno donosi podatak o organizacijama koje su obrađivanim osobama služile kao mjesto publiciranja, uglavnom znanstvenih radova (65%).

Lokacije kao nazivi predstavljaju standardizirane unificirane faktografske podatke vezane uz enciklopedička djela. Odnose se na mjesto rođenja (100%) i smrti (55%). Određen broj članaka odnosi se na živuće osobe, ali članci mrežnog izdanja Hrvatskoga biografskog leksikona predstavljaju izvornike tiskanoga izdanja, tako da se određen dio podataka nije ažurirao, odnosno članci se nisu dopunjavali. To se odnosi i na nazive koji predstavljaju vremenska obilježja. Ovo istraživanje pokazalo je da je ta vrsta podataka najtemeljitije zastupljena i da predstavlja standard ovoga izdanja, od kojega se ne odstupa. Gotovo svi podatci zabilježeni u tekstovima imaju pojavnost veću od 50% u uzorku. Zastupljeni su podatci o datumima rođenja (95%) i smrti (55%), godinama rođenja (100%) i smrti (55%), godinama postizanja diplome (70%) i doktorata znanosti (50%), umirovljenja (25%).

Zanimljivost vezana uz ispitivana mrežna izdanja je kako jedini donosi vremensko obilježje o završetku srednjoškolskog obrazovanja osoba (55%). Kako projekt Hrvatskoga biografskog leksikona sadrži velik broj biografskih članaka, do sada više od 11 000, a znatan broj tek slijedi, jasno je kako ova vrsta podataka koju ovaj projekt sustavno donosi predstavlja njegovu iznimnu vrijednost. Tomu pridonosi i činjenica kako neke pojmove obrađuje prvi put i stvara sadržaj koji do tada nije bio dostupan u sređenom obliku i predstavljen javnosti.

Dodatne informacije o obrađivanim osobama predstavljaju naslov doktorske disertacije (40%), postignuti izumi ili patenti (20%), dodijeljene nagrade (55%) i počasna zvanja (10%) te znamenitosti koje podsjećaju na djelovanje osoba (20%). Dodatnu vrijednost ovoga projekta predstavlja i sustavno donošenje objavljenih djela (75%), odnosno cjelokupnog opusa publicistički aktivnih osoba. Uz to, svaki članak nadograđen je detaljnim popisom literature vezanim uz pojam, koji obuhvaća najrelevantniju dostupnu građu koja omogućuje dodatno informiranje o pojmu i mogući nastavak istraživanja. Statistički prikaz vrsta i tipova naziva vezanih uz životopise osoba Hrvatskoga biografskog leksikona prikazani su grafičkim prikazom (slika 31).



Slika 31. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Hrvatskoga biografskog leksikona

Ukupno je ispitano 20 članaka mrežnog izdanja Hrvatskoga biografskog leksikona, 1438 redaka enciklopedičkoga teksta, te zabilježeno ukupno 2169 entiteta. To je ukupno 1,51 entitet po retku, što ovo izdanje čini najgušćim informacijama, odnosno najinformativnijim. To se i odrazilo na povećanu zastupljenost unificiranih faktografskih podataka koji su karakterizirani kao takvi. Velika je zastupljenost naziva koji označavaju vremenska obilježja (1029), što se odrazilo i na povoljne rezultate kvalitativne analize sadržaja te vrste podataka.

Projekt Hrvatskoga biografskog leksikona predstavlja iznimno povoljno mrežno enciklopedičko izdanje za stvaranje strukture metodom označavanja unificiranih faktografskih podataka u cilju stvaranja bogate baze podataka koja bi bila temelj funkcionalnosti i razvoja mrežne aplikacije na korist korisnika mrežnog izdanja.

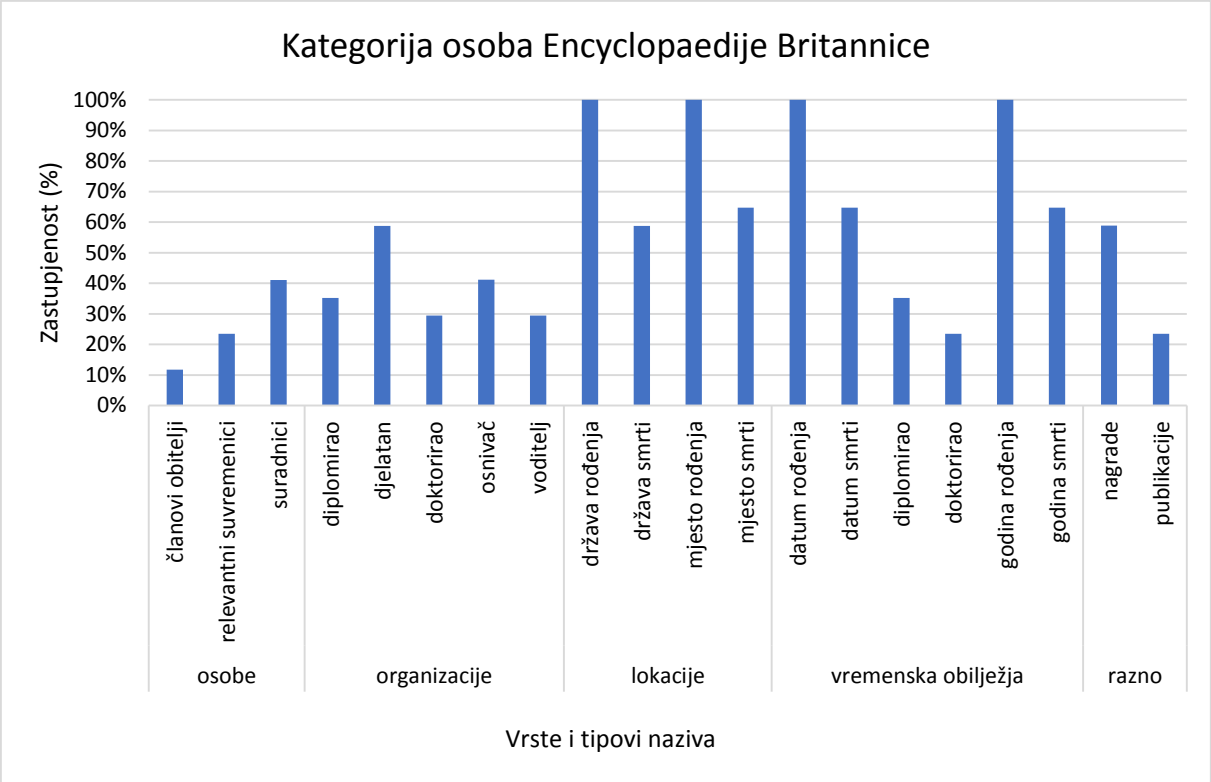
4.3.1.6. Encyclopaedia Britannica

Ukupno je u cjelokupnoj analizi sadržaja podvrgnuto 50 natuknica Encyclopaedije Britannice. Od toga je ispitano 17 biografskih članaka, 21 članak koji se odnosi na organizacije te 12 natuknica općega tipa (prilog 6).

Najveća zastupljenost unificiranih faktografskih podataka utvrđena je kod biografskih natuknica, gdje je uočeno ukupno 20 tipova naziva. U vrstu naziva vezanih uz prepoznavanje osoba, u određenome broju članaka navode se podatci o osobama kao suradnicima u vezi s područjem djelovanja (41,1%), članovima njihovih obitelji (11,7%) te relevantnim suvremenicima (23,5%). Kao vrste naziva vezanih uz prepoznavanje naziva organizacija ističu se organizacije na kojima je kandidat diplomirao (35,2%) i doktorirao (29,4%), gdje je ostvarivao svoju djelatnost (58,8%), te organizacije koje je osoba osnovala (41,2%) i vodila (11,7%).

Od vrsta naziva vezanih uz lokacije, svi ispitivani članci donose podatke o mjestu i državi rođenja (100%), ali redovito i o mjestu (64,7%) i državi (58,8%) smrti. Unificirani podatci vezani uz vremenska obilježja također su zastupljeni u formi datuma i godine smrti (64,7%) te godine diplomiranja (35,2%) i doktoriranja (23,5%) obrađivane osobe.

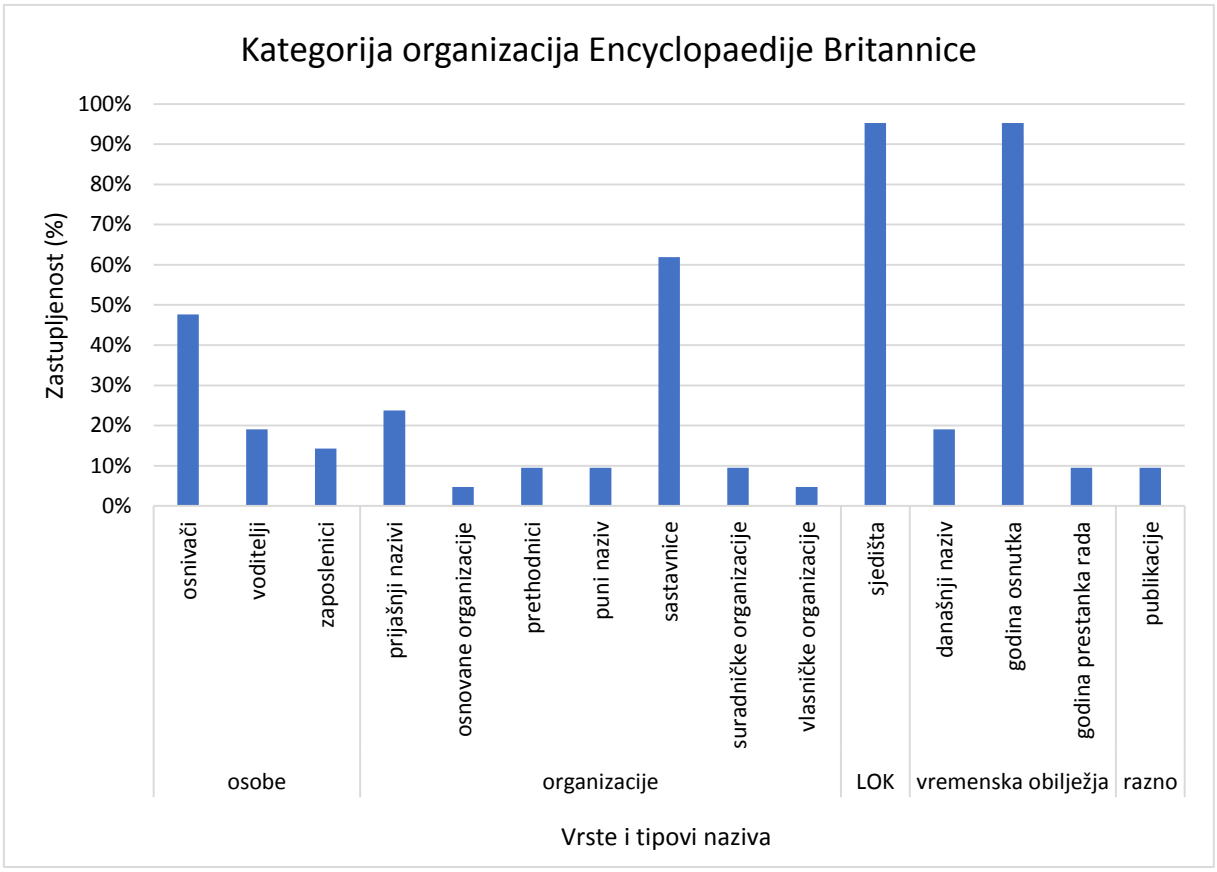
Od tipova naziva klasificiranih pod naziv razno zabilježeni su podatci o dodijeljenim nagradama (58,8%) i publikacijama (23,5%) vezani uz obrađivane osobe. Statistički prikaz vrsta i tipova naziva vezanih uz tip članka osobe Encyclopaedije Britannice prikazani su histogramom (slika 32).



Slika 32. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Encyclopaedije Britannice

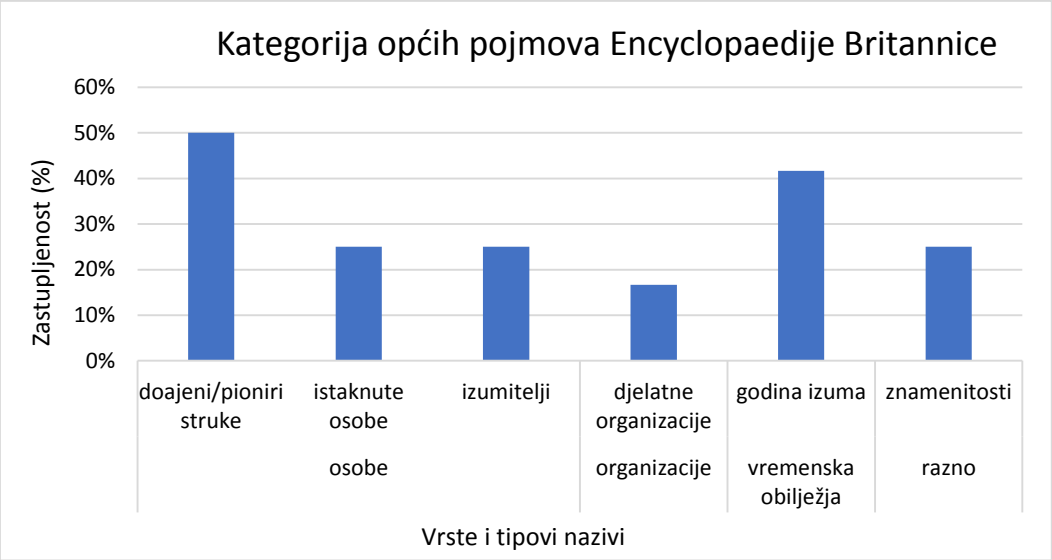
Kategorija natuknica organizacija Encyclopaedije Britannice od naziva vezanih uz prepoznavanje osoba donosi podatke osnivača (47,6%), zaposlenika (14,3%) i voditelja (19%) organizacija o kojima je riječ. Više zabilježenih podataka vezano je uz organizacije koje se javljaju u tekstovima članaka, u koje spadaju prijašnji nazivi promatranih organizacija (23,8%), njihovi prethodnici (9,5%), puni nazivi (9,5%), sastavnice (61,9%) te osnovane (4,8%), suradničke (9,5%) i vlasničke (4,7%) organizacije.

Od lokacija visokozastupljen je podatak o sjedištima organizacija (95,2%), a od vremenskih obilježja godina osnivanja (95,2%), uz prisutnost godine prestanka rada (9,5%) i preuzimanja današnjega naziva (19%). Zabilježen je i podatak o publikacijama koje su vezane uz obrađivane pojmove (9,5%). Statistički prikaz vrsta i tipova naziva vezanih uz tip članka organizacija Encyclopaedije Britannice prikazani su histogramom (slika 33).



Slika 33. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija Encyclopaedije Britannice

Nazivi koji opisuju veći broj natuknica najmanje su zastupljeni kod kategorije članaka općih pojmova. Prisutni su podatci o doajenima ili pionirima struke ili područja (50%), istaknutim osobama (25%) i osobama koje su ostvarile izum ili patent (25%). Od organizacija ističu se djelatne organizacije vezane uz pojmove (16,6%). Vremenska obilježja uključena su kao podatci o godinama postizanja izuma koji su vezani uz pojam (41%). Navode se i znamenitosti vezane za pojmove (25%). Statistički prikaz vrsta i tipova naziva vezanih uz tip članka organizacija Encyclopaedije Britannice prikazani su histogramom (slika 34).



Slika 34. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka općih pojmova Encyclopaedije Britannice

Uzorak Encyclopaedije Britannice namijenjen kvantitativnoj analizi sadržaja sadržavao je ukupno 9033 retka enciklopedičkoga teksta i zabilježena je prisutnost 3661 naziva. Nazivi su razmjerno jednoliko raspoređeni po vrstama koje ih opisuju. Razmjerno visok udio naziva koji se opisuju kao osobe u kategorijama članaka osoba rezultat je navođenja osoba u trećem licu.

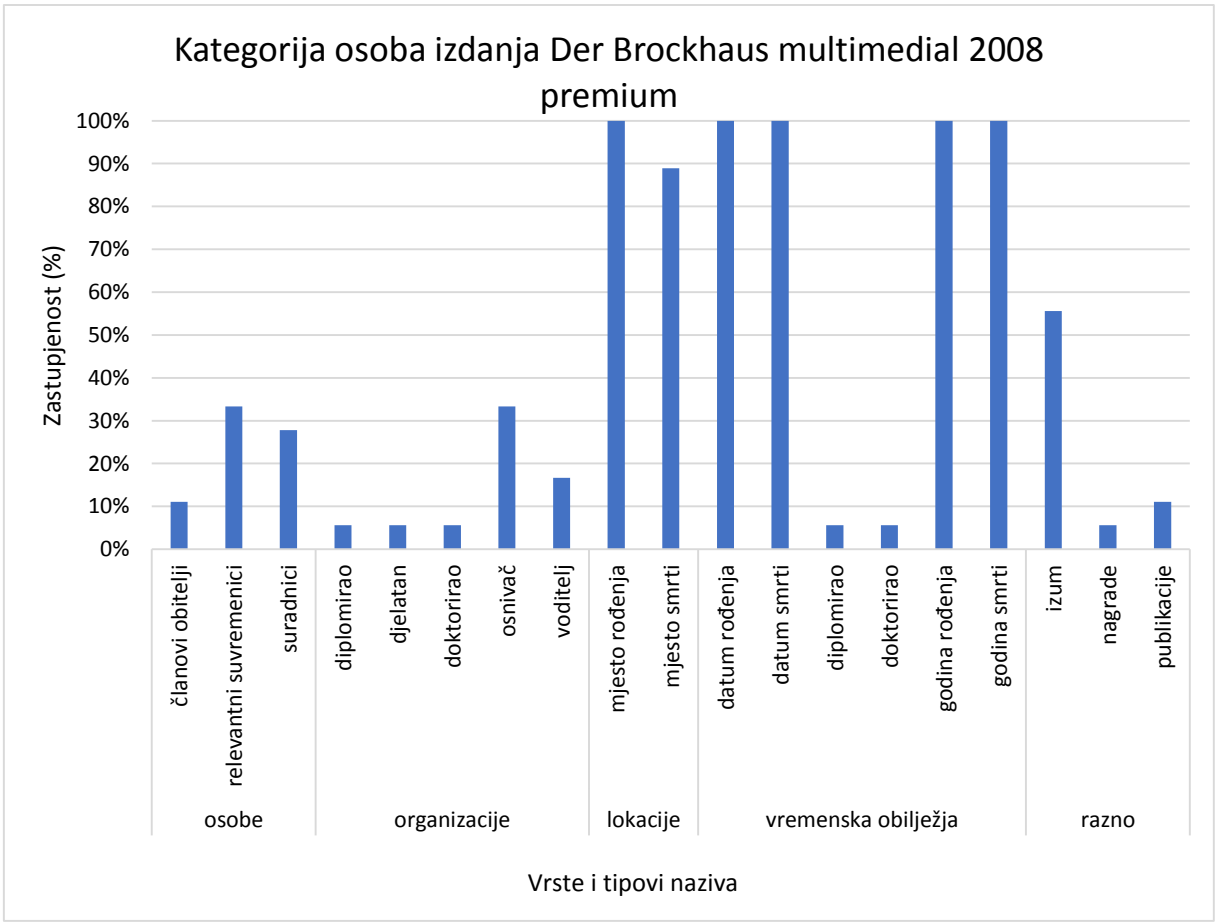
Mrežno izdanje Encyclopaedije Britannice sadrži članke veće opsegom, ali ne i toliko brojem naziva, što upućuje na slobodniji način sastavljanja toga sadržaja u odnosu na neke od sastavnica enciklopedičkoga koncepta. Ovo izdanje sadrži povećan raspon tipova naziva, pogotovo kod kategorija članaka osoba i općih pojmova, koji pokazuju i konstantnu povećanu zastupljenost istih.

4.3.1.7. Der Brockhaus multimedial 2008 premium

Ispitano je 50 članaka digitalnog izdanja Der Brockhaus multimedial 2008 premium, od čega 18 natuknica osoba, 16 organizacija i 16 općih pojmova. Iako nije riječ o mrežnom enciklopedičkom izdanju koje je javno dostupno korisnicima, uvršteno je u istraživanje kako bi se povećala raznovrsnost uvrštenih djela, te samim time jasnije dočarao postojeći standard enciklopedičkih djela s faktografskoga gledišta. On je kod ovog izdanja vidljiv i u rezultatima kvalitativne analize sadržaja, koja u konačnici pokazuje velik raspon tipova naziva, međutim, veći je dio smanjene zastupljenosti. To potvrđuje i uvid u kvantitativnu analizu sadržaja uzorka ovoga izdanja, koja pokazuje kako neke jedinice uzorka ne sadrže ni jedan naziv (prilog 7).

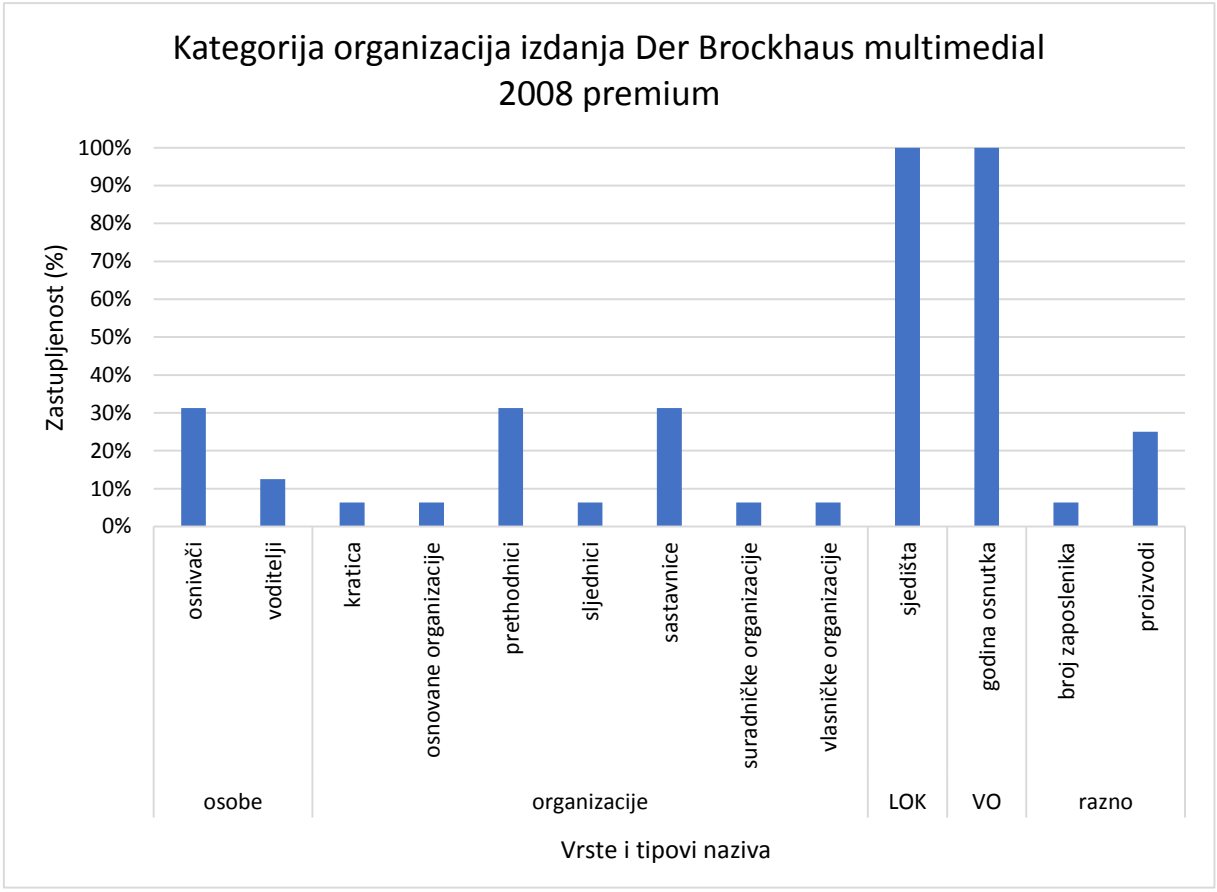
Najviše naziva donose članci osoba, njih 19. Kao i kod drugih izdanja, najveću zastupljenost pokazuju nazivi vezani uz lokacije i vremenska obilježja. Tu spadaju podatci o mjestu rođenja (100%) i smrti (88,9%), te datumima i godinama rođenja i smrti (100%). Zabilježeni su još podatci o godinama diplomiranja i postizanja doktorata znanosti (5,6%). Povišene su vrijednosti zastupljenosti naziva koji se odnose na osobe, odnosno na relevantne suvremenike (33,3%), suradnike (27,8%) i članove obitelji (11,1%) obrađivanih osoba. Tipova naziva koji se odnose na organizacije zabilježeno je pet, a opisuju obrađivane osobe kao osnivače (33,3%) i voditelje (16,7%) organizacija, te predstavljaju ustanove i poduzeća koja su bila mjesta obavljanja djelatnosti (5,6%) te postizanja diploma (5,6%) i doktorata znanosti (5,6%).

Nazivi koji se ne mogu egzaktno razvrstati u neke od standardnih vrsta koje se najviše pojavljuju u enciklopedičkim tekstovima, a pojavljuju se u člancima uzorka ovoga izdanja su navedeni izumi (55,6%), publikacije (11,1%) i nagrade (5,6%) vezane uz obrađivane osobe. Statistički prikaz vrsta i tipova naziva vezanih uz kategoriju članaka osoba izdanja Der Brockhaus multimedial 2008 premium prikazani su histogramom (slika 35).



Slika 35. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba izdanja Der Brockhaus multimedial 2008 premium

Kategorija natuknica organizacija smanjene je faktografske vrijednosti. To se ne odnosi na podatke o sjedištima i godinama osnutka organizacija (100%) koji su zastupljeni u svim ispitivanim izdanjima. Od ostalih naziva svih vrsta zabilježena je pojavnost podataka o osobama kao osnivačima (31,3%) i voditeljima (12,5%) obrađivanih organizacija, zatim organizacija koje predstavljaju kratice obrađivanih pojmova (6,3%), osnovanih organizacija (6,3%), prethodnike (31,3%), slijednike (6,3%), sastavnice (31,3%), suradničke i vlasničke organizacije (6,3%) obrađivanih pojmova. Također podatci koji se razmjerno nalaze u ovoj kategoriji natuknice su broj zaposlenika organizacija (6,3%) te proizvodi (25%). Statistički prikaz vrsta i tipova naziva vezanih uz kategoriju članaka organizacija izdanja Der Brockhaus multimedial 2008 premium prikazani su histogramom (slika 36).



Slika 36. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija izdanja Der Brockhaus multimedial 2008 premium

Kategorija članaka općih pojmova ne predstavlja enciklopedičke članke pogodne za strukturiranje putem faktografskih podataka. Zabilježena je manja zastupljenost podataka koji se odnose na doajene ili pionire struka ili područja koja predstavljaju obrađivani pojmovi te

podatci o osobama kao izumiteljima ako je riječ o pojmu koji donosi takvu vrstu podatka (12,5%).

Ukupno je ispitano 1760 redaka enciklopedičkoga teksta digitalnog izdanja Der Brockhaus multimedial 2008 premium te je utvrđeno postojanje 1064 naziva sukladno ovom istraživanju. Iako ispitivani uzorak sadrži natuknice većega opsega, znatan broj članaka predstavlja samo definicije pojmova, te ne sadrži nazive potrebne za stvaranje strukture. To je posebno vidljivo u kategoriji članaka općih pojmova, ali i u ostalim dvjema kategorijama. Promatrajući zasebne jedinice sadržaja, zabilježen je nedostatak naziva u određenom dijelu uzorka, na što možda ne upućuje ukupni omjer broja naziva i ispitanog opsega teksta.

Digitalno izdanje Der Brockhaus multimedial 2008 premium sadrži velik broj enciklopedičkih natuknica te, kao i mrežna izdanja Proleksis i Hrvatske enciklopedije, potencijalno predstavlja vrijedan sadržaj za vezanje strukture putem interoperabilnosti s drugim izdanjima.

4.3.1.8. Wikipedija

Uzorak Wikipedije sačinjavali su članci izdanja hrvatske, engleske i njemačke inačice, koje su promatrane kao zasebna mrežna enciklopedička izdanja. Svaka inačica u istraživanju bila je zastupljena s 50 članaka, od čega 22 članka koji se odnose na pojmove organizacija, 17 životopisa osoba te 11 članaka općih pojmova. Važno je napomenuti kako su obrađivani isti pojmovi, ali s gledišta hrvatske, engleske ili njemačke inačice ove enciklopedije. Uzorak svake inačice ovoga slobodnog internetskog izdanja koje svatko može uređivati povećanog je opsega, međutim, kad je riječ o pojavnosti naziva, taj podatak nije utjecao na isticanje Wikipedije u faktografskom smislu koje je predmet ovoga istraživanja.

Kategorija natuknica osoba donosi ukupno 25 tipova naziva svih ispitivanih vrsta. Njihova zastupljenost u skladu je s izdanjima u istraživanju koja su pokazala bolje rezultate, pogotovo ako promatramo ispitivane inačice Wikipedije kao cjelinu.

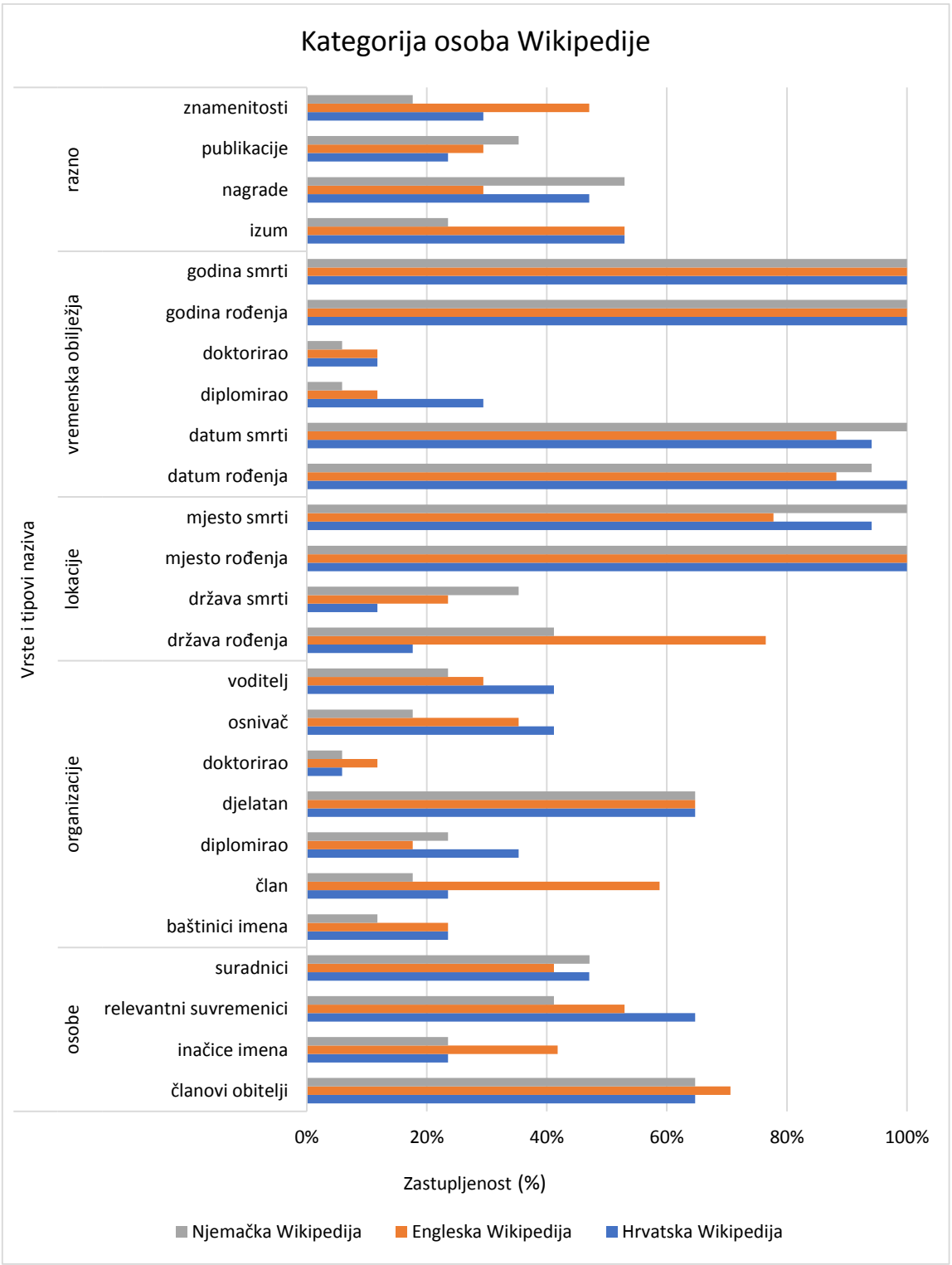
Vrste naziva koje se odnose na prisutnost podataka o osobama u tekstovima članaka Wikipedije bolje su zastupljenosti u odnosu na druga izdanja. To se pogotovo odnosi na članove

obitelji obrađivanih osoba (64,7%; 70,6%; 64,7%)[198], suradnike (47,1%; 41,2%; 47,1%), relevantne suvremenike (64,7%; 52,9%; 41,2%) te inačice imena obrađivanih osoba (23,5%; 41,8%; 23,5%). Podatci koji se odnose na prisutnost naziva vezanih uz navođenje organizacija su oni koji označavaju nositelje djelatnosti osoba (64,7%; 64,7%; 64,7%), baštinike imena osoba (23,5%; 23,5%; 11,7%), osobe kao osnivače (41,2%; 35,3%; 17,6%) i voditelje (41,2%; 29,4%; 23,5%) te članove (23,5%; 58,8%; 17,6%) organizacija, ali i organizacije kao ustanove postizanja diploma (35,3%; 17,6%; 23,5%) i doktorata znanosti (5,9%; 11,7%; 5,9%).

Već uvriježeno, najveća zastupljenost zabilježena je kod naziva koji se odnose na lokacije i vremenska obilježja. Značajno zastupljeni podatci ove kategorije enciklopedičkoga članka su mjesta rođenja (100%; 100%; 100%) i smrti (94,1%; 82,3%; 100%), datumi rođenja (100%; 88,2%; 94,1%) i smrti (94,1%; 88,2%; 100%) osoba te godine rođenja i smrti (100%; 100%; 100%). Mrežna izdanja Wikipedije donose još podatke o državama u kojima se nalaze mjesta rođenja (17,7%; 76,5%; 41,2%) i smrti (11,8%; 23,5%; 35,3%), te godinama postizanja diploma (29,4%; 11,8%; 5,9%) i doktorata znanosti (11,8%; 11,8%; 5,9%).

Pod skupinom razno nalaze se nazivi u vezi s dodijeljenim nagradama vezanima za osobe (47,1%; 29,4%; 52,9%), znamenitosti (29,4%; 47,1%; 17,6%), postignuti izumi ili patenti (52,9%; 52,9%; 23,5%) te objavljene publikacije (23,5%; 9,4%; 35,3%). Statistički prikaz vrsta i tipova naziva vezanih uz tip članka osoba izdanja hrvatske, engleske i njemačke Wikipedije i njihova usporedba prikazani su histogramom (slika 37).

[198] Vrijednosti navedene u zagradama odnose se na postotke vezane uz zastupljenost određenog tipa podatka u člancima uzorka za hrvatsku, englesku i njemačku inačicu Wikipedije. Npr., navedena formulacija navodi zastupljenost entiteta osoba koji se odnose na članove obitelji, i to 64,7% u Hrvatskoj Wikipediji, 70,6% u Engleskoj Wikipediji te 64,7% u Njemačkoj Wikipediji.



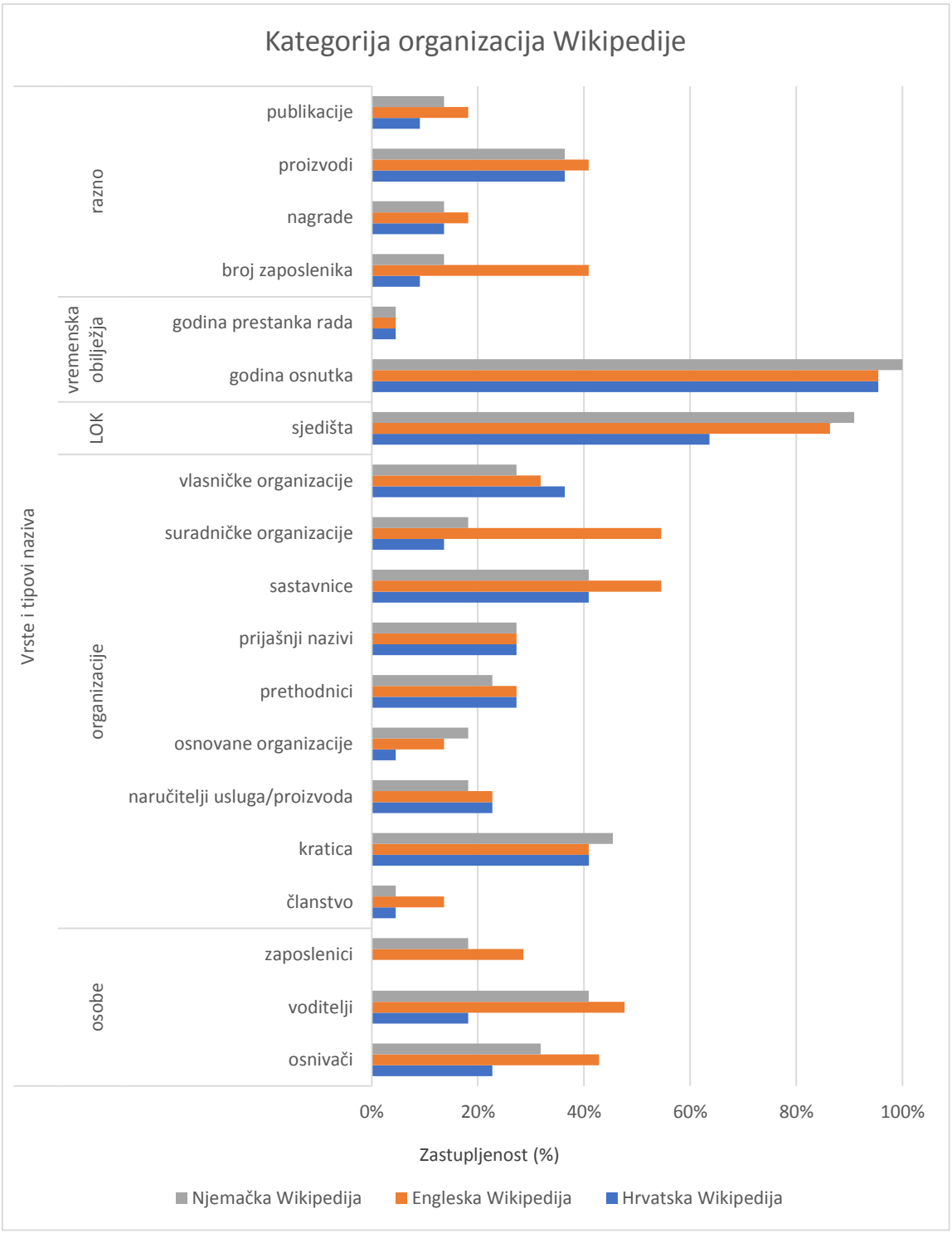
Slika 37. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba izdanja hrvatske, engleske i njemačke inačice Wikipedije

Kategorija natuknica u koje spadaju organizacije opisana je u 19 tipova naziva. Faktografski najinformativnijom pokazala se engleska inačica Wikipedije, u kojoj su zastupljeni svi tipovi podataka s povišenim vrijednostima.

Podatci vezani uz prepoznavanje osoba odnose se na osnivače (22,7%; 42,9%; 31,8%), voditelje (18,2%; 47,6%; 40,1%) i zaposlenike (–; 28,6%; 18,2%) organizacija. Kao i kod rezultata ostalih ispitivanih izdanja, najviše naziva ove kategorije članaka odnosi se na prisutnost podataka o drugim organizacijama. U odnosu na obrađivane pojmove, pojavljuju se kao njihove kratice (40,9%; 40,9%; 45,5%), prethodnici (27,3%; 27,3%; 22,7%), prijašnji nazivi (27,3%; 27,3%; 27,3%), sastavnice (40,9%; 54,5%; 40,9%) te suradničke (13,6%; 54,5%; 18,2%), vlasničke (36,4%; 21,8%; 27,3%) ili osnovane organizacije (4,5%; 13,6%; 18,2%). Također, ovo izdanje sustavno donosi podatke o organizacijama kao naručiteljima usluga ili proizvoda (22,7%; 22,7%; 18,2%) te pružateljima članstva (4,5%; 13,6%; 4,5%).

Lokacije i vremenska obilježja najzastupljeniji su nazivi ove kategorije enciklopedičkoga članka. Odnose se na podatke o sjedištima organizacija (63,6%; 86,4%; 90,9%) te godinama osnutka (95,5%; 95,5%; 100%), uz moguće donošenje podatka o godini prestanka rada (4,5%; 4,5%; 4,5%).

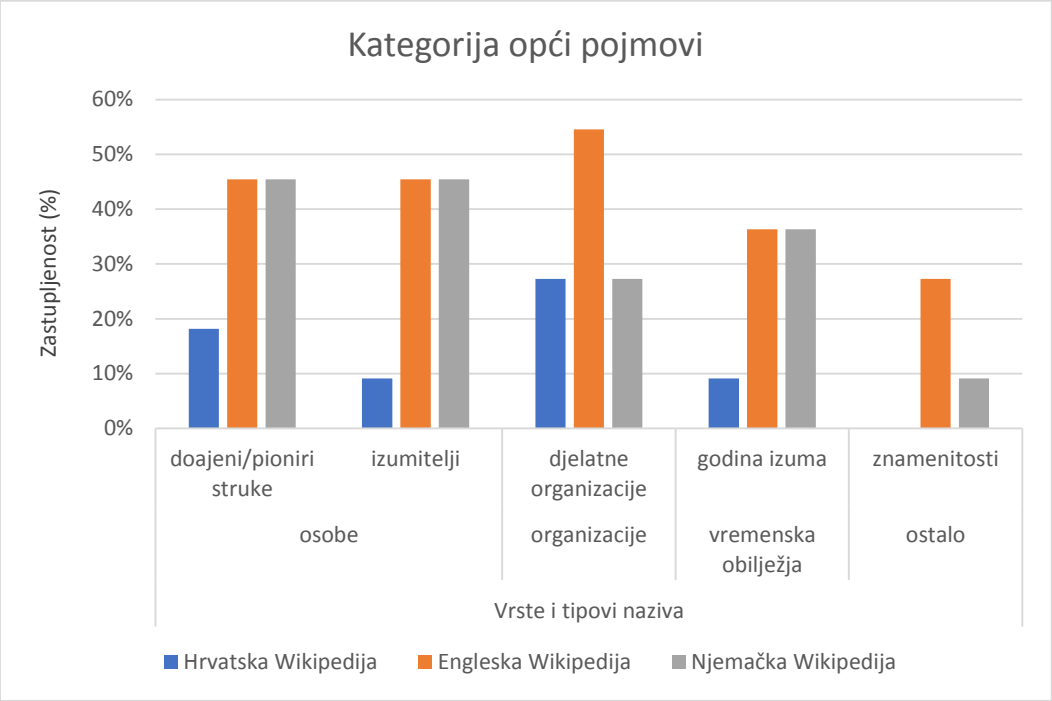
Također, zabilježeno je postojanje podataka o brojevima zaposlenika obrađivanih organizacija (9,1%; 40,9%; 13,6%), te vezanim nagradama (13,6%; 18,2%; 13,6%), publikacijama (9,1%; 18,2%; 13,6%) i proizvodima (36,4%; 40,9%; 36,4%). Statistički prikaz vrsta i tipova naziva vezanih uz tip članka organizacija izdanja hrvatske, engleske i njemačke Wikipedije prikazani su histogramom (slika 38).



Slika 38. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija izdanja hrvatske, engleske i njemačke inačice Wikipedije

Kategorija općih pojmova smanjene je faktografske vrijednosti. Ističu se podatci o doajenima ili pionirima struke ili područja (18,2%; 45,5%; 45,5%) te izumiteljima (9,1%; 45,5%; 45,5%) kao pojavnice vezane uz osobe, o djelatnim organizacijama (27,3%; 54,5%; 27,3%), godinama izuma (9,1%; 36,4%; 36,4%) te znamenitostima (–; 27,3%; 9,1%) koji se

odnose na obrađivanu problematiku. Statistički prikaz vrsta i tipova naziva vezanih uz tip članka općih pojmova izdanja hrvatske, engleske i njemačke Wikipedije prikazani su histogramom (slika 39).



Slika 39. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka općih pojmova izdanja hrvatske, engleske i njemačke inačice Wikipedije

Ukupno je ispitano 150 članaka mrežnih izdanja Wikipedije, po 50 hrvatske, engleske i njemačke inačice. Od toga, opseg ispitanih enciklopedičkih tekstova na hrvatskom jeziku iznosio je 5021 redak uz zabilježenih 3669 naziva, na engleskom jeziku ispitano je 12 267 redaka uz zabilježenih 7998 naziva, dok je njemačka inačica sudjelovala s ispitana 9992 retka i zabilježenih 6013 naziva.

Analiza ovih triju inačica Wikipedije, koje predstavljaju tri zasebne mrežne enciklopedije zasnovane na slobodnom principu, upućuje na to kako velike razlike u opsegu ovih triju izdanja ne utječu na raznolikost tipova podataka. Zastupljenost tipova naziva u tekstu najopširnije engleske Wikipedije povećana je, međutim, ako te rezultate usporedimo s rezultatima pojedinih izdanja zastupljenih u ovom ispitivanju, poput Hrvatske tehničke enciklopedije kao specijaliziranog enciklopedičkoga izdanja vezanog za tehniku, Wikipedija predstavlja slabije informativno izdanje u smislu donošenja unificiranih faktografskih podataka karakterističnih za određenu kategoriju članaka, unatoč puno većem opsegu i gotovo dva puta većem broju zastupljenih naziva u tekstu.

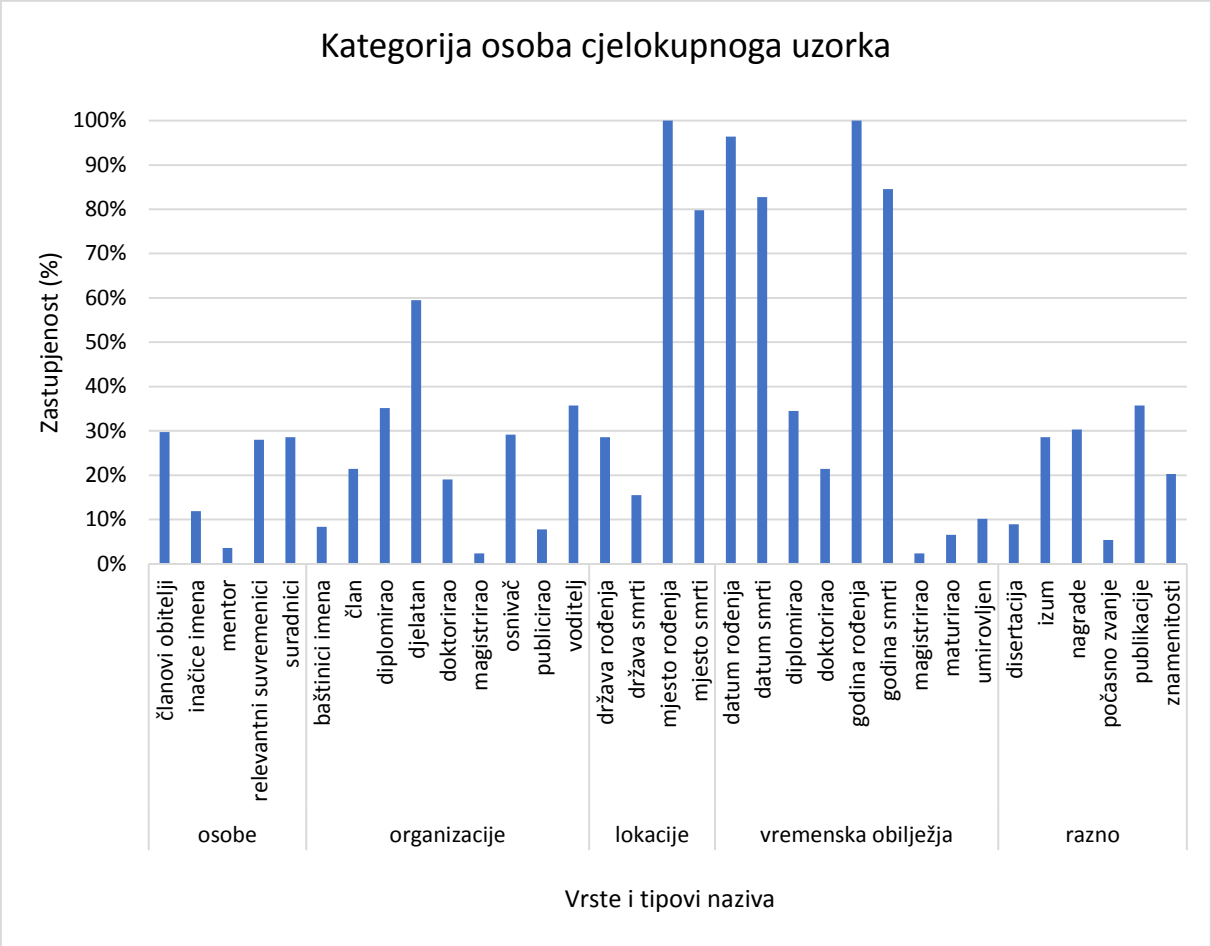
Osim slobodnijeg načina sastavljanja sadržaja i neujednačenijih članaka, kod Wikipedijinih inačica prisutne su jedinice sadržaja s povećanim brojem naziva koji nisu unificirani u smislu redovitog donošenja podataka i mogućeg strukturiranja u cilju razvoja korisničkih aplikacija i postizanja interoperabilnosti. Oni utječu na mogući zaključak vezan uz uporabljivost ovog projekta putem strukturiranosti. Primjer toga je natuknica *Jadrolinija* hrvatske inačice koja sadrži 893 naziva, od kojih se polovica odnosi isključivo na luke povezane trajektnim linijama, što očito nije standardizirani enciklopedički tip podatka. Iako, Wikipedijini članci bogato su strukturirani i sadrže infookvire koji donose strukturirane podatke vezane uz pojmove koji se ne nalaze u tekstovima članaka, a koji kao metapodatci nalaze druge namjene (poglavlje 3. 3. 4.)

Uzimajući u obzir tekstove Wikpedijinih članaka kao sastavnica ovog istraživanja, slobodni pristup sastavljanju sadržaja i odmak od uvriježenih principa enciklopedičkoga koncepta ne umanjuje vrijednost ovog projekta u smislu mogućeg strukturiranja sadržaja na temelju unificiranih faktografskih podataka, a već postignuta organizacija i struktura enciklopedičkih članaka Wikipedije mogu biti primjer kako pristupiti dodatnom povećanju vrijednosti mrežnog enciklopedičkog sadržaja.

4.3.2. ANALIZA SADRŽAJA JEDINSTVENOGA UZORKA I USPOREDBA REZULTATA ZASEBNIH IZDANJA

Jedinstveni uzorak namijenjen analizi sadržaja tekstova enciklopedičkih članaka činilo je ukupno 455 jedinica sadržaja kao dijelova deset mrežnih enciklopedičkih izdanja. Preliminarnom analizom, koja je obuhvaćala analizu sadržaja jedne zaokružene enciklopedičke cjeline, članci uzorka podijeljeni su u tri kategorije na temelju kriterija donošenja unificiranih faktografskih podataka pogodnih za stvaranje strukture mrežnih enciklopedičkih projekata u cilju pospješavanja funkcionalnosti korisničkih aplikacija i postizanja intropera bilnosti.

Slika 40 predstavlja rezultate analize sadržaja jedinstvenoga uzorka. U tom smislu, prikazuje uniju svih tipova naziva zabilježenih ovim istraživanjem i upućuje na postojeći faktografski standard enciklopedičkih tekstova. Predstavljaju prosječne vrijednosti zastupljenosti te na taj način mogu poslužiti kao polazište za usporedbu i valorizaciju zasebno ispitanih izdanja u uzorku. Kategorija članaka osoba ukupno opisuje 33 tipa naziva svih ispitivanih izdanja.



Slika 40. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba cjelokupnog uzorka

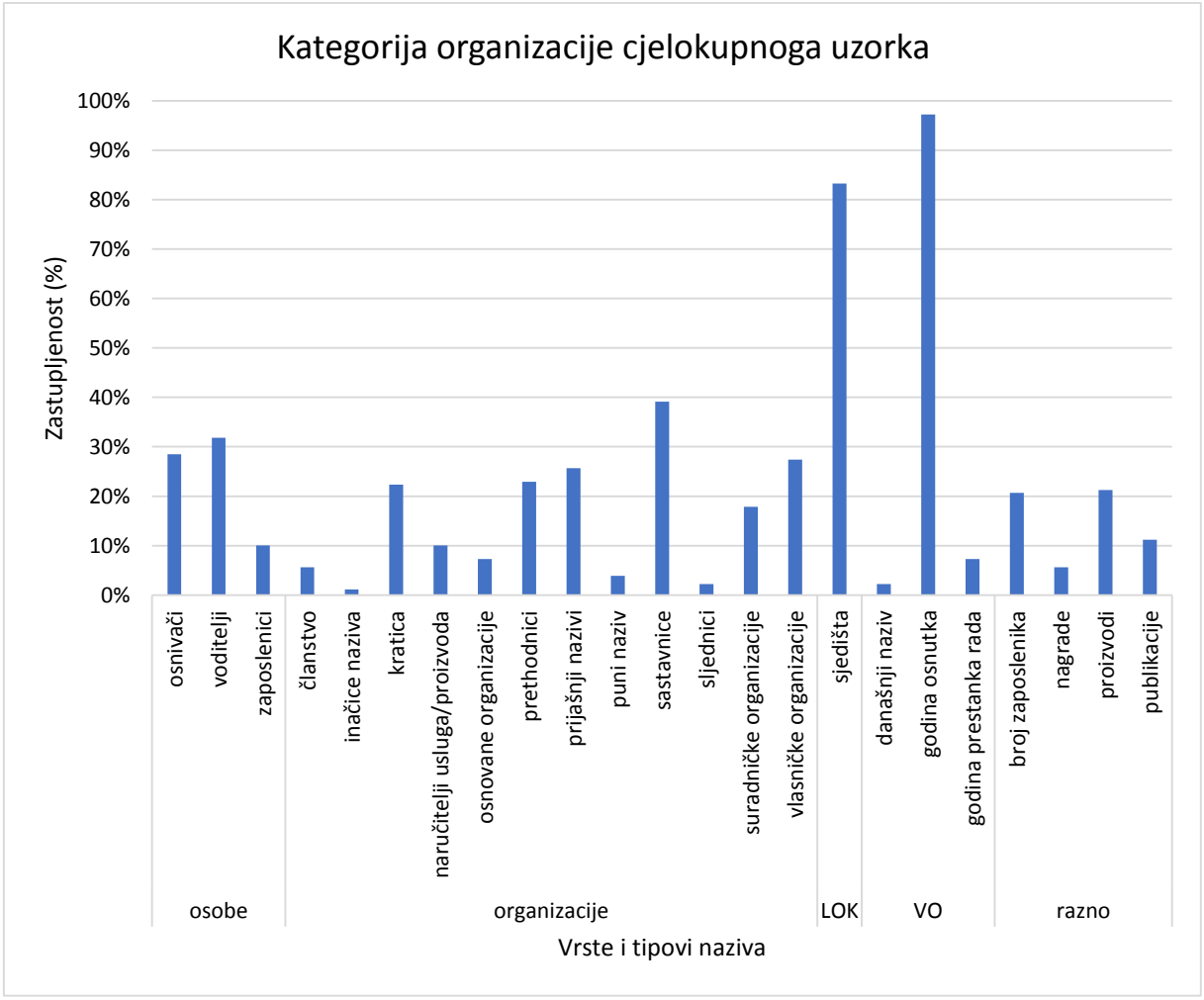
Tablica 11 donosi prikaz svih vrsta i tipova naziva vezanih uz kategoriju članaka osoba i njihovu zastupljenost u svim enciklopedičkim izdanjima uključenima u ispitivanje, uz prikaz srednjih vrijednosti kao jedinstvenog uzorka.

Tablica 11. Zastupljenost tipova naziva kategorije članaka osoba svih zastupljenih izdanja uzorka uz navedene srednje vrijednosti

Vrste i tipovi naziva		Zastupljenost entiteta (%)										
		HTE	HE	PE	IE	HBL	EB	BRO	W_{hr}	W_{en}	W_{de}	x
osobe	članovi obitelji	6,3	6,3	5,3	9,1	40,0	11,7	11,1	64,7	70,6	64,7	29,8
	inačice imena	25,0	6,3	5,3	9,1	25,0			23,5	41,8	23,5	11,9
	mentor	37,5										3,6
	relevantni suvremenici	18,8	12,5	15,8		10,0	23,5	33,3	64,7	52,9	41,2	28,0
	suradnici	18,8	6,3	21,1	18,2	15,0	41,1	27,8	47,1	41,2	47,1	28,6
organizacije	baštinici imena	12,5				10,0			23,5	23,5	11,8	8,3
	član	37,5	37,5	5,3	36,4	35,0			23,5	58,8	17,6	21,4
	diplomirao	62,5	68,8	10,5	36,4	60,0	35,2	5,6	35,3	17,6	23,5	35,1
	djelatan	81,3	81,3	26,3	81,8	80,0	58,8	5,6	64,7	64,7	64,7	59,5
	doktorirao	50,0	43,8			35,0	29,4	5,6	5,9	11,8	5,9	19,1
	magistrirao	25,0										2,4
	osnivač	50,0	43,8	5,3	9,1	15,0	41,2	33,3	41,2	35,3	17,6	29,2
	publicirao					65,0						7,7
	voditelj	50,0	68,8	21,1	45,5	40,0	29,4	16,7	41,2	29,4	23,5	35,7
lokacije	država rođenja	18,8	18,8	10,5			100		17,7	76,5	41,2	28,6
	država smrti		6,3	5,3	9,1	5,0	58,8		11,8	23,5	35,3	15,5
	mjesto rođenja	100	100	100	100	100	100	100	100	100	100	100
	mjesto smrti	62,5	81,3	100	72,7	55,0	64,7	88,9	94,1	77,8	100	79,8
vremenska obilježja	datum rođenja	93,8	100	94,7	100	95,0	100	100	100	88,2	94,1	96,4
	datum smrti	62,5	81,3	100	81,8	55,0	64,7	100	94,1	88,2	100	82,7
	diplomirao	62,5	75,0		63,6	70,0	35,2	5,6	29,4	11,8	5,9	34,5
	doktorirao	43,5	50,0		9,1	50,0	23,5	5,6	11,8	11,8	5,9	21,4
	godina rođenja	100	100	100	100	100	100	100	100	100	100	100
	godina smrti	62,5	81,3	100	81,8	55,0	64,7	100	100	100	100	84,5
	magistrirao	25,0										2,4
	maturirao					55,0						6,6
	umirovljen	43,5	18,8		18,2	25,0						10,1
razno	disertacija	43,5				40,0						8,9
	izum	18,8	2,5	36,8		20,0		55,6	52,9	52,9	23,5	28,6
	nagrade	43,5	43,5	5,3	18,2	55,0	58,8	5,6	47,1	29,4	52,9	30,3
	počasno zvanje	25,0	18,8			10,0						5,4
	publikacije	62,5	68,8	5,3	18,2	75,0	23,5	11,1	23,5	29,4	35,3	35,7
	znamenitosti	43,5	18,8	10,5	18,2	20,0			29,4	47,1	17,6	20,2

Povišene vrijednosti zastupljenosti i broja tipova naziva pokazuju izdanja Hrvatske tehničke enciklopedije, Hrvatske enciklopedije, Hrvatskoga biografskog leksikona i inačica Wikipedije, dok izdanja Proleksis enciklopedije, Istarske enciklopedije, Encyclopaedije Britannice i izdanja Der Brockhaus multimedial 2008 premium pokazuju smanjene vrijednosti zastupljenosti naziva i manji broj tipova podataka, naravno, izuzevši već spomenute standardne tipove podataka prisutnih gotovo isključivo u ovoj kategoriji enciklopedičkoga članka u svim izdanjima.

Kategorija članaka koja okuplja organizacije opisana je ovim istraživanjem u ukupno 23 tipa naziva svih vrsta uključenih u istraživanje (slika 41).



Slika 41. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija cjelokupnog uzorka

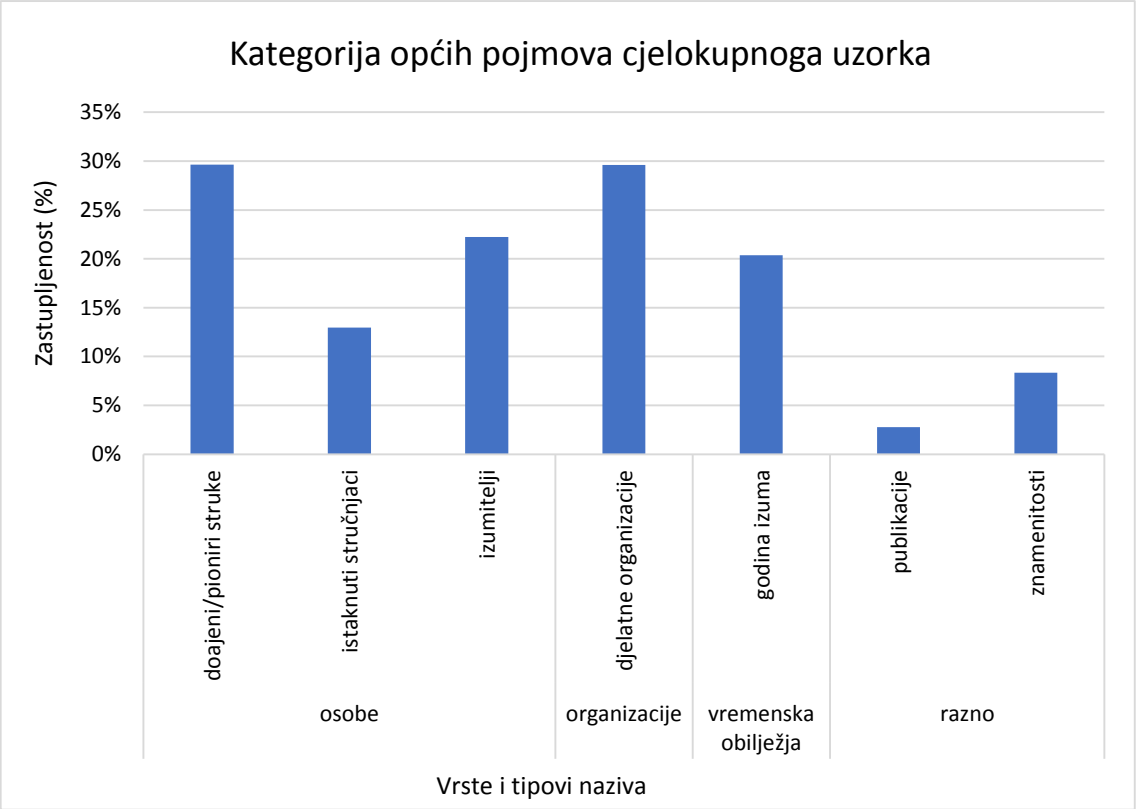
Tablica 12 donosi prikaz svih vrsta i tipova naziva vezanih za kategoriju članaka organizacija i njihovu zastupljenost u svim enciklopedičkim izdanjima uključenima u ispitivanje, uz prikaz srednjih vrijednosti kao jedinstvenog uzorka.

Tablica 12. Zastupljenost tipova naziva kategorije članaka organizacija svih zastupljenih izdanja uzorka uz navedene srednje vrijednosti

Vrste i tipovi naziva		Zastupljenost entiteta (%)									
		HTE	HE	PE	IE	EB	BRO	W_{hr}	W_{en}	W_{de}	x
osobe	osnivači	16,0	20,0	18,2	26,7	47,6	31,3	22,7	42,9	31,8	28,5
	voditelji	72,0	12,0		46,7	19,0	12,5	18,2	47,6	40,9	31,8
	zaposlenici	16,0			6,7	14,3			28,6	18,2	10,1
organizacije	članstvo	20,0						4,5	13,6	4,5	5,6
	inačice naziva				13,3						1,1
	kratica	28,0	8,0		13,3		6,3	40,9	40,9	45,5	22,4
	naručitelji usluga/proizvoda	12,0			6,7			22,7	22,7	18,2	10,1
	osnovane organizacije	12,0				4,8	6,3	4,5	13,6	18,2	7,3
	prethodnici	28,0	20,0	18,2	20,0	9,5	31,3	27,3	27,3	22,7	22,9
	prijašnji nazivi	56,0	20,0	9,1	20,0	23,8		27,3	27,3	27,3	25,7
	puni naziv		8,0	27,3		9,5					3,9
	sastavnice	48,0	24,0		26,7	61,9	31,3	40,9	54,5	40,9	39,1
	slijednici		4,0		13,3		6,3				2,2
	suradničke organizacije	28,0			20,0	9,5	6,3	13,6	54,5	18,2	17,9
	vlasničke organizacije	44,0	36,0	9,1	33,3	4,8	6,3	36,4	31,8	27,3	27,4
LOK	sjedišta	84,0	76,0	81,8	73,3	95,2	100	63,6	86,4	90,9	83,2
VO	današnji naziv					19,0					2,2
	godina osnutka	100	100	90,9	93,3	95,2	100	95,5	95,5	100,0	97,2
	godina prestanka rada	20,0	4,0		13,3	9,5		4,5	4,5	4,5	7,3
razno	broj zaposlenika	28,0	28,0	18,2	40,0		6,3	9,1	40,9	13,6	20,7
	nagrade							13,6	18,2	13,6	5,6
	proizvodi	20,0	8,0	9,1	6,7		25,0	36,4	40,9	36,4	21,2
	publikacije	24,0	8,0		6,7	9,5		9,1	18,2	13,6	11,2

Faktografski bogatija izdanja unificiranim podatcima vezanima uz organizacije u tehničkome području su Hrvatska tehnička enciklopedija i inačice Wikipedije, iako i ostala izdanja sadrže pojedine tipove naziva u razmjerno povišenim vrijednostima, budući da ova kategorija enciklopedičkoga članka pokazuje manju faktografsku vrijednost unificiranim podatcima koji su predmet ovog istraživanja.

Kategorija članaka općih pojmova, ako se promatra kao jedinstven uzorak, donosi ukupno sedam tipova naziva i ne predstavlja povoljni enciklopedički sadržaj za strukturiranje putem unificiranih faktografskih podataka u odnosu na kategorije osoba i organizacija (slika 42).



Slika 42. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka općih pojmova cjelokupnog uzorka

Tablica 13 donosi prikaz svih vrsta i tipova naziva vezanih uz kategoriju članaka organizacija i njihovu zastupljenost u svim enciklopedičkim izdanjima uključenima u ispitivanje, uz prikaz srednjih vrijednosti kao jedinstvenog uzorka.

Tablica 13. Zastupljenost tipova naziva kategorije članaka općih pojmova svih zastupljenih izdanja uzorka uz navedene srednje vrijednosti

Vrste i tipovi naziva		Zastupljenost entiteta (%)									
		HTE	HE	PE	IE	EB	BRO	W_{hr}	W_{en}	W_{de}	x
OS	doajeni/pioniri struke	44,4	55,6	15,0		50,0	12,5	18,2	45,5	45,5	29,6
	istaknuti stručnjaci	77,8	11,1		33,3	25,0					13,0
	izumitelji	22,2	22,2	25,0		25,0	12,5	9,1	45,5	45,5	22,2
ORG	djelatne organizacije	88,9	55,6		55,6	16,7		27,3	54,5	27,3	29,6
VO	godina izuma	22,2	22,2	15,0		41,7		9,1	36,4	36,4	20,4
razno	publikacije	33,3			22,2						2,8
	znamenitosti					25,0			27,3	9,1	8,3

Članci općega tipa u uzorku najviše osciliraju opsegom i informativnošću. Od isključivo definicija pojmova leksikonskoga tipa do opširnih prikaza određenih područja. Enciklopedička izdanja koja donose faktografski bogatije članke ove kategorije su Hrvatska tehnička enciklopedija, Hrvatska enciklopedija, Encyclopaedia Britannica i inačice Wikipedije. Također, članci ove kategorije, prema preliminarnoj analizi, najviše su zastupljeni u sastavljanju enciklopedičkih projekata.

Međutim, postizanjem interoperabilnosti među navedenim mrežnim enciklopedičkim izdanjima, nazivi kao metapodatci koji predstavljaju sadržaj postaju dio jedinstvene cjeline i u tom kontekstu ih također treba razmatrati. Nazivi prisutni u više izdanja potencijalnom interakcijom omogućuju povećanu zastupljenost takvog tipa podataka u svim interoperabilnim izdanjima, pospješujući druge važne sastavnice enciklopedičkoga koncepta poput ažurnosti i točnosti, budući da objedinjuju rad više autora, urednika, uredništava. S druge strane, strukturirani faktografski podatci rjeđe razmatrani u enciklopedičkim projektima predstavljaju nadopunu svim povezanim izdanjima koja ih ne donose, povećavajući njihove epistemološke vrijednosti koje proizlaze iz strukturiranoga sadržaja.

5. ISPITIVANJE ALATA ZA PREPOZNAVANJE NAZIVLJA NA ENCIKLOPEDIČKIM TEKSTOVIMA

Automatsko prepoznavanje nazivlja (engl. *Named Entity Recognition*, NER) dio je šireg područja koje se naziva obradom prirodnog jezika (OPJ). Obrada prirodnog jezika (engl. *Natural Language Processing*, NLP) kao interdisciplinarno područje omogućava automatizaciju procesa razumijevanja, stvaranja i raspoznavanja prirodnog jezika.[199] Također, prema Khurana et. al.[200] obrada prirodnog jezika dio je umjetne inteligencije i lingvistike posvećen računalnom razumijevanju izjava ili riječi napisanih ljudskim (prirodnim) jezikom. Cilj joj je obrada i analiziranje velike količine podataka na prirodnom jeziku koristeći se računalnim sustavima. Kao takva predstavlja interdisciplinarno područje čijim se razvojem i uslugama bave stručnjaci i znanstvenici širokog područja interesa.

5.1. OBRADA PRIRODNOG JEZIKA

Većina tehnika obrade prirodnog jezika oslanja se na strojno učenje (engl. *Machine Learning*, ML) [201] u cilju dohvaćanja značenja [202] iz prirodnog jezika ili neke druge reprezentacije sadržaja. Strojno učenje temelji se na statističkom zaključivanju kako bi se usvojila pravila automatiziranim postupkom kroz analizu opsežnih korpusa sastavljenih od uobičajenih primjera prirodnog jezika. Obrada prirodnog jezika stoga podrazumijeva primjenu računalnih algoritama za prepoznavanje i izdvajanje pravila prirodnog jezika tako da se nestrukturirani jezični podatci pretvaraju u strojno razumljiv oblik. Prilikom obavljanja tih zadataka koriste se sintaktička i semantička analiza kao glavne tehnike.

[199] Prirodnim se jezikom ljudi koriste za govor, pisanje ili općenito u svrhu komunikacije.

[200] Khurana et al.: Natural Language Processing: State of The Art, Current Trend and Challenges (2006) https://arxiv.org/abs/1708.05148v1 (29. 5. 2020.)

[201] Vidi: Jain, A., Kulkarni, G., Shah, V.: Natural Language Processing. International Journal of Computer Sciences and Engineering. 6(2018) 1, str. 161–167.

[202] Značenje ili semantičko obilježje predstavlja pojam što ga označava neka riječ. Značenje izraza u prirodnom jeziku može se prikazati kao logički oblik nakon njegova parsiranja i lišenja sintaktičke ovisnosti. Vidi natuknicu: značenje. *Hrvatska enciklopedija, mrežno izdanje*. Leksikografski zavod Miroslav Krleža, 2020. http://www.enciklopedija.hr/Natuknica.aspx?ID=67349 (5. 6. 2020.)

Sintaktička analiza koristi se za procjenu usklađenosti prirodnog jezika s gramatičkim pravilima. Algoritmi primjenjuju gramatička pravila na skupini riječi i iz njih izvode značenja. Neke od metoda koje se koriste su lematizacija (svođenje riječi na kanonski oblik), morfološka segmentacija (razdvajanje riječi na morfeme i identificiranje razreda morfema), segmentacija riječi (razdvajanje teksta na pojavnice ili opojavničenje), parsiranje (gramatička analiza dane rečenice), gramatičko označavanje (označavanje riječi u tekstu koje odgovaraju pojedinoj kategoriji sličnih gramatičkih svojstava), korjenovanje (dobivanje korijena riječi) i dr.

Semantička analiza predstavlja primjenu računalnih algoritama za razumijevanje značenja i interpretacije riječi i strukture rečenice. Neke od metoda semantičke analize su automatsko prepoznavanje nazivlja (prepoznavanje dijelova teksta koji se mogu prepoznati i kategorizirati u unaprijed određene skupine), stvaranje prirodnog jezika (pretvaranje strukturiranih podataka, npr. iz baza podataka u prirodni jezik), strojno prevođenje (automatizirano prevođenje teksta s jednoga prirodnog jezika na drugi), analiza sentimenta (izdvajanje subjektivnih informacija, odnosno mišljenja kao osjećaja osobe prema nekom predmetu ili događaju) i dr.

5.2. AUTOMATSKO PREPOZNAVANJE NAZIVLJA

Prema Ljubešiću et al.[203], automatsko prepoznavanje nazivlja dio je crpenja (ekstrakcije) informacija (engl. *Information Extraction*, IE) čime se želi pronaći i klasificirati elemente u prethodno određene kategorije koristeći se statističkim modelima ili modelima temeljenima na pravilima. Nastoji locirati i klasificirati nazive (entitete) u nestrukturiranom tekstu u unaprijed definirane kategorije kao što su imena osoba, nazivi organizacija, lokacije, vremenska obilježja, količine, novčane vrijednosti, postotci i dr. Proces ekstrakcije podataka nastoji nestrukturirane informacije sadržane u tekstu preobličiti u strukturirane podatke ili kodirani sadržaj. Prvi korak većine zadataka ekstrakcije informacija upravo je prepoznavanje nazivlja. U procesu automatskog prepoznavanja nazivlja, naziv se odnosi na one dijelove (raspone) teksta koji se u najvećoj mjeri odnose na vlastite imenice, ali i na druge vrste riječi poput brojeva, koji se prema mogućnostima sustava (alata, algoritama) mogu tipizirati ili kategorizirati.

[203] Ljubešić, N. et al.: Combining Available Datasets for Building Named Entity Recognition Models of Croatian and Slovene. Slovenščina 2.0, 2(2013) 1, str. 36.

Aplikacije namijenjene automatskom prepoznavanju nazivlja, ovisno o izvedbi, mogu nazive obrađivanih tekstova klasificirati u više vrsta ili klasa, a osnovne su osobe, organizacije i lokacije. Specijalizirani algoritmi, ovisno o izvedbi, u obzir mogu uzimati kategorije naziva poput vremenskih obilježja (datumi, godine, vremenski rasponi), proizvoda, oružja, umjetnina, ili sadržavati kategoriju razno koja služi za klasificiranje naziva koji ne nalaze mjesta u osnovnim tipovima. Tablica 14 donosi primjere označenoga teksta enciklopedičkih članaka iz tehničkoga područja zastupljenih u istraživanju alatom CroNER (poglavlje 5. 2. 1. 1.) koji prikazuju prisutnost pojedinih naziva.

Tablica 14. Kategorije naziva s pripadajućim primjerima relevantne za ispitivanje označene alatom naziva CroNER

Kategorije naziva	Oznake	Primjer
Osobe	PERSON	Ondje je <Person>D. Novak</Person> uzletio zrakoplovom, te postao prvim hrvatskim pilotom.
Organizacije	ORGANIZATION	… proglasen pocasnim doktorom <Organization>Sveucilista u Zagrebu</Organization>.
Lokacije	LOCATION	… u dvoristu izmedu <Location>Gunduliceve</Location> i <Location>Preradoviceve ulice</Location> s nazivom …
Geopolitički entiteti	GPE[204]	Povijesni razvoj visokoskolske nastave strojarstva u <Location>Hrvatskoj</Location> obiljezila je …
Vremenska obilježja	DATE; TIME; DURATION	God. <Date>2005.</Date> dovrsava se vijadukt preko doline rijeke …

Tipovi naziva vezani uz osobe u tekstu članaka javljaju se uz imena osoba, uključujući imena, prezimena, nazive obitelji, nadimaka, umjetničkih imena. Tipovi naziva vezani uz organizacije javljaju se uz nazive kompanija, upravnih tijela (vlade, agencije, ministarstva i dr.), obrazovnih institucija, udruga, sportskih klubova, muzeja, knjižnica i dr. U tu skupinu uključeni su i nazivi tiskovina, odnosno periodičnih publikacija (časopisa). Tip naziva lokacija odnosi se na nazive kao geopolitička obilježja (države, gradovi, savezne države, provincije) te nazive lokacija u koje spadaju npr. planinski masivi, obale, mora, jezera i dr. Vremenska obilježja odnose se na datume (godine) i vremenska razdoblja.

[204] Geopolitička obilježja poput naziva država, gradova, provincija i dr. uglavnom se odnose na lokacije, te su ih alati korišteni u ovom istraživanju označavali kao lokacije.

Problem koji se javlja pri procesu računaloga prepoznavanja naziva u tekstu jesu karakteristične formulacije javljanja naziva u prirodnom jeziku (tekstu) i utvrđenog skupa kategorija naziva određenog sustava koji ga analizira. S gledišta algoritama koji analiziraju tekst, to se odnosi na pojave više vrsta višeznačnosti prigodom kategorizacije naziva. Primjerice, pri ispitivanju uzorka tekstova na engleskom jeziku javlja se naziv *Jobs* što označava prezime (*Steve Jobs*), ali algoritam može pod istu kategoriju uvrstiti pojavnicu *jobs* (množina riječi *job*, hrv. *posao*) koja se ne razmatra ovim vrstama analiza. Također, pri utvrđivanju naziva, određeni naziv može se pridružiti u više kategorija. Primjerice, pri ispitivanju enciklopedičkoga članka uzorka na hrvatskom jeziku pojavljuje se naziv *Penkala*, koji se može odnositi na osobu (*Slavoljub Eduard Penkala*), ali i na poduzeće, odnosno organizaciju (*Penkala-Edmund Monster & Co.*). Taj naziv u istom se članku javlja više puta te je kategoriziran i kao osoba i kao organizacija, što upućuje na problem prepoznavanja konteksta od strane sustava.

Sustavi namijenjeni automatskom prepoznavanju nazivlja temelje se na prepoznavanju gramatičkih svojstava jezika za koji su namijenjeni i statističkim modelima poput strojnog učenja. Statistički modeli zahtijevaju veliku količinu ručno zabilježenih podataka (anotacija), te su prisutni poluautomatizirani pristupi kako bi se izbjegao dio napora. Pri strojnom učenju aplikacije koriste više tipova klasifikatora, od kojih je najzastupljeniji model uvjetnih nasumičnih polja (engl. *Conditional Random Fields*, CRF), odnosno metoda statističkoga modeliranja koja se često primjenjuje u prepoznavanju uzoraka i strojnom učenju te se koristi za strukturno predviđanje.[205] Takav softver predstavlja Stanford Named Entity Recognizer (Stanford NER)[206], koji je dostupan za potrebe znanstvene zajednice pod licencom GNU General Public Licence.

Prepoznavanje i klasificiranje naziva primjenjuje se primjerice za kategorizaciju novinskih članaka koji koriste portali namijenjeni informiranju, kako bi što bolje organizirali i iskoristili veliku količinu mrežno objavljenog sadržaja, zatim pri dizajniranju što učinkovitijih

[205] Više o modeliranju, implementaciji i primjeni CRF modela u radovima: Sutton, C., McCallum, A.: An Introduction to Conditional Random Fields. Foundations and Trend sin Machine Learning, 4(2012) 4, str. 267–373.; Laffetry, J., McCallun, A., Pereira F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Dana. U: Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001). San Francisco, 2001., str. 282–289.; Finkel, J., Grenager T., Manning, C.: Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. U: Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005). Stroudsburg, 2005., str. 363–370.

[206] https://nlp.stanford.edu/software/CRF-NER.shtml (22. 1. 2020.)

računalnih algoritama za pretraživanje, za analizu povratnih informacija klijenata pojedinih gospodarskih subjekata, ali i za strukturiranje (tagiranje) mrežnoga sadržaja na temelju označenih naziva. Važnije područje primjene automatskog prepoznavanja naziva jest izvedba sustava za preporučivanje sadržaja, odnosno vezanja relevantnih mrežnih entiteta (npr. novinskih članaka) na temelju zastupljenosti sličnih naziva.[207]

5.2.1. METODOLOGIJA

Kako bi se omogućilo ispitivanje alata namijenjenih automatskom prepoznavanju nazivlja, određen je reprezentativni uzorak sastavljen od jedinica sadržaja svih izdanja uključenih u analizu sadržaja (poglavlje 5. 2. 3. 1.). Tekstovi odabranih članaka ručno su označeni kako bi se te oznake mogle usporediti s izlaznim rezultatima alata korištenih u istraživanju.

Pri ručnom označavanju obuhvaćene su četiri najvažnije skupine podataka koji se pojavljuju u enciklopedičkim tekstovima, odnosno imena osoba, nazivi organizacija, podatci vezani uz prostorne odrednice (lokacije) te vremenska obilježja. Zahtjevniji kriteriji postavljeni su uz mogućnosti alata prigodom označavanja skupine naziva organizacija, u koju su se ubrajali nazivi poduzeća, ustanova, udruga, muzeja, časopisa i sl., te vezano uz nazive vremenskih obilježja koji se u enciklopedičkim tekstovima javljaju u većoj mjeri, a uključuju podatke vezane uz nadnevke, godine, vremenske raspone (npr. *1960–62.* ili *XIX. st.*), ali i specifična vremenska razdoblja poput termina *I.* ili *II. svj. rat, Domovinski rat,* odnosno onih koji se u vremenskom smislu mogu egzaktno odrediti.

Uzorak je sastavljen od dva dijela. Evaluacija alata CroNER i ReLDI provedena je na ručno označenome uzorku teksta na hrvatskom jeziku koji se sastojao od članaka Hrvatske tehničke enciklopedije, Hrvatske enciklopedije, Proleksis enciklopedije, Istarske enciklopedije i Hrvatskoga biografskog leksikona, dok su svojstva alata Stanford CoreNLP ispitivana na uzorku teksta na engleskom jeziku sastavljenom iz izdanja Encyclopaedije Britannice i engleske inačice Wikipedije.

[207] Više o primjeni sustava za automatsko prepoznavanje nazivlja u: Bokka, K. R. et al.: Deep Learning for Natural Language Processing: Solve your natural language processing problems with smart deep neural networks. Birmingham, 2019., str. 57–58.

Na temelju parametara koji karakteriziraju označene nazive u tekstu enciklopedičkih članaka (poglavlje 5. 2. 2.) doneseni su izračuni standardnih evaluacijskih mjera (preciznost, odziv, točnost, F mjera) za svaki alat zasebno razmatrajući svaku jedinicu sadržaja, kategorija enciklopedičkih članaka te vrste prepoznavanih naziva. Na temelju izračuna evaluacijskih mjera i primijećenih zakonitosti koje se javljaju prigodom rada alata na enciklopedičkim tekstovima, doneseni su zaključci o primjenjivosti dostupnih alata za prepoznavanje nazivlja u enciklopedičkim tekstovima.

5.2.1.1. Korišteni alati

Članci hrvatske enciklopedike ispitivani su s pomoću alata namijenjenog automatskom prepoznavanju nazivlja razvijenog posebno za potrebe hrvatskog jezika CroNER te također s pomoću mrežne aplikacije ReLDI. Za tekstove na engleskom jeziku korišten je alat naziva Stanford CoreNLP.

CroNER je proizvod stručnjaka i znanstvenika s Fakulteta elektrotehnike i računarstva u Zagrebu organiziranih u Laboratorij za analizu teksta i inženjerstvo znanja (TakeLab).[208] Nije javno dostupan, ali se može koristiti u istraživačke svrhe, poput izrade ove disertacije. Iako je razvijen ponajprije za hrvatski jezik, uspješno se primjenjuje i na tekstovima srodnih jezika, poput slovenskoga, srpskoga, bosanskoga, crnogorskoga i makedonskoga.

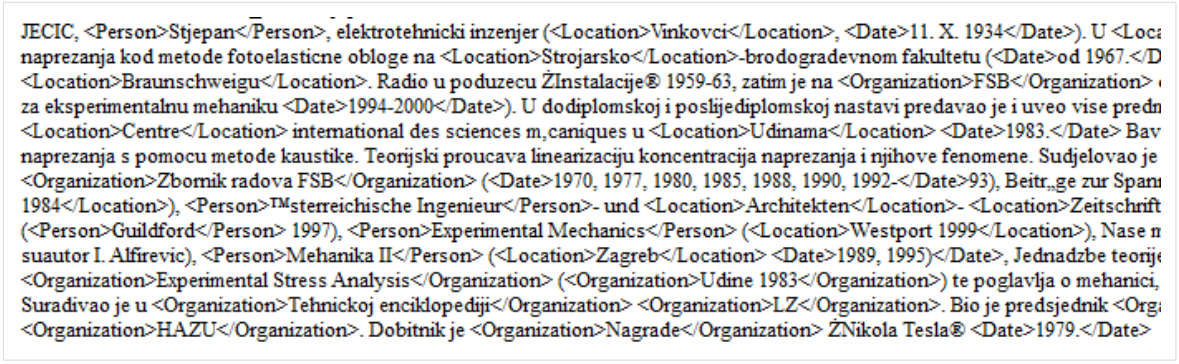
Za izgradnju (treniranje) CRF modela namijenjenog prepoznavanju naziva korišten je korpus koji sadrži ukupno 591 novinski članak s ukupno 310 000 pojavnica, proizašlih iz novina Vjesnik koji su objavljeni u razdoblju od 1999. do 2009. godine. Za anotaciju (označavanje) toga sadržaja upotrijebljeno je sedam standardnih tipova oznaka[209] u koje spadaju osobe, organizacije, lokacije, vremenska obilježja (vrijeme i datum kao dvije različite oznake), valuta i postotci. U konačnici je dodana i oznaka etničke pripadnosti, tako da je alat u konačnici osposobljen za prepoznavanje osam tipova naziva u prirodnome jeziku[210], a navedene

[208] http://takelab.fer.hr/ (16. 9. 2019.)

[209] Kako bi se pospješila ekstrakcija podataka, do sada je organizirano sedam skupova na kojima je začeta tema o prepoznavanju naziva (engl. *Message Understanding Conferences*, MUC) u cilju prihvaćanja jednakih standarda, odnosno da se pri strojnom prepoznavanju naziva u neoznačenome tekstu istaknu nazivi koji mogu biti značajni u procesu ekstrakcije informacija.

[210] Više o izvedbi CroNER alata za obradu prirodnog jezika u: Glavaš, G. et al.: CroNER: A State-of-the-Art Named Entity Recognition and Classification for Croatian. Proceedings of the Eighth Language Technologies Conference. Ljubljana, 2012., str. 73–78.

vrijednosti evaluacijskih mjera dobivene ispitivanjem na vrsti tekstova na kojima je razvijan iznose više od 90%. Slika 43. prikazuje označeni tekst nakon procesa automatskog prepoznavanja nazivlja na enciklopedičkom tekstu s pomoću alata CroNER.


JECIC, <Person>Stjepan</Person>, elektrotehnicki inzenjer (<Location>Vinkovci</Location>, <Date>11. X. 1934</Date>). U <Loca
naprezanja kod metode fotoelasticne obloge na <Location>Strojarsko</Location>-brodogradevnom fakultetu (<Date>od 1967.</D
<Location>Braunschweigu</Location>. Radio u poduzecu ŽInstalacije® 1959-63, zatim je na <Organization>FSB</Organization>
za eksperimentalnu mehaniku <Date>1994-2000</Date>). U dodiplomskoj i poslijediplomskoj nastavi predavao je i uveo vise predm
<Location>Centre</Location> international des sciences m,caniques u <Location>Udinama</Location> <Date>1983.</Date> Bav
naprezanja s pomocu metode kaustike. Teorijski proucava linearizaciju koncentracija naprezanja i njihove fenomene. Sudjelovao je
<Organization>Zbornik radova FSB</Organization> (<Date>1970, 1977, 1980, 1985, 1988, 1990, 1992-</Date>93), Beitr„ge zur Span
1984</Location>), <Person>™sterreichische Ingenieur</Person>- und <Location>Architekten</Location>- <Location>Zeitschrift
(<Person>Guildford</Person> 1997), <Person>Experimental Mechanics</Person> (<Location>Westport 1999</Location>), Nase m
suautor I. Alfirevic), <Person>Mehanika II</Person> (<Location>Zagreb</Location> <Date>1989, 1995)</Date>, Jednadzbe teorije
<Organization>Experimental Stress Analysis</Organization> (<Organization>Udine 1983</Organization>) te poglavlja o mehanici,
Suradivao je u <Organization>Tehnickoj enciklopediji</Organization> <Organization>LZ</Organization>. Bio je predsjednik <Orga
<Organization>HAZU</Organization>. Dobitnik je <Organization>Nagrade</Organization> ŽNikola Tesla® <Date>1979.</Date>

Slika 43. Primjer označenoga dijela teksta enciklopedijskoga članka Jecić, Stjepan nakon procesa automatskog prepoznavanja nazivlja s pomoću alata CroNER

Određena je i evaluacija alata ReLDI[211] koji je razvijen na Odsjeku za informacijske i komunikacijske znanosti Filozofskog fakulteta u Zagrebu pod vodstvom Nikole Ljubešića, a u projektu su sudjelovale Hrvatska, Srbija i Švicarska.[212] Dostupan je putem organizacije CLARIN.SI, odnosno slovenske članice europske istraživačke infrastrukture koja objedinjava digitalne jezične resurse i alate kroz jedinstveno mrežno okruženje kao podršku istraživačima u humanističkim i društvenim znanostima (engl. *Common Language Resources and Technology Infrastructure*, CLARIN)[213], a na poticaj Europske komisije u okviru inicijative razvoja europskih istraživačkih infrastruktura (engl. *European Research Infrastructures Consortium*, ERIC). Za njegov razvoj korišten je softver Stanford Named Entity Recognizer, a za njegovu izgradnju vezanu uz automatsko prepoznavanje naziva na hrvatskom jeziku

[211] Više o ReLDI alatu i njegovoj izvedbi u: Ljubešić, N. et al.: Combining Available Datasets for Building Named Entity Recognition Models of Croatian and Slovene. Slovenščina 2.0, 2(2013) 1, str. 35–57.

[212] Više o projektu Regional Linguistic Data Initiative (ReLDI) na adresi https://reldi.spur.uzh.ch/hr-sr/o-projektu/ (5. 6. 2020.)

[213] https://www.clarin.eu/ (16. 9. 2019.)

korištene su domene nacional.hr, jutarnji.hr, bug.hr i poslovni.hr [214] , zatim portala SETimes.com[215] te korpus proizašao iz novina Vjesnik.[216]

Uz automatsko prepoznavanje nazivlja, ReLDI omogućuje morfosintaktičko označavanje, lematizaciju te parsiranje univerzalnih ovisnosti na uzorku teksta. Pogodan je za ispitivanje hrvatskog, slovenskog i srpskog jezika. Prigodom pronalaženja i klasificiranja naziva koristi pet raspoloživih kategorija: osobe, izvedenice osoba, lokacije, organizacije i razno. F mjera (poglavlje 5. 2. 2. 4.) ove aplikacije kao evaluacijska mjera, prema navodima stručnjaka koji su ga razvijali, a navode podatke za slovenski jezik[217], iznosi 91% za osobe, 79% za lokacije te 57% za organizacije. Slika 44. prikazuje primjer označenoga dijela teksta enciklopedijskoga članka *Bazjanac, Davorin* nakon procesa automatskog prepoznavanja nazivlja s pomoću alata ReLDI.

[214] Građa preuzeta iz širokokorištenih domena .hr, sakupljena putem korpusa hrWaC. Korišteno je 59 212 pojavnica.

[215] Korpus SETimes sastavljen je od novinskih članaka objavljenih na portalu SETimes.com, koji objavljuje vijesti relevantne za područje jugoistočne Europe na deset jezika, od kojih i na hrvatskom. Sadrži 178 982 pojavnice.

[216] Korpus Vjesnik proizašao je iz tekstova koji pokrivaju širok raspon tema. Sadrži 104 494 pojavnice.

[217] https://www.clarin.si/info/k-centre/web-services-documentation/ (30. 5. 2020.)

	Surface	Tags	Lemma	Entity	Paragraph	Sentence	Token	Start char	End char
1.	Bazjanac	Npmsn	Bazjanac	B-per	1	1	1	1	8
2.	,	Z	,	O	1	1	2	9	9
3.	Davorin	Npmsn	Davorin	B-per	1	1	3	11	17
4.	,	Z	,	O	1	1	4	18	18
5.	hrvatski	Agpmsny	hrvatski	O	1	1	5	20	27
6.	inženjer	Ncmsn	inženjer	O	1	1	6	29	36
7.	strojarstva	Ncnsg	strojarstvo	O	1	1	7	38	48
8.	(	Z	(	O	1	1	8	50	50
9.	Beravci	Npmsn	Beravci	B-loc	1	1	9	51	57
10.	,	Z	,	O	1	1	10	58	58
11.	kraj	Ncmsn	kraj	O	1	1	11	60	63
12.	Slavonskog	Agpmsgy	slavonski	B-loc	1	1	12	65	74
13.	Broda	Ncmsg	brod	I-loc	1	1	13	76	80
14.	,	Z	,	O	1	1	14	81	81
15.	22.	Mdo	22.	O	1	1	15	83	85
16.	IX	Mrc	IX	O	1	2	1	87	88
17.	.	Z	.	O	1	2	2	89	89
18.	1902	Mdc	1902	O	1	2	3	91	94
19.	–	Z	–	O	1	2	4	96	96
20.	Zagreb	Npmsn	Zagreb	B-loc	1	2	5	98	103
21.	,	Z	,	O	1	2	6	104	104
22.	21.	Mdo	21.	O	1	2	7	106	108
23.	IX	Mrc	IX	O	1	3	1	110	111
24.	.	Z	.	O	1	3	2	112	112
25.	1988	Mdc	1988	O	1	3	3	114	117
26.	)	Z	)	O	1	3	4	118	118
27.	.	Z	.	O	1	3	5	119	119
28.	Diplomirao	Vmp-sm	diplomirati	O	1	4	1	121	130
29.	i	Cc	i	O	1	4	2	132	132
30.	doktorirao	Vmp-sm	doktorirati	O	1	4	3	134	143
31.	strojarstvo	Ncnsa	strojarstvo	O	1	4	4	145	155
32.	u	Sl	u	O	1	4	5	157	157
33.	Zürichu	Npmsl	Zürich	B-loc	1	4	6	159	165

Slika 44. Primjer označenoga dijela teksta enciklopedijskoga članka Bazjanac, Davorin nakon procesa automatskog prepoznavanja nazivlja s pomoću alata ReLDI

Alat Stanford CoreNLP[218] razvila je skupina stručnjaka i znanstvenika sa Sveučilišta u Stanfordu koja se bavi razvojem algoritama za obradu prirodnog jezika. Pruža većinu osnovnih postupaka obrade prirodnog jezika, od tokenizacije (podjele teksta elementarne jedinice sadržaja ili pojavnice) do pronalaženja svih izraza koji se odnose na isti naziv u tekstu.

Kao i ReLDI, objedinjuje više alata namijenjenih jezičnoj analizi prirodnog teksta, u koje spadaju označavanje vrsta riječi (engl. *Part-Of-Speech Tagger, POS Tagger*), automatsko prepoznavanje nazivlja, pronalaženje izraza i njihovih ovisnosti (engl. *Coreference Resolution System*), analiza sentimenta (engl. *Sentiment Analysis*) i ekstrakcija trojki koje predstavljaju relacije (odnose) u obliku subjekt-relacija-objekt (engl. *Open Information Extraction, OpenIE*). Uz to, algoritmi su, osim za engleski jezik, razvijeni i za ispitivanje tekstova na arapskom, kineskom, francuskom, njemačkom i španjolskom jeziku. Slikom 45 prikazan je primjer označenoga dijela teksta enciklopedijskoga članka Apple Inc. nakon procesa automatskog prepoznavanja nazivlja s pomoću alata Stanford CoreNLP.[219]

[218] Više o izvedbi i mogućnostima Stanford CoreNLP alata za obradu prirodnog teksta u: Manning, C. D. et al.: The Stanford CoreNLP Natural Language Processing Toolkit. U: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Stroudsburg, 2014., str. 55–60.

[219] Mrežnom sučelju alata Stanford CoreNLP omogućen je slobodan pristup. Više o projektu na https://nlp.stanford.edu/, a samo sučelje za rad dostupno je na http://corenlp.run/ (17. 9. 2019.).



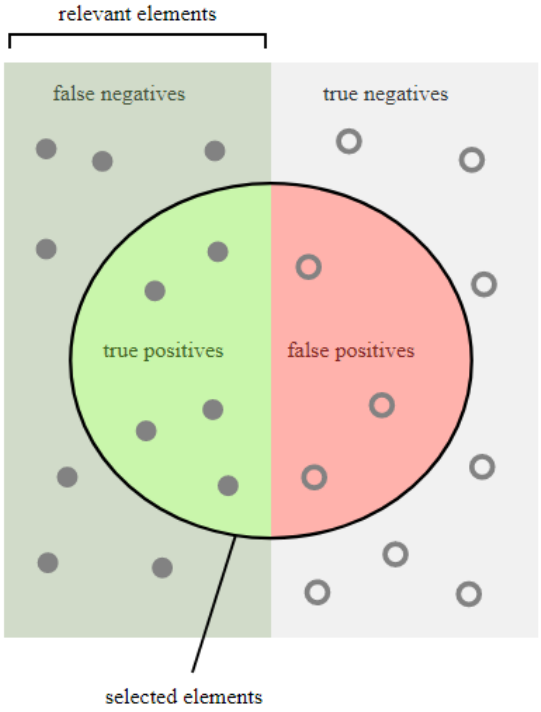
Slika 45. Primjer označenoga teksta dijela enciklopedijskoga članka Apple Inc. nakon procesa automatskog prepoznavanja nazivlja s pomoću alata Stanford CoreNLP

Alat Stanford CoreNLP koristi Stanford Named Entity Recognizer, odnosno kombinaciju triju CRF označitelja treniranih na više korpusa. Raznolikost prepoznavanja naziva uvjetovana je odabirom jezika na kojem se ispituje sadržaj. Za engleski prepoznaje imenske (osobe, lokacije, organizacije i razno) i vremenske (datum, vrijeme, razdoblje) nazive, a uključivanjem dodatnog označitelja proširuju se klase naziva na internetske adrese, gradove, države, nacionalnost, religiju, ideologije i dr. Time broj naziva vezanih uz engleski jezik koje ovaj alat može prepoznati iznosi ukupno 23.

5.2.2. EVALUACIJSKE MJERE

Kako bi se mogle izračunati vrijednosti evaluacijskih mjera korištenih alata namijenjenih automatskom prepoznavanju nazivlja, potrebno je usporediti nazive prepoznate s pomoću alata u odnosu na ručno označeni tekst donesen analizom sadržaja za određeni uzorak. Evaluacija određenog alata određuje se njegovom mogućnošću da relevantne nazive obuhvaćene istraživanjem označi ispravno (engl. *True Positives*, TP) ili ih izostavi (engl. *False Negatives*, FN), ali i da one koje ne bi trebao razmatrati također uključi traženim oznakama

naziva (engl. *False Positives*, FP) ili ih izostavi, što također djeluje pozitivno na njegovu izvedbu (engl. *True Negatives*, TN). Slika 46. donosi odnos svih elemenata (naziva) u nekom uzorku koji utječu na izvedbu sustava za automatsko prepoznavanje nazivlja.



Slika 46. Grafički prikaz odnosa svih elemenata uzorka koji utječu na izvedbu sustava za automatsko prepoznavanje nazivlja (Izvor: https://en.wikipedia.org/wiki/Precision_and_recall)

Na temelju navedenih parametara koji predstavljaju usporedbu između ručno i strojno označenih naziva u tekstu, izračunavaju se vrijednosti preciznosti, odziva, točnosti i F mjere kojima se vrednuje rad alata na određenom uzorku.

5.2.2.1. Preciznost (P)

U prepoznavanju naziva, dohvaćanju informacija i klasifikaciji, preciznost predstavlja omjer relevantnih označenih elemenata i ukupnog broja dohvaćenih elemenata.

$$P = \frac{TP}{\mathrm{TP} + \mathrm{FP}}$$

U ovom slučaju preciznost predstavlja omjer ispravno dohvaćenih naziva u tekstovima enciklopedičkih članaka uzorka i svih označenih naziva. Ako su svi nazivi koji su prepoznati

relevantni (P = 1), ne postoji informacija jesu li svi relevantni dohvaćeni, što predstavlja nedostatak ove mjere ako se razmatra zasebno.

5.2.2.2. Odziv (R)

Odziv kao mjera evaluacije sustava za dohvaćanje elemenata (informacija) predstavlja omjer relevantnih označenih elemenata i ukupnog broja relevantnih elemenata.

$$R = \frac{\mathrm{TP}}{TP + FN}$$

U ovom slučaju predstavlja omjer ispravno označenih naziva i svih relevantnih naziva koji su trebali biti označeni. Ako je vrijednost odziva najviša moguća (R = 1) svi su relevantni nazivi dohvaćeni, ali nije moguće znati koliko je nerelevantnih naziva (pogrešno označenih) također uključeno u izračun.

5.2.2.3. Točnost (A)

Točnost predstavlja određivanje udjela elemenata u nekom uzorku koji su točno označeni, razmatrajući i relevantne i nerelevantne elemente, odnosno sve elemente uključene u ispitivanje.

$$A = \frac{TP + TN}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}$$

To znači da ova mjera, za razliku od preciznosti i odziva koji razmatraju samo pozitivne primjere i predviđanja, razmatra i mogućnosti alata u prepoznavanju negativnih, nerelevantnih slučajeva vezanih uz prepoznavanje naziva. U tom smislu predstavlja potpuniju mjeru pri donošenju zaključaka o evaluaciji alata.

Međutim, mjera točnosti nije prikladna za analiziranje rezultata svih skupova podataka, odnosno ne može poslužiti kao mjerodavna u svim vrstama uzoraka. Ako većina elemenata podvrgnuta vrednovanju pripada nerelevantnoj skupini, čak i dobri sustavi prepoznavanja rezultata neke će nazive svrstati u relevantnu skupinu, što će rezultirati većim brojem pogrešno označenih naziva (FP) koja će narušiti konačan rezultat točnosti, iako je zapravo izvedba sustava na puno višoj razini za velik broj drugih uzoraka drugačijeg omjera elemenata.

5.2.2.4. F mjera (F1)

Kako je mjera točnosti uvjetovana omjerom elemenata uzorka, ti nedostatci nadilaze se izračunom F mjere, koja predstavlja harmonijsku sredinu preciznosti i odziva.

$$F1 = \frac{2 * \mathrm{P} * \mathrm{R}}{\mathrm{P} + \mathrm{R}}$$

F mjera predstavlja siguran pokazatelj evaluacije alata, budući da su mjere preciznosti i odziva obrnuto proporcionalne. Maksimalan odziv može se postići jednostavnim dohvaćanjem svih elemenata, ili u ovom slučaju naziva svih vrsta, međutim, povećanjem uključenih naziva u ispitivanje pada preciznost. Sukladno tomu, sustav za automatsko prepoznavanje nazivlja učinkovit je ako uz pouzdan odziv naziva uključuje i manji udio pogrešno označenih naziva koji se mogu tolerirati, što prikazuje jedinstvena F mjera.

5.2.3. REZULTATI ISTRAŽIVANJA

5.2.3.1. Analiza uzorka

Kako je potrebno provesti ispitivanje alata namijenjenih tekstovima na hrvatskom i engleskom jeziku, uzorak je podijeljen na dva dijela, a svaki uključuje sve kategorije članaka uključene u analizu sadržaja. Uzorak namijenjen ispitivanju na hrvatskom jeziku sadržavao je 15, a na engleskom jeziku sedam članaka. Pri tome, broj pojavnica obaju uzoraka bio je približno jednak, kako međusobno tako i pojedinačnih kategorija unutar zasebnoga uzorka. Uzorak na hrvatskom jeziku činilo je ukupno 7720, a na engleskom jeziku 8125 pojavnica svih triju kategorija enciklopedičkih članaka uključenih u istraživanje (tablice 15 i 16).

Pojavnice obuhvaćene uzorkom ručno su označene oznakama za četiri vrste naziva koji su najznačajniji u zastupanju unificiranih faktografskih podataka u enciklopedičkim tekstovima iz tehničkoga područja, odnosno kao osobe (OS), organizacije (ORG), lokacije (LOK) i vremenska obilježja (VO). Pri tome je važno kazati kako više pojavnica može sačinjavati jedan naziv, primjerice dio teksta *Davorin Bazjanac* čine dvije, a *Fakultet strojarstva i brodogradnje* četiri pojavnice koje predstavljaju jedan naziv. Nakon ispitivanja enciklopedičkih tekstova s pomoću alata uključenih u istraživanje, njihovi rezultati (označeni tekst) uspoređivali su se s ručno označenim tekstom uzorka. Neki alati korišteni u ovom istraživanju imaju mogućnost prepoznavanja i drugih tipova i vrsta naziva, ali oni zbog manje relevantnosti za ovo istraživanje

i općenito faktografiju enciklopedičkih tekstova iz tehničkoga područja nisu uključeni u vrednovanje i nisu se razmatrali u konačnim rezultatima.

Tako je za uzorak enciklopedičkih tekstova na hrvatskom jeziku kategorija osoba uključena s ukupno 2413 pojavnice, organizacije s 2672 pojavnice, a općih pojmova s 2635. Tablicom 15 prikazan je broj naziva svake vrste za pojedinu kategoriju enciklopedičkoga članka, ali i za zasebne jedinice sadržaja.

Tablica 15. Broj ručno označenih naziva prema kategorijama i jedinicama sadržaja enciklopedičkih članaka, vrstama i tipovima naziva i postotak njihove zastupljenosti u tekstu ovisno o broju pojavnica za svaku kategoriju i jedinicu sadržaja uzorka na hrvatskom jeziku

Natuknica (izdanje)		Nazivi				Ukupno	Postotak u tekstu	Pojavnica u tekstu
		OS	ORG	LOK	VO			
osobe	Bazjanac, Davorin (PRO)	1	1	5	7	14	13,3	105
	Lugnani, Giuseppe de (IE)	1	10	10	17	38	13,8	275
	Bošnjaković, Fran (HTE)	2	21	16	27	66	11,4	580
	Penkala, Slavoljub Eduard (HE)	10	2	13	13	38	7,1	538
	Ivan Krstitelj Rabljanin (HBL)	29	6	29	32	96	10,5	915
	Ukupno	43	40	73	96	252	10,4	2413
organizacije	Društvo za plastiku i gumu (HTE)	1	10	1	5	17	9,6	177
	Brodarski institut(HE)	0	6	2	5	13	8,3	156
	Citroën SA (PRO)	3	6	1	10	20	13,0	154
	Tvornica duhana Rovinj d. o. o. (IE)	1	9	4	13	27	6,9	392
	Tehnički fakultet u Rijeci (HTE)	36	54	6	85	181	10,1	1793
	Ukupno	41	85	14	118	258	9,7	2672
opći pojmovi	cement (PRO)	0	0	3	1	4	2,4	164
	trabakul (HTE)	0	0	2	3	5	1,3	394
	površinski kop (HE)	0	0	0	0	0	0,0	279
	most (IE)	0	0	53	26	79	9,8	804
	računalo (HE)	16	0	2	21	39	3,9	994
	Ukupno	16	0	60	51	127	4,8	2635
Ukupno		100	125	147	265	637	8,3	7720

Udio naziva u tekstu uzorka najveći je za kategoriju osoba (10,4%), slijede organizacije s 9,7% i opći pojmovi s 4,8%. Najveći broj naziva od njih ukupno 637 ipak je zabilježen kod kategorije organizacija (258), zatim kod osoba (252) i općih pojmova (127). Kad je riječ o vrstama naziva, najzastupljenija su vremenska obilježja (265), zatim lokacije (147), organizacije (125) te osobe (100).

Sukladno analizi uzorka na hrvatskom jeziku, provedena je i analiza uzorka na engleskom jeziku koji je poslužio za vrednovanje alata Sanford CoreNLP na enciklopedičkim tekstovima (tablica 16).

Tablica 16. Broj ručno označenih naziva prema kategorijama i jedinicama sadržaja enciklopedičkih članaka, vrstama i tipovima naziva i postotak njihove zastupljenosti u tekstu ovisno o broju pojavnica za svaku kategoriju i jedinicu sadržaja uzorka na engleskom jeziku

Natuknica (izdanje)		Nazivi				Ukupno	Postotak u tekstu	Pojavnica u tekstu
		OS	ORG	LOK	VO			
osobe	Eads, James B. (EB)	22	2	35	10	69	5,2	1318
	Robert Whitehead (W_{eng})	69	18	36	38	161	10,2	1576
	Ukupno	91	20	71	48	230	7,9	2894
organizacije	The International Society for Knowledge Organization (W_{eng})	1	19	20	4	44	10,3	426
	Intel (EB)	25	81	6	32	144	8,6	1682
	California Institute of Technology (EB)	5	8	12	5	30	9,8	305
	Ukupno	31	108	38	41	218	9,0	2413
opći pojmovi	Wi-Fi (EB)	1	4	0	4	9	1,7	530
	Pneumatics (W_{eng})	1	0	1	3	5	0,5	1068
	Bionics (EB)	0	0	0	1	1	0,1	1220
	Ukupno	2	4	1	8	15	0,5	2818
Ukupno		124	132	110	97	463	5,7	8125

Uzorak proizašao iz izdanja Encyclopaedije Britannice i engleske inačice Wikipedije donosi ukupno 2894 pojavnice vezane uz kategoriju članaka osoba, 2413 vezane uz kategoriju članaka organizacija te 2818 vezanih uz kategoriju članaka općih pojmova. Zastupljenost naziva najveća je kod organizacija (9%) i osoba (7,9%), dok je kod kategorije članaka općih pojmova gotovo izostala (0,5%). Općenito je broj naziva manji (463) i podjednako raspoređen po vrstama, te se kreće oko 100 za kategorije članaka osoba i organizacija.

Analiza uzorka na hrvatskom i engleskom jeziku upućuje na to kako je za vrednovanje i moguću prilagodbu i izgradnju alata za automatsko prepoznavanje nazivlja namijenjenoga enciklopedičkim tekstovima iz tehničkoga područja ključno razmatrati korpuse enciklopedičkih tekstova koji se klasificiraju u kategorije osoba i organizacija, budući da članci općih pojmova u informativnom smislu donošenja naziva znatno zaostaju. Na to upućuju rezultati analize sadržaja cjelokupnog uzorka, odnosno prisutnih sedam tipova podataka ili u ovom slučaju naziva u enciklopedičkim tekstovima članaka općega tipa sa smanjenom zastupljenošću (tablica 13).

5.2.3.2. Ispitivanje alata

Kako bi se dobio što bolji uvid u mogućnosti korištenja alata namijenjenih automatskom prepoznavanju nazivlja na enciklopedičkim tekstovima, doneseni su rezultati kao mjerilo evaluacije za zasebne jedinice sadržaja, pojedine kategorije članaka te za označavanje pojedinih vrsta naziva u jedinstvenom uzorku za hrvatski i engleski jezik. Uz evaluacijske mjere navedeni su i parametri karakterizacije označenih naziva.

5.2.3.2.1. Kategorije članaka i zasebne jedinice sadržaja

CroNER je korišten je za ispitivanje četiriju vrsta naziva (osobe, organizacije, lokacije i vremenska obilježja), dok ReLDI ne sadrži mogućnost razmatranja vremenskih obilježja, što je u konačnici rezultiralo manjim brojem ispitanih naziva u istom uzorku.

CroNER pokazuje podjednake iznose mjera za pojedine kategorije članaka. U analizu su bile uključene sve četiri vrste naziva te je pri tome preciznost ispitanih naziva za kategorije članaka iznosila oko 80%, odziv neznatno manje, točnost 67%, a F mjera 78%. Sukladno tomu, rezultati evaluacijskih mjera CroNER alata na cjelokupnom uzorku kretali su se u tim vrijednostima (tablica 17).

Tablica 17. Rezultati ispitivanja evaluacijskih mjera CroNER alata na enciklopedičkome tekstu

CroNER									
Natuknica (izdanje)		Evaluacijske mjere				Vrednovanje naziva			
		P	R	A	F1	TP	FP	FN	TN
osobe	Bazjanac, Davorin (PRO)	0,80	0,67	0,57	0,73	8	2	4	0
	Lugnani, Giuseppe de (IE)	0,73	0,63	0,54	0,68	19	7	11	2
	Bošnjaković, Fran (HTE)	0,93	0,90	0,86	0,92	54	4	6	8
	Penkala, Slavoljub Eduard (HE)	0,81	0,91	0,76	0,86	30	7	3	2
	Ivan Krstitelj Rabljanin (HBL)	0,73	0,68	0,58	0,71	61	22	29	10
	Ukupno	0,80	0,76	0,67	0,78	172	42	53	22
organizacije	Društvo za plastiku i gumu (HTE)	0,89	0,85	0,78	0,87	17	2	3	1
	Brodarski institut(HE)	0,67	0,67	0,54	0,67	6	3	3	1
	Citroën SA (PRO)	0,67	0,71	0,55	0,69	10	5	4	1
	Tvornica duhana Rovinj d. o. o. (IE)	0,60	0,68	0,53	0,64	15	10	7	4
	Tehnički fakultet u Rijeci (HTE)	0,83	0,78	0,70	0,80	134	27	38	17
	Ukupno	0,79	0,77	0,67	0,78	182	47	55	24
opći pojmovi	cement (PRO)	1,00	0,75	0,80	0,86	3	0	1	1
	trabakul (HTE)	0,33	0,50	0,50	0,40	1	2	1	2
	površinski kop (HE)	–	–	–	–	0	0	0	0
	most (IE)	0,85	0,77	0,70	0,81	56	10	17	7
	računalo (HE)	0,64	0,87	0,61	0,74	27	15	4	3
	Ukupno	0,76	0,79	0,67	0,78	87	27	23	13
Ukupno		0,79	0,77	0,67	0,78	441	116	131	59

ReLDI pokazuje veća odstupanja navedenih mjera za pojedinu kategoriju enciklopedičkoga članka. Preciznost je najniža za kategoriju osoba (69%), veća za kategoriju organizacija (75%) te najveća kod općih pojmova (80%). Za razliku od CroNER-a, ReLDI pokazuje veći odziv koji kod osoba iznosi i 91%. Točnost je i u ovom slučaju manja, oko 65% za osobe i organizacije te 76% za opće pojmove. Sukladno tim rezultatima, ukupne vrijednosti evaluacijskih mjera za ispitane kategorije enciklopedičkih članaka kreću se od 75 do 84%. Ukupni rezultat približan je CroNER-u, uz nešto veći odziv, ali i manju preciznost (tablica 18).

Tablica 18. Rezultati ispitivanja evaluacijskih mjera ReLDI alata na enciklopedičkome tekstu

ReLDI									
Natuknica (izdanje)		Evaluacijske mjere				Vrednovanje entiteta			
		P	R	A	F1	TP	FP	FN	TN
osobe	Bazjanac, Davorin (PRO)	0,67	1,00	0,67	0,80	6	3	0	0
	Lugnani, Giuseppe de (IE)	0,39	0,82	0,41	0,53	9	14	2	2
	Bošnjaković, Fran (HTE)	0,71	0,94	0,72	0,81	30	12	2	6
	Penkala, Slavoljub Eduard (HE)	0,84	0,88	0,75	0,86	21	4	3	0
	Ivan Krstitelj Rabljanin (HBL)	0,72	0,92	0,71	0,81	57	22	5	8
	Ukupno	0,69	0,91	0,67	0,79	123	55	12	16
organizacije	Društvo za plastiku i gumu (HTE)	0,82	0,64	0,63	0,72	9	2	5	3
	Brodarski institut(HE)	0,67	0,57	0,50	0,62	4	2	3	1
	Citroën SA (PRO)	0,90	1,00	0,91	0,95	9	1	0	1
	Tvornica duhana Rovinj d. o. o. (IE)	0,53	0,83	0,54	0,65	10	9	2	3
	Tehnički fakultet u Rijeci (HTE)	0,77	0,74	0,68	0,76	71	21	25	25
	Ukupno	0,75	0,75	0,66	0,75	103	35	35	33
opći pojmovi	cement (PRO)	0,75	1,00	0,75	0,86	3	1	0	0
	trabakul (HTE)	0,25	1,00	0,40	0,40	1	3	0	1
	površinski kop (HE)	–	–	–	–	0	0	0	0
	most (IE)	0,87	0,89	0,79	0,88	47	7	6	2
	računalo (HE)	0,74	0,88	0,75	0,80	14	5	2	7
	Ukupno	0,80	0,89	0,76	0,84	65	16	8	10
Ukupno		0,73	0,84	0,68	0,78	291	106	55	59

Najučinkovitijim se pokazao alat Stanford CoreNLP na uzorku teksta na engleskom jeziku. To se posebno pokazalo na rezultatima kategorije organizacija, gdje sve vrijednosti evaluacijskih mjera iznose više od 90%. Kategorija osoba također pokazuje visoke vrijednosti vezane za odziv (96%) te F mjeru (91%) uz vrijednosti preciznosti i točnosti više od 85%. U rasponu od 80 do 90% kreću se vrijednosti mjera vezanih uz opće pojmove, a ukupan rezultat Stanford CoreNLP-a kreće se za sve vrijednosti oko 90%, uz odziv naziva od 96% (tablica 19).

Tablica 19. Rezultati ispitivanja evaluacijskih mjera Stanford CoreNLP alata na enciklopedičkome tekstu

Stanford CoreNLP									
Natuknica (izdanje)		Evaluacijske mjere				Vrednovanje entiteta			
		P	R	A	F1	TP	FP	FN	TN
osobe	Eads, James B. (EB)	0,77	0,95	0,77	0,85	58	17	3	10
	Robert Whitehead (W_{eng})	0,89	0,97	0,89	0,93	151	18	5	30
	Ukupno	0,86	0,96	0,85	0,91	209	35	8	40
organizacije	The International Society for Knowledge Organization (W_{eng})	0,96	1,00	0,96	0,98	43	2	0	2
	Intel (EB)	0,90	0,98	0,91	0,94	149	16	3	38
	California Institute of Technology (EB)	0,94	1,00	0,94	0,97	31	2	0	2
	Ukupno	0,92	0,99	0,92	0,95	223	20	3	42
opći pojmovi	Wi-Fi (EB)	0,90	0,75	0,78	0,82	9	1	3	5
	Pneumatics (W_{eng})	0,86	0,80	0,78	0,83	12	2	3	6
	Bionics (EB)	1,00	1,00	1,00	1,00	1	0	0	1
	Ukupno	0,88	0,79	0,79	0,83	22	3	6	12
Ukupno		0,89	0,96	0,88	0,92	454	58	17	94

Analizom enciklopedičkih tekstova u vezi s ispitivanjem alata namijenjenih automatskom prepoznavanju nazivlja uočeno je nekoliko pojava koje utječu na ipak smanjene vrijednosti evaluacijskih mjera, pogotovo u odnosu na tekstove općega tipa na kojima su razvijani i na kojima ostvaruju vrijednosti istih veće od 90% prema testiranjima stručnjaka koji su ih razvili.

Smanjene vrijednosti preciznosti vezane su uza specifične formulacije enciklopedičkih tekstova koje aplikacije ovog tipa krivo procjenjuju. Riječ je o svojevrsnom metajeziku koji je zastupljen u enciklopedici, poput korištenja rimskih brojki, korištenja raznih kratica, ali i konstrukcije rečenica nerijetko su drugačije nego u tekstovima drugog tipa. Budući da su korišteni alati namijenjeni ispitivanju uzorka teksta na hrvatskom jeziku razvijani na korpusima koji sadrže npr. novinske članke, u ovom stanju ne sadržavaju dovoljan broj anotacija kako bi učinkovito označavali nazive u enciklopedičkim tekstovima. To je vidljivo u smanjenim vrijednostima odziva, gdje znatan broj naziva ostaje neoznačen. Sukladno ovim zaključcima, smanjene su i vrijednosti točnosti i F mjere za ove uzorke.

U prilog ovome idu i bolji ostvareni rezultati alata Stanford CoreNLP koji je razvijan na više korpusa i sadrži bogatije baze podataka kojima se koristi u svome radu te također prepoznaje puno više tipova naziva. Jedna od mogućnosti boljega postizanja automatskog označavanja nazivlja na enciklopedičkim tekstovima bila bi dodatna nadogradnja postojećih aplikacija (alata) s pomoću enciklopedičkih sadržaja kako bi se proširile njihove mogućnosti označavanja naziva koje takav sadržaj donosi. Također, potrebno je razmotriti i jednostavniji način sastavljanja enciklopedičkih tekstova, odnosno približiti ga formi kojom se donose drugi sadržaji općega tipa na kojima su alati razvijani, u cilju što boljega funkcioniranja već razvijenih aplikacija na enciklopedičkome sadržaju. Na taj način omogućilo bi se bolje iskorištavanje već postojećih, slobodno dostupnih resursa s gledišta enciklopedike.

Tome u prilog idu i izračuni evaluacijskih mjera za zasebne jedinice sadržaja u oba uzorka. Primjerice, članci *Bošnjaković, Fran* Hrvatske tehničke enciklopedije (CroNER) ili *Citroën SA* Proleksis enciklopedije (ReLDI) uvelike nadmašuju rezultate mjera vezanih uz druge jedinice sadržaja u uzorku (više od 90%). To znači da i u ovom stanju alati mogu odraditi prepoznavanje nazivlja na razini i većoj od tekstova na kojima su razvijani, ako izostanu njihovi nedostatci nepoznavanja i krivog tumačenja naziva.

Smanjene vrijednosti evaluacijskih mjera alata također su rezultat problema dvoznačnosti naziva, odnosno pogrešnog kategoriziranja naziva u posebnim slučajevima. Primjeri toga su formulacije u obliku *Nagrada Nikola Tesla* ili *Međunarodni simpozij Peter Salcher & Ernst Mach* kojih su pojedine pojavnice pogrešno svrstane pod kategoriju naziva osoba, iako je očito da nije riječ o osobama, nego o nagradi i međunarodnome znanstvenom skupu. Međutim, problem semantičkoga razlučivanja temeljni je nedostatak ovog područja, a u ovom slučaju marginalno je utjecao na ukupne rezultate jer se u većini jedinica sadržaja niti ne pojavljuje, a neki primjeri su i točno kategorizirani. Primjer članka gdje su se istaknuli ovakvi slučajevi jest *Tehnički fakultet u Rijeci*. Vrednovanjem ovih slučajeva u korist alata (ReLDI) ispitivanjem teksta tog članka preciznost bi iznosila 89%, točnost 75%, a F mjera 82%.

5.2.3.2.2. Nazivi

Sukladno analizi pojedinih kategorija i zasebnih članaka, vrijednosti preciznosti, odziva, točnosti i F mjere za pojedine vrste naziva ispitanih tekstova na hrvatskom jeziku kreću se u manjim vrijednostima od očekivanih prije ispitivanja. Ističe se visoka preciznost CroNER-a kod prepoznavanja vremenskih obilježja (93%), međutim uz smanjeni odziv koji tomu

pridonosi. Posebno se ističe evaluacijska mjera kao kombinacija preciznosti i odziva (F mjera) označavanja naziva lokacija (97%) i vremenskih obilježja (94%) alata Stanford CoreNLP, uz ionako visoku pouzdanost pri ispitivanju enciklopedičkih tekstova. ReLDI pokazuje visoke vrijednosti odziva za osobe (95%) i lokacije (93%) što u nekim vrstama ispitivanja također predstavlja vrijednu odliku. Statistički prikaz vrednovanja svih vrsta naziva u tekstovima uzorka prikazan je tablicom 20.

Tablica 20. Rezultati ispitivanja evaluacijskih mjera za određene vrste naziva alata CroNER, ReLDI i Stanford CoreNLP

Alat	Vrsta naziva	Evaluacijske mjere				Vrednovanje entiteta			
		P	R	A	F1	TP	FP	FN	TN
CroNER	osobe	0,65	0,82	0,59	0,73	74	40	16	6
	organizacije	0,57	0,80	0,55	0,66	94	72	24	23
	lokacije	0,77	0,89	0,71	0,83	126	37	16	6
	vremenska obilježja	0,93	0,68	0,65	0,78	143	11	68	5
	Ukupno	0,73	0,78	0,63	0,75	437	160	124	40
ReLDI	osobe	0,66	0,95	0,64	0,78	97	51	5	2
	organizacije	0,67	0,66	0,60	0,67	78	38	40	37
	lokacije	0,76	0,93	0,74	0,84	115	36	8	9
	Ukupno	0,70	0,85	0,66	0,77	290	125	53	48
CoreNLP	osobe	0,82	0,99	0,83	0,90	112	24	1	14
	organizacije	0,86	0,91	0,84	0,88	138	23	13	51
	lokacije	0,94	1,00	0,94	0,97	108	7	0	9
	vremenska obilježja	0,89	1,00	0,91	0,94	98	12	0	23
	Ukupno	0,87	0,97	0,87	0,92	456	66	14	97

Moguća minimalna razlikovanja rezultata koji se odnose na ukupne rezultate ispitivanja uzorka za pojedini alat u odnosu na rezultate prikazane tablicama 17, 18 i 19 posljedica su dvostrukog vrednovanja naziva u pojedinim slučajevima pri ispitivanju evaluacijskih mjera za određene vrste entiteta. Npr, označeni tekst alata CroNER sadržava naziv označen na način *<Organization>Penkala</Organization>* koji se razmatra i kod vrednovanja vrsta naziva osoba i organizacija, budući da je riječ o pogrešnoj oznaci osobe koja se klasificira kao organizacija. Budući da takvih slučajeva dvostrukog vrednovanja nema kod vrednovanja evaluacijskih mjera zasebnih jedinica sadržaja gdje se takvi nazivi razmatraju jednom, moguće su razlike u iznosima parametara koji su uključeni u izračune mjera.

Ova analiza može poslužiti kao mjerodavna za najbolji odabir alata ako je predmet dohvaćanja određena vrsta naziva. Budući da je provedena detaljna kvantitativna analiza sadržaja (prilozi 1−10), moguće je utvrditi koja je vrsta naziva najzastupljenija u određenoj kategoriji enciklopedičkoga članka za određena izdanja te upotrijebiti alat koji pokazuje najbolje vrijednosti prigodom njihova označavanja (prepoznavanja).

6. USPOSTAVA MODELA INTEROPERABILNOSTI MREŽNOGA ENCIKLOPEDIČKOG SADRŽAJA IZ TEHNIČKOGA PODRUČJA

Pojam model nosi više značenja i može se različito definirati ovisno o gledištu pojedinoga područja ljudskoga djelovanja. U širem smislu predstavlja oblik, obrazac, uzorak, predložak, uzor ili primjer planiranoga sustava ili funkcionalne cjeline, napravljen u svrhu njihova prikazivanja i daljnjeg proučavanja, boljega razumijevanja i eksperimentiranja. Modeli se mogu implementirati (provesti, izvršiti, uvrstiti, primijeniti) kao konceptualni (shematski prikaz, opis riječima), matematički (sustav jednadžbi) i simulacijski (sustav varijabli upravljanjem kojima se prikazuju svojstva ili ponašanje nekog fizičkog ili apstraktnog sustava).

U svrhu ovoga istraživanja relevantna je definicija u znanstvenom smislu, koja opisuje model kao skup pretpostavki s pomoću kojih se teorijski opisuje neki sustav te se sastoji od opće teorije i opisa objekta ili sustava na koji se teorija odnosi.[220]

Model enciklopedičke interoperabilnosti implementiran na temelju rezultata istraživanja ove disertacije predstavlja simulacijski model (prototip) izveden u obliku računalnog sustava u cilju oponašanja potencijalnog sustava enciklopedičke interoperabilnosti primjenjivog u praksi. Manipulacijom strukturnih elemenata omogućava simuliranje situacija s pomoću promjenjivih ulaznih podataka u koje može biti doveden planirani sustav, odnosno provođenje eksperimentalnog ispitivanja.

6.1. METODOLOGIJA IZRADE MODELA

Model enciklopedičke interoperabilnosti zasniva se na mogućnosti informatičkih sustava za međusobnu razmjenu metapodataka, koji predstavljaju vrijednosti unificiranih elemenata strukture više mrežnih enciklopedičkih projekata. Stoga, temelj izrade modela kao

[220] Prema natuknici: model. *Hrvatska enciklopedija, mrežno izdanje.* Leksikografski zavod Miroslav Krleža, 2020. <http://www.enciklopedija.hr/Natuknica.aspx?ID=41453>. (1. 5. 2020.)

središnjeg sustava (baze podataka) čini analiza sadržaja enciklopedičkih djela u cilju bilježenja unificiranih faktografskih podataka pogodnih za stvaranje strukturnih elemenata.

Razvijen je prototip modela, koji omogućuje razmjenu podataka (metapodataka) među izdanjima Hrvatske tehničke enciklopedije, Proleksis enciklopedije i Encyclopaedije Britannice, odnosno uzorka koji čine njihovi članci koji su također bili ispitani analizom sadržaja enciklopedičkih tekstova ovim istraživanjem. Svako zasebno izdanje predstavlja bazu podataka, a stvaranje sadržaja omogućeno je posebno razvijenim sučeljima za ovu namjenu.

Kako je svrha modela omogućavanje interoperabilnosti podataka među enciklopedičkim projektima, od kojih je svaki strukturiran u određenoj mjeri i na specifičan način opisuje strukturne elemente, s gledišta modela interoperabilnosti enciklopedičkoga sadržaja ključno je uključivanje svih vrijednosti strukturnih elemenata kako bi njima mogli raspolagati svi mrežni enciklopedički projekti. Sukladno tomu, model enciklopedičke interoperabilnosti čini središnju bazu podataka koja donosi uniju vrijednosti strukturnih elemenata svih povezanih mrežnih izdanja, uz mogućnost proširenja na druga područja. Model predstavlja sustav kojim upravlja administrator, odnosno nadzire njegov rad, pa je razvijeno posebno sučelje koje to omogućuje.

Testiranjem modela, odnosno uvođenjem novih vrijednosti elemenata strukture, dokazano je postojanje interoperabilnosti među zasebnim mrežnim enciklopedičkim sadržajima uključenima u njegovu izvedbu.

6.2. IMPLEMENTACIJA MODELA ENCIKLOPEDIČKE INTEROPERABILNOSTI

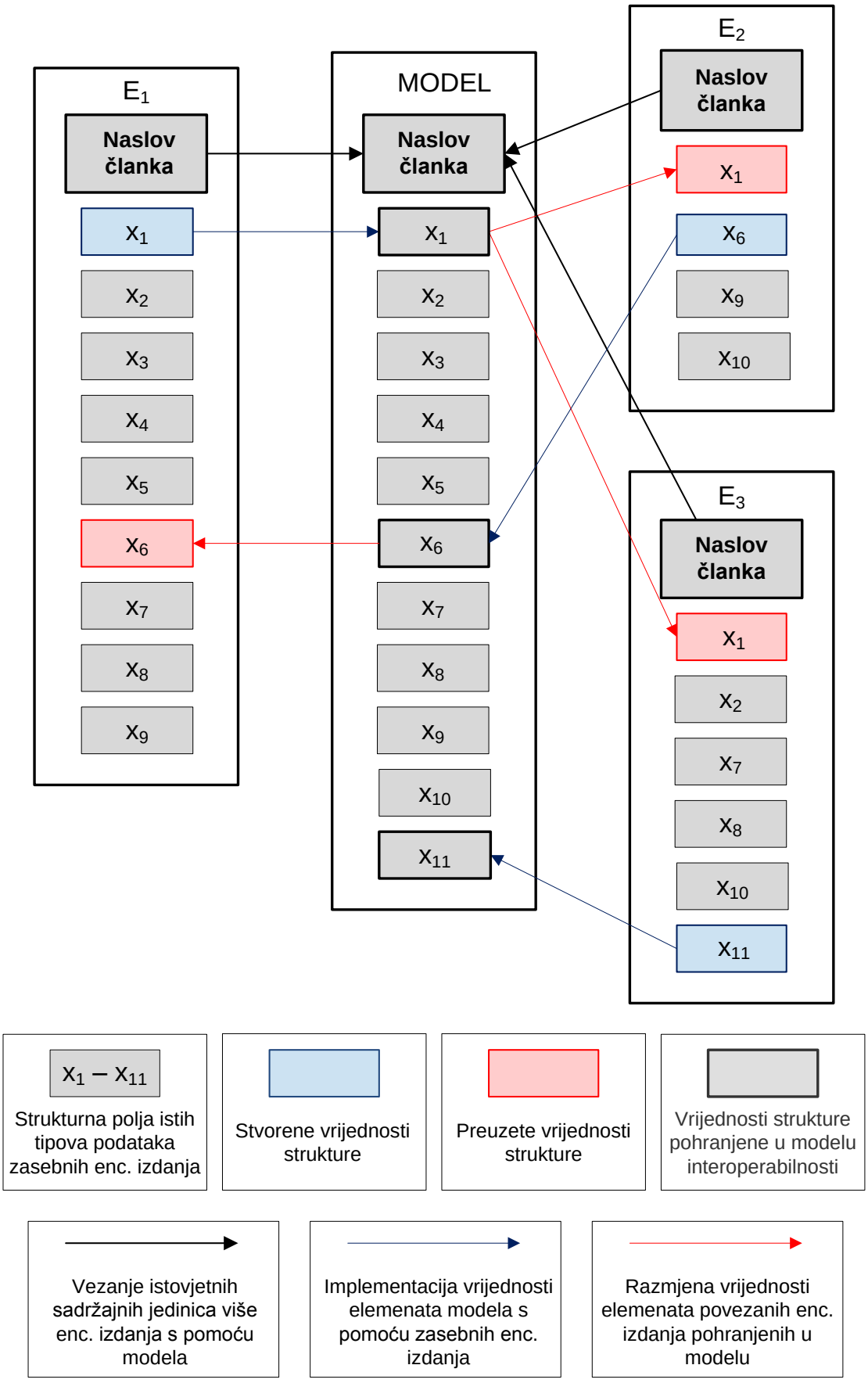
6.2.1. ZNAČAJKE I SHEMATSKI PRIKAZ MODELA ENCIKLOPEDIČKE INTEROPERABILNOSTI

Model enciklopedičke interoperabilnosti razvijen je na temelju mogućnosti infomatičkih sustava s gledišta postizanja interoperabilnoga sadržaja (poglavlje 3. 4.) te saznanja (rezultata) dobivenih analizom sadržaja mrežnih enciklopedičkih sadržaja (poglavlje 4.). Temelji se na interoperabilnosti metapodataka kao elemenata strukture mrežnih enciklopedičkih izdanja koji su implementirani prema sadržajnim potrebama svakoga zasebnog enciklopedičkoga projekta. Metapodatci su ključni u provedbi interoperabilnosti jer

predstavljaju temeljni enciklopedički sadržaj, poglavito informacije ili podatke sadržane u tekstovima članaka. Strukturiranjem enciklopedičkoga sadržaja[221], odnosno implementacijom metapodatka, od bogato informativnog enciklopedičkoga sadržaja u obliku prirodnog jezika stvaraju se enciklopedičke baze podataka. To znači da se enciklopedička interoperabilnost postiže među mrežnim enciklopedičkim izdanjima kao baza podataka, odnosno među organiziranim i računalno obradivim sustavima logički povezanih, pretraživih i međusobno ovisnih podataka (metapodataka).

Sukladno tomu, model enciklopedičke interoperabilnosti prestavlja središnju bazu podataka povezanu s istovjetnim elementima strukture (metapodatcima). Ovakav sustav omogućuje implementaciju vrijednosti elemenata strukture iz pojedinih vezanih enciklopedičkih projekata, kao i dijeljenje (razmjenu) tih istih vrijednosti među njima, ovisno o korištenim strukturnim elementima, odnosno podatcima kojima se koriste u svojoj izvedbi. Shema modela enciklopedičke interoperabilnosti prikazana je slikom 47.

[221] Prigodom razvoja modela enciklopedičke interoperabilnosti određen je pristup kroz analizu sadržaja enciklopedičkih tekstova i potvrđivanja visoke faktografske vrijednosti (informativnosti) u cilju stvaranja strukture koja će omogućiti njegovu izvedbu (poglavlje 6. 2.). Stoga se termin strukturiranje pri izradi modela enciklopedičke interoperabilnosti u ovoj disertaciji koristi u tu namjenu, iako predstavlja širi raspon djelatnosti u informacijskom smislu (poglavlje 3. 2.)



Slika 47. Shemtski prikaz i značajke modela enciklopedičke interoperabilnosti

Interoperabilnost enciklopedičkoga sadržaja koja se postiže ovim pristupom temelji se na razmjeni strukturnih elemenata vezanih uz određenu enciklopedičku sadržajnu jedinicu (članak, natuknicu). Budući da enciklopedička izdanja uglavnom obrađuju istovjetne pojmove, odnosno u procesu lematizacije dodjeljuju im iste naslove članaka, olakšan je pristup vezanja strukturnih vrijednosti istih natuknica. Shematskim prikazom prikazana su tri mrežna enciklopedička projekta (E_1–E_3) i model koji kao središnji sustav omogućava interoperabilnost sadržaja.

Kao informatički sustavi, enciklopedička izdanja i model sadrže strukturna polja (x_1–x_{11}) vrijednosti kojih omogućuju razmjenu informacija vezanih uz encikopedički sadržaj. Budući da svako enciklopedičko izdanje nastaje i djeluje zasebno, zastupljenost i tipovi strukturnih elemenata među njima mogu biti različiti. Unatoč tome, određeni tipovi podataka pojedinih kategorija enciklopedičkih članaka predstavljaju svojevrsni enciklopedički standard i za očekivati je kako će ti podatci biti uključeni u članke određene kategorije. Toj tezi u prilog idu već strukturirani članci pojedine kategorije izdanja Encyclopaedije Britannice, Wikipedije i novostrukturirane Hrvatske tehničke enciklopedije, ali i rezultati analize sadržaja provedene u okviru istraživanja vezanoga uz ovu disertaciju (poglavlje 4. 3. 1.).

Model interoperabilnosti predstavlja središnju bazu podataka koji ima mogućnost pohranjivanja svih strukturnih elemenata prisutnih u svim povezanim mrežnim enciklopedičkim projektima. Predstavljaju uniju metapodataka korištenih u povezanim projektima koji sudjeluju u pohranjivanju strukturnih vrijednosti u model, te ujedno i koriste njegov sadržaj za izgradnju vlastitoga. Shema predstavlja enciklopedička izdanja E_1 s ukupno devet elemenata strukture (x_1–x_9), E_2 s ukupno četiri strukturna elementa (x_1,x_6, x_9, x_{10}) te E_3 s ukupno šest strukturnih elemenata (x_1, x_2, x_7, x_8, x_{10}, x_{11}). Svi sadrže isti tip podatka x_1 koji se iz E_1 pohranjuje u model, budući da je to izdanje donijelo taj podatak, a koji se s pomoću modela razmjenjuje s izdanjima E_2 i E_3. Definiranje veza među strukturnim poljima koja su uvjet interoperabilnosti njihovih vrijednosti, a nalaze se u modelu te povezanim enciklopedičkim izdanjima E_1, E_2 i E_3, ostvareno je mapiranjem (Poglavlje 3. 4. 4.). U razmjeni strukturnih vrijednosti sudjeluju izdanja koja donose jednake tipove podataka, neovisno o svim povezanim projektima (x_6). Model je kao sustav proširiv i omogućuje implementaciju strukturnih elemenata nekih izdanja koje ne koriste drugi povezani projekti, kako bi se i ti podatci mogli upotrijebiti putem interoperabilnosti sadržaja u bilo kojem trenutku (x_{11}).

6.2.2. USPOSTAVA MODELA ENCIKLOPEDIČKE INTEROPERABILNOSTI U MREŽNOME PROSTORU

Uspostava modela enciklopedičke interoperabilnosti u mrežnome prostoru izvedena je na temelju znanstvenih spoznaja dobivenih iz pregleda dosadašnjeg stanja strukturiranosti i interoperabilnosti mrežnih enciklopedički sadržaja te tehničkih mogućnosti računalnih sustava uz provedenu analizu faktografske vrijednosti sadržaja postojećih mrežnih enciklopedičkih djela. Korišten je softver otvorenoga koda Airtable[222], koji omogućuje kolaborativno tablično manipuliranje sadržajem (podatcima) sa svojstvima njegove organizacije u obliku baze podataka. Tehničku izvedbu sustava omogućili su stručnjaci karlovačkoga poduzeća Infinum d. o. o. Softver pruža pohranu podataka u polja nalik tabličnim ćelijama nekih standardnih aplikacija[223], uz puno raznovrsniju pohranu tipova podataka. Standardni tipovi podataka su u obliku teksta (i cijelih članaka), privitaka datoteka, datuma i vremena, telefonskih brojeva, elektroničke pošte, URL-a, brojeva, valuta, postotaka, formula, barkodova i dr.

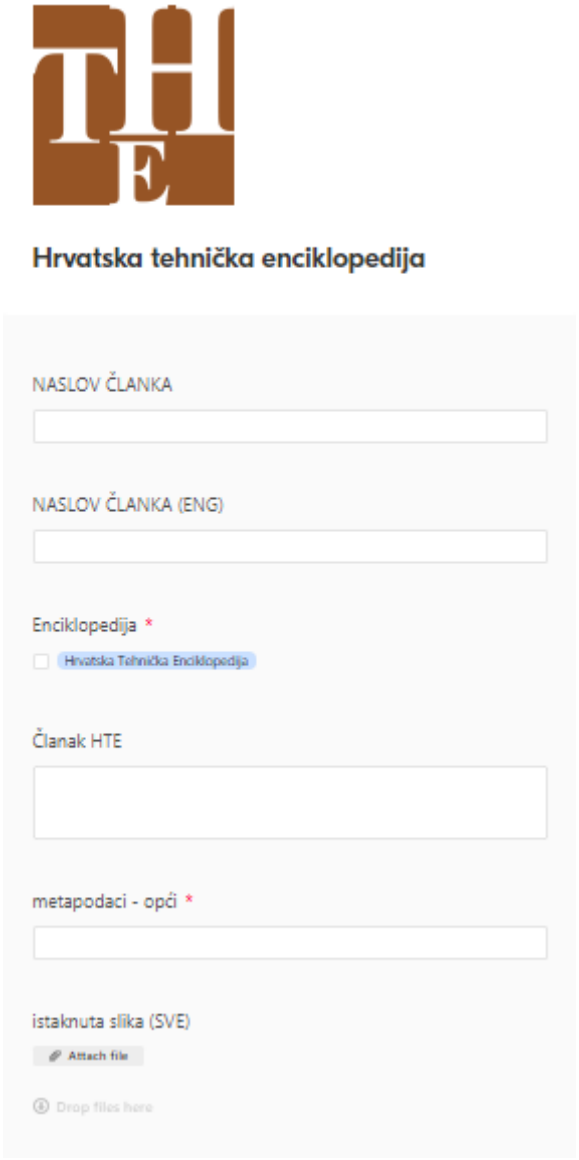
S gledišta organizacije sadržaja baze podataka ovaj sustav predstavlja relacijski model, odnosno omogućava korisnicima stvaranje baze podataka koju čine stupci (varijable) i redci (zapisi ili vrijednosti) tablica ili organiziranih skupova podataka, zatim međusobno povezivanje tablica, kolaborativni pristup u stvaranju i korištenju sadržaja, sortiranje zapisa i objavljivanje u okviru eksternih mrežnih sustava.

Kako bi se prikazala interoperabilnst enciklopedičkoga sadržaja u mrežnome prostoru, implementirani su sustavi namijenjeni unosu podataka enciklopedičkih izdanja te središnja baza podataka, u ovom kontekstu model enciklopedičke interoperabilnosti jer omogućuje pohranjivanje i razmijenu podataka svih enciklopedičkih izdanja povezanih preko njega. Stvaranje mrežnoga enciklopedičkog sadržaja omogućeno je s pomoću posebnih obrazaca za unos (prikupljanje) podataka (engl. *Airtable Form*), koji pruža organizirana polja za unos vrijednosti elemenata strukture koji se automatski organiziraju u bazi podataka (slika 48). Ovim alatom u izvedbu sustava interoperabilnosti enciklopedičkoga sadržaja uključeni su podatci

[222] https://airtable.com/ (28. 3. 2020.)

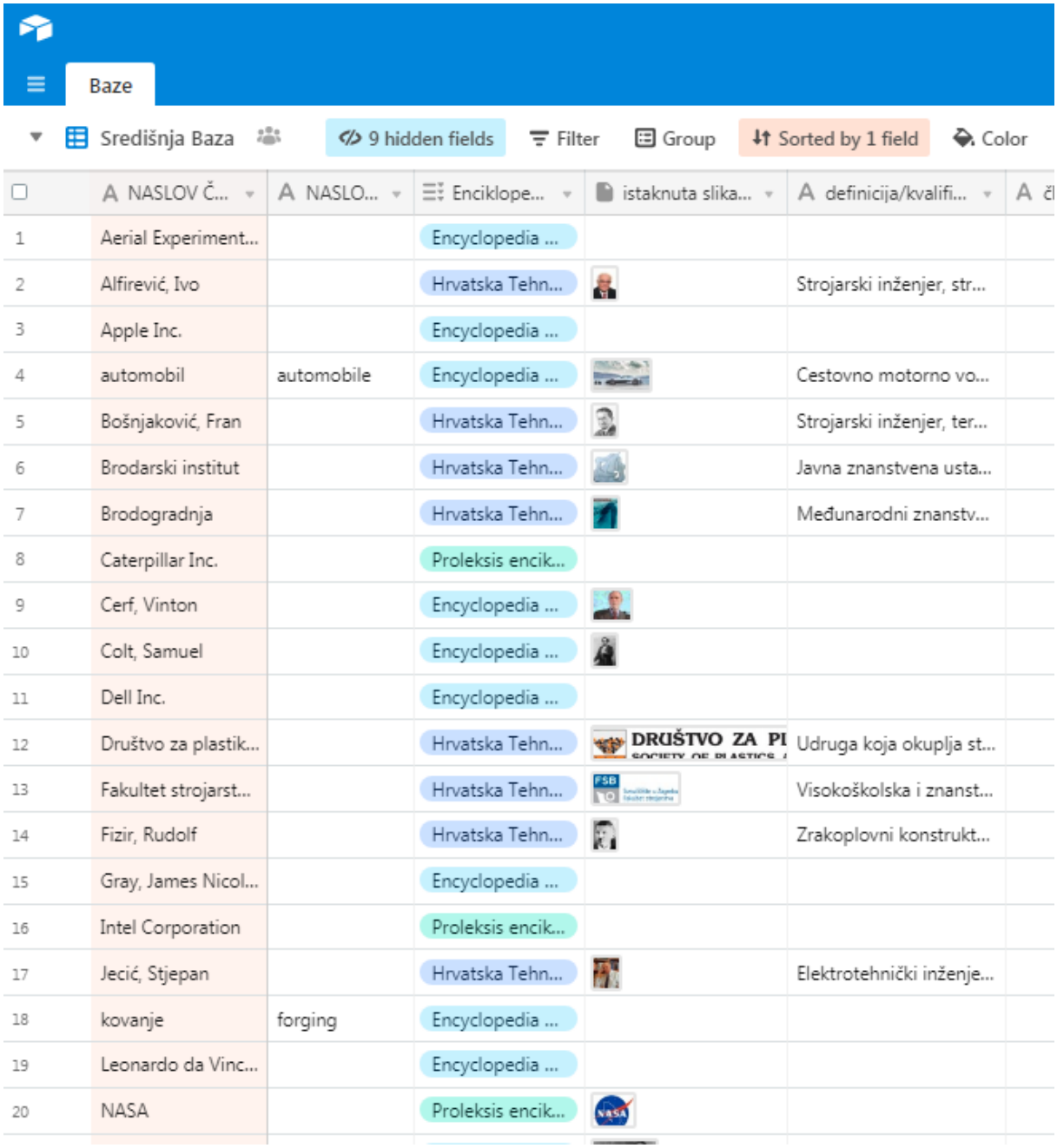
[223] Prepoznatljive pod terminom proračunske tablice (engl. *Spreadsheet*), a služe za organizaciju, analizu i pohranu podataka u tabličnom obliku. Ćelije tablica mogu sadržavati numeričke i tekstualne podatke te vrijednosti izvedene (izračunate) na temelju vrijednosti drugih ćelija. Osim izvršavanja osnovnih aritmetičkih i matematičkih funkcija, moderni sustavi pružaju funkcije financijskih te statističkih operacija i prikaza. Najpoznaiji primjer takva sustava jest široko korišteni Microsoft Excel.

dobiveni iz uzoraka Hrvatske tehničke enciklopedije, Proleksis enciklopedije te Encyclopaedije Britannice.



Slika 48. Obrazac za unos podataka (engl. *Airtable Form*) Hrvatske tehničke enciklopedije

Enciklopedička izdanja predstavljaju zasebne baze podataka međusobna interoperabilnost kojih se postiže implementacijom posebno izvedenoga modela kao središnje baze podataka ovisne o svim povezanim izdanjima. Razina strukturiranja pojedinog enciklopedičkog izdanja ovisi o zastupljenosti vrsta i tipova naziva u tekstovima članaka, koji su temeljni nositelji podataka. Manipulacija sadržajem modela provodi se s pomoću tabličnoga prikaza karakterističnoga za Airtable aplikacije (slika 49).



	NASLOV Č...	NASLO...	Enciklope...	istaknuta slika...	definicija/kvalifi...
1	Aerial Experiment...		Encyclopedia ...		
2	Alfirević, Ivo		Hrvatska Tehn...		Strojarski inženjer, str...
3	Apple Inc.		Encyclopedia ...		
4	automobil	automobile	Encyclopedia ...		Cestovno motorno vo...
5	Bošnjaković, Fran		Hrvatska Tehn...		Strojarski inženjer, ter...
6	Brodarski institut		Hrvatska Tehn...		Javna znanstvena usta...
7	Brodogradnja		Hrvatska Tehn...		Međunarodni znanstv...
8	Caterpillar Inc.		Proleksis encik...		
9	Cerf, Vinton		Encyclopedia ...		
10	Colt, Samuel		Encyclopedia ...		
11	Dell Inc.		Encyclopedia ...		
12	Društvo za plastik...		Hrvatska Tehn...	DRUŠTVO ZA PI	Udruga koja okuplja st...
13	Fakultet strojarst...		Hrvatska Tehn...		Visokoškolska i znanst...
14	Fizir, Rudolf		Hrvatska Tehn...		Zrakoplovni konstrukt...
15	Gray, James Nicol...		Encyclopedia ...		
16	Intel Corporation		Proleksis encik...		
17	Jecić, Stjepan		Hrvatska Tehn...		Elektrotehnički inženje...
18	kovanje	forging	Encyclopedia ...		
19	Leonardo da Vinc...		Encyclopedia ...		
20	NASA		Proleksis encik...		

Slika 49. Prikaz dijela sučelja modela enciklopedičke interoperabilnosti

Strukturiranje, odnosno izbor polja za unos strukturnih vrijednosti enciklopedičkih izdanja omogućen je na temelju rezultata analize sadržaja (poglavlje 4. 3. 2.) te je jednak za sva izdanja. Sukladno tomu, model pruža enciklopedičku interoperabilnost te donosi jednake strukturne vrijednosti kao i enciklopedička izdanja uključena u sustav, budući da predstavlja njihov objedinjeni sadržaj te omogućava interakciju među njihovim strukturnim poljima istoga tipa.

Takvo rješenje omogućava nadogradnju sustava (modela) strukturnim elemenima pojedinih kategorija članaka drugih područja leksikografskoga rada, osim tehničkoga. Proširivost u smislu dodavanja novih strukturnih elemenata središnjeg modela interoperabilnosti omogućuje potencijalno vezanje i drugih mrežnih projekata, neenciklopedičkih, a koji enciklopedički dio sadržaja pronalaze relevantnim. Na taj način moguća je nadgradnja enciklopedičkih sadržaja s pomoću strukture drugih izvora informacija, a koji se ne moraju obrađivati kroz enciklopedički rad.

Administratoru sustava omogućeno je povezivanje naslova istih članaka pristiglih iz više enciklopedičkih izdanja kako bi se među njima razmijenile vrijednosti strukturnh polja istoga tipa. Na taj način otklanjaju se potencijalni problemi uvjetovani homonimijom i sinonimijom, čemu su podložni naslovi enciklopedičkih članaka, ali i stvaranje jedinstvenoga enciklopedičkog sustava sa svim jedinicama sadržaja (zajednički abecedarij) i strukturnim vrijednostima.

Također, središnji model kao baza podataka pruža sve mogućnosti takva sustava, poput naprednog pretraživanja, pregledavanja i sortiranja sadržaja. Omogućene su izmjene unificianih struktunih vrijednosti svih enciklopedičkih izdanja putem modela, odnosno korekcija razlikovanja vrijednosti strukturnih polja istog tipa (ujednačavanje vrijednosti pojedinih strukturnih elemenata). U epistemološkome smislu, takvo unapređenje odnosi se na temeljne odrednice mrežnih enciklopedičkih izdanja poput točnosti i ažurnosti te ujednačavanja strukturnih vrijednosti ili podataka namijenjenih informiranju korisnika. Interoperabilnošću istog tipa podataka povezanih izdanja, sustav pruža mogućnost odabira relevantne vrijednosti koja nakon toga postaje važeća. Time dolazi do otklanjanja netočnih podataka prihvaćanjem zajedničke vrijednosti toga strukturnog elementa, a promjena vrijednosti strukturnih elemenata putem modela značit će ažuriranje podataka svih sudionika u procesu. Također, dodavanje određenih vrijednosti strukturnih elemenata značit će i implementaciju tih vrijednosti svim povezanim izdanjima, što će podići faktografsku vrijednost tih mrežnih enciklopedičkih sadržaja.

6.2.3. TESTIRANJE MODELA ENCIKLOPEDIČKE INTEROPERABILNOSTI I REZULTATI

Nakon uspostave mrežnoga modela enciklopedičke interoperabilnosti pristupilo se njegovu testiranju, odnosno eksperimentalnim pristupom ispitana je treća hipoteza o

mogućnosti postizanja interoperabilnoga mrežnog enciklopedičkoga sadržaja. Implementirani uzorak namijenjen funkcioniranju modela interoperabilnosti čine enciklopedički članci Hrvatske tehničke enciklopedije, Proleksis enciklopedije te Encyclopaedije Britannice (tablice 21, 22 i 23) nad kojim je također provedena i analiza sadržaja (poglavlja 4. 3. 1. 1., 4. 3. 1. 3. te 4. 3. 1. 6.).

Svako enciklopedičko izdanje uključeno je s ukupno 15 natuknica te je svakoj od njih pristupljeno strukturiranju sadržaja. Udio strukturiranoga sadržaja, odnosno zastupljenost vrijednosti strukturnih elemenata kao kvalitativnih opisa pojedinih naziva zastupljenih u enciklopedičkim člancima ovisio je o izvedbi pojedinih članaka, odnosno enciklopedičkih izdanja s gledišta faktografske vrijednosti. Tablicama 21, 22 i 23 prikazani su omjeri ukupno uočenih uvriježenih ili najzastupljenijih (standardnih) vrsta i tipova naziva koji se javljaju u enciklopedičkim tekstovima, zatim naziva implementiranih kao strukturne vrijednosti korištene u postizanju enciklopedičke interoperabilnosti te ukupan broj preuzetih strukturnih vrijednosti pojedinog enciklopedičkog članka zahvaljujući interoperabilnosti postignutoj među zasebnim enciklopedičkim izdanjima implementiranima kao dio modela interoperabilnosti.

Također, kodiranjem sadržaja i stvaranjem strukture korišteni su i podatci pojedinih izdanja koji se ovim istraživanjem nisu pokazali kao u potpunosti standardizirani u enciklopedici. Analizom sadržaja ta vrsta naziva okarakterizirana je pod naslovom *razno*, ali joj se mogu dodati i dijelovi sadržaja pojedinih izdanja poput izdvajanja noseće slike članka, definicije članka, donošenja podataka vezanih uz klasifikaciju članaka (organizacija znanja, npr. tezaurus ili taksonomija) i dr.

Potreba za implementacijom ovakvog sustava ili enciklopedičke paradigme najviše je uočena na slučaju Proleksis enciklopedije, koja se ovim istraživanjem pokazala faktografski smanjene vrijednosti. Od ukupno uočena 174 standardizirana naziva, implementirano je 79 strukturnih vrijednosti, a preuzeto njih čak 260 (tablica 21). To upućuje na višestruki rast strukturiranoga sadržaja uzorka Proleksis enciklopedije koji može koristiti u izgradnji svoga sustava. Kad je riječ o dodatnim tipovima podataka stavljenih na raspolaganje s pomoću modela interoperabilnosti, a koji ne spadaju u standardizirane podatke i ovim istraživanjem nisu kvantitativno razmatrani, sadržaj strukturiranih vrijednosti Proleksis enciklopedije porastao je za dodatnih 68.

Tablica 21. Kvantitativni pokazatelji omjera naziva i strukturnih vrijednosti Proleksis enciklopedije implementiranih u model enciklopedičke interoperabilnosti

Proleksis enciklopedija						
Natuknica	**Opseg (redaka)**	**Ukupno naziva u tekstu članka/strukturiranih vrijednosti/preuzetih vrijednosti strukture putem modela interoperabilnosti**				
		OS	**ORG**	**LOK**	**VO**	**Σ**
Bošnjaković, Fran	5	1/0/1	1/1/12	5/2/0	11/4/3	18/7/16
Colt, Samuel	5	1/0/2	0/0/1	5/3/0	6/4/0	12/7/3
Fizir, Rudolf	4	1/0/0	0/0/11	2/2/0	6/4/2	9/6/13
Penkala, Slavoljub Eduard	13	2/1/2	0/0/3	8/3/0	10/4/1	20/8/6
Puretić, Mario	6	2/1/–[224]	0/0/–	4/2/–	6/4/–	12/7/–
Tesla, Nikola	13	2/1/6	0/0/1	7/2/2	6/4/0	15/7/9
Tonković, Kruno	5	1/0/–	2/2/–	5/2/–	5/5/–	13/9/–
Vrančić, Faust	13	2/1/8	0/0/3	9/2/0	9/3/1	20/6/12
Caterpillar Inc.	13	1/1/–	4/3/–	2/1/–	4/1/–	11/6/–
Intel Corporation	8	0/0/–	1/1/–	2/1/–	3/1/–	6/3/–
NASA	9	0/0/–	3/1/–	7/1/–	2/1/–	12/3/–
Polaroid Corporation	4	2/2/–	1/0/–	0/0/–	1/1/–	4/3/–
automobil	28	7/7/43	0/0/84	8/–/–	7/0/0	22/7/127
strojarstvo	11	0/0/48	0/0/17	0/–/–	0/0/0	0/0/65
žičara	7	0/0/2	0/0/7	0/–/–	0/0/0	0/0/9
Σ	**144**	**22/14/107**	**12/8/139**	**64/21/2**	**76/36/7**	**174/79/260**

Broj preuzetih strukturnih vrijednosti na temelju uzorka Hrvatske tehničke enciklopedije znatno je manji u odnosu na Proleksis enciklopediju. Uzorak Hrvatske tehničke enciklopedije preuzeo je ukupno 91 strukturnu vrijednost (tablica 22) uvriježenih strukturnih elemenata, te sedam dodatnih. Međutim, velik broj izvorno strukturiranih elemenata (402) upućuje na to kako su članci Hrvatske tehničke enciklopedije faktografski visoke vrijednosti te mogu znatno unaprijediti druge enciklopedičke sadržaje putem interoperabilnoga strukturiranog sadržaja.

[224] Broj strukturiranih ili ramijenjenih vrijednosti ne prikazuju se jer navedene vrste naziva nisu kvalitativno okarakterizirane analizom sadržaja ili pojedini članak ne ostvaruje interakciju zbog odsutnosti istovjetnih sadržajnih jedinica u drugim enciklopedičkim izdanjima uključenima u model interoperabilnosti.

Tablica 22. Kvantitativni pokazatelji omjera naziva i strukturnih vrijednosti Hrvatske tehničke enciklopedije implementiranih u model enciklopedičke interoperabilnosti

Hrvatska tehnička enciklopedija						
Natuknica	**Opseg (redaka)**	**Ukupno naziva u tekstu članka/strukturiranih vrijednosti/preuzetih vrijednosti strukture putem modela interoperabilnosti**				
		OS	**ORG**	**LOK**	**VO**	**Σ**
Alfirević, Ivo	39	3/2/–	11/9/–	3/1/–	34/6/–	51/18/–
Bošnjaković, Fran	64	2/1/0	21/13/0	16/2/0	27/7/0	66/23/0
Fizir, Rudolf	92	3/0/0	21/11/0	23/2/0	44/6/0	91/19/0
Jecić, Stjepan	48	4/3/–	18/13/–	9/1/–	32/6/–	63/23/–
Penkala, Slavoljub Eduard	238	17/2/1	9/3/0	35/3/0	38/5/0	99/13/1
Ressel, Josip	65	7/2/–	3/2/–	32/3/–	20/5/–	62/12/–
Senjanović, Ivo	38	2/1/–	17/11/–	6/1/–	32/6/–	57/19/–
Vrančić, Faust	268	37/9/0	5/3/0	40/2/0	48/4/0	130/18/0
Brodarski institut	241	49/23/–	48/16/–	16/1/–	73/1/–	186/41/–
Brodogradnja	43	11/11/–	11/4/–	5/1/–	29/1/–	56/17/–
Društvo za plastiku i gumu	27	1/1/–	10/6/–	1/1/–	5/1/–	17/9/–
Fakultet strojarstva i brodogradnje	256	57/41/–	97/29/–	31/1/–	144/1/–	329/72/–
automobil	271	21/15/31	36/25/59	69/–/–	88/0/0	214/44/90
strojarstvo	687	116/48/0	161/17/0	125/–/–	244/0/0	646/65/0
žičara	149	3/2/0	19/7/0	78/–/–	42/0/0	142/9/0
Σ	**2526**	**333/156/32**	**487/169/59**	**489/19/0**	**900/49/0**	**2209/402/91**

Encyclopaedia Britannica također predstavlja bogato strukturirani enciklopedički sadržaj, međutim, zbog razlikovanja u sastavnicama uzorka s drugim enciklopedičkim izdanjima uzorka, ostvarila je umanjenu razmjenu strukturnih vrijednosti sadržaja. Preuzela je 34 strukturne vrijednosti, ali izvorno strukturirala čak 250 (tablica 23). Uspostava strukturiranoga sadržaja Encyclopaedije Britannice upućuje na to kako je vezanje strukturnih elemenata više enciklopedičkih izdanja moguće i s gledišta višejezičnosti. Univerzalna organizacija enciklopedičkoga sadržaja omogućava sustavu učinkovito vezanje istovjetnih sadržajnih jedinica, postojana je i prisutnost istih vrsta i tipova podataka (naziva), a ovakav model interoperabilnosti kao računalni sustav omogućava ujednačavanje strukturnih vrijednosti (poglavlje 3. 4. 3) te stvaranje konzistetntnog skupa podataka.

Tablica 23. Kvantitativni pokazatelji omjera naziva i strukturnih vrijednosti Encyclopaedije Britannice implementiranih u model enciklopedičke interoperabilnosti

Encyclopaedia Britannica						
Natuknica	**Opseg (redaka)**	**Ukupno naziva u tekstu članka/strukturiranih vrijednosti/preuzetih vrijednosti strukture putem modela interoperabilnosti**				
		OS	**ORG**	**LOK**	**VO**	**Σ**
Cerf, Vinton	71	26/2/–	27/8/–	7/2/–	21/3/–	81/15/–
Colt, Samuel	39	3/2/0	4/1/0	9/3/0	14/4/0	30/10/0
Gray, James Nicolas	45	9/0/–	14/12/–	8/3/–	33/5/–	64/20/–
Noyce, Robert	80	31/4/–	18/6/–	11/4/–	16/4/–	76/18/–
Stevens, John Frank	46	11/2/–	3/6/–	8/4/–	9/4/–	31/16/–
Tesla, Nikola	109	36/7/0	6/1/0	24/4/0	19/4/0	85/16/0
Torvalds, Linus	48	4/0/–	17/3/–	4/2/–	9/3/–	34/8/–
Wozniak, Steve	76	19/2/–	16/5/–	11/2/–	17/2/–	63/11/–
Aerial Experiment Association	62	24/6/–	12/1/–	5/0/–	12/2/–	53/9/–
Apple Inc.	286	30/8/–	96/6/–	4/1/–	41/1/–	171/16/–
Dell Inc.	39	1/1/–	18/3/–	4/1/–	14/2/–	37/7/–
Leonardo da Vinci Museum of Science and Technology	14	4/0/–	1/2/–	2/1/–	3/0/–	10/3/–
automobile	1676	122/39/9	143/59/25	111/–/–	212/0/0	588/100/34
forging	26	0/0/–	0/0/–	0/–/–	0/0/–	0/0/–
Wi-Fi	50	1/0/–	4/3/–	0/–/–	4/0/–	9/3/–
Σ	**2667**	**321/73/9**	**379/116/25**	**208/27/0**	**424/34/0**	**1332/252/34**

Analizom modela kao središnje baze podataka ili strukturiranoga enciklopedičkog sadržaja koja omogućava interakciju istovjetnih strukturnih elemenata među svim povezanim izdanjima utvrđena je prisutnost 35 sadržajnih jedinica s ukupno 671 strukturnom vrijednosti. To potvrđuje tezu o modelu kao bazi podataka koja čini uniju sadržaja svih povezanih enciklopedičkih izdanja. Donosi skupni aecedarij te ukupni strukturirani sadržaj stavljen na raspolaganje i budućim enciklopedičkim i potencijalno neenciklopedičkim projektma, koji mogu biti uvedeni u ovakav sustav enciklopedičke interoperabilnosti te takav tip sadržaja pronalaze relevantnim.

7. ZAKLJUČAK

7.1. EPISTEMOLOŠKE ODREDNICE ENCIKLOPEDIČKOGA SADRŽAJA – ENCIKLOPEDIČKI KONCEPT

Porastom ljudskih spoznaja javlja se stalna potreba za njihovim sređivanjem, popisivanjem, sintetiziranjem, oblikovanjem u enciklopedijski oblik. Enciklopedije postaju pogodne kao priručnici znanja o određenom području. Epistemološki postav enciklopedičkih izdanja, kao pokazatelj njihove uloge u pružanju i prenošenju znanja, svrstava enciklopedije u tercijarne izvore znanja sintetskoga tipa. Ističu se sveobuhvatnošću promatranoga područja, temeljenom na velikoj stručnosti i visokoj relevantnosti za najširi krug korisnika, objektivnošću, ograničavanjem bilo kakva utjecaja na sadržaj svojih natuknica te visokom točnošću koja ih čini veoma pouzdanim izvorom informacija. U konačnici, enciklopdička djela predstavljaju konsolidirane izvore informacija namijenjene širokoj publici. Omogućavaju razumijevanje sadržaja većini korisnika te predstavljaju premosnicu između ponekad laiku teško razumljivog stručnog ili znanstvenog sadržaja i njegova korištenja.

Digitalizacija je, osim rada leksikografa, promijenila i konceptualne značajke enciklopedičkih djela. Način na koji korisnici dolaze do novih spoznaja promijenio se, jer se enciklopedije sada kao mrežna izdanja nalaze okružene mnoštvom podataka koji kolaju internetom. Korištenjem računalnih alata omogućena je neograničenost njihova opsega, kolaborativnost s korisnicima, povećana pretraživost i dohvat informacija iz područja interesa, umreženost s drugim izvorima znanja, hipertekstualnost, ažurnost, adaptivnost i dr., čineći ih učinkovitim javnim servisom.

Enciklopedije su se promijenile te se nastavljaju mijenjati, odnosno razvijati. Tu pojavu zasigurno možemo nazvati razvojem, budući da uz kvalitete tiskanih izdanja, mrežna enciklopedička izdanja nose niz pogodnosti, čime postaju mjesta jednostavnog dohvata i razmjene pouzdanog, umreženoga sadržaja, a zbog svojih karakteristika čine epistemološki oslonac u digitalnom informiranju. Digitalizacija i globalizacija sigurno vode k umreženom, povezanom znanju, u kojem više ne postoje zapreke za brzo pronalaženje informacija. Budućnost mrežnih enciklopedija je u što učinkovitijem povezivanju vlastita sadržaja te u umrežavanju s drugim izvorima znanja. I sama odgovornost enciklopedija kao informacijskih

izvora u epistemološkom smislu postaje veća, budući da postaju dostupne svima putem uređaja povezanoga na internet.

7.2. STRUKTURIRANOST I INTEROPERABILNOST MREŽNIH ENCIKLOPEDIČKIH SADRŽAJA

Stavke enciklopedičkoga koncepta koje su bile u fokusu ovog istraživanja, strukturiranost i interoperabilnost, omogućavaju stvaranje enciklopedičkoga sadržaja razumljivoga i ljudima i računalima, sa svoga gledišta promatraju enciklopedije kao potencijalne baze podataka koje će podići funkcionalnosti, ali i omogućiti razmjenu enciklopedičkoga sadržaja s drugim informacijskim sustavima koji ga pronalaze relevantnim. To znači postavljanje enciklopedičkog sadržaja iznad granica enciklopedike.

Strukturiranost s gledišta ove disertacije odnosi se na unificirane faktografske podatke izdignute iznad tekstualnog oblika pririodnog jezika dodavanjem kontekstualnih obilježja s računarskog gledišta i pretvaranja u metapodatke namijenjene strojnoj obradi. Kao bogato strukturirana mrežna enciklopedička izdanja navedene su Wikipedija, Hrvatska tehnička enciklopedija i Encyclopaedia Britannica. Ta izdanja donose unificirane faktografske podatke organizirane u obliku infookvira, visokostrukturiranih prikaza podataka (metapodataka) u formulaciji atribut-vrijednost pogodnih za njhovu ekstrakciju kao u slučaju DBpedije, ali i općenito za organizaciju funkcionalnosti vezanih uz dohvaćanje sadržaja što je izvedeno u sklopu ovog istraživanja na Hrvatskoj tehničkoj enciklopediji.

Strukturiranjem mrežnih enciklopedičkih sadržaja po načelu tabličnih prikaza kao nadogradnje tekstualnog oblika sadržaja olakšana je normalizacija relacijskih baza podataka, odnosno tablica kao forme pohranjivanja podataka. Jasan tablični prikaz u kojem se za svaku natuknicu donose podatci (varijable) i njihove vrijednosti, tvore također retke (entiteti, natuknice) i stupce (varijable) tablica baza podataka koji tvore polja u kojima se nalaze te iste vrijednosti. Budući da se strukturiranje mrežnih stranica i pohranjivanje podataka u bazi podataka odvija po istom principu, veoma su ograničene pojave redundantnosti i međuzavisnosti podataka uz stvaranje konzistentnog modela, što tabličnu strukturu čini pogodnom za izgradnju uređene baze podataka u službi enciklopedičkoga sadržaja.

Strukturiranje mrežnih enciklopedičkih projekata na temelju unificiranih faktografskih podataka nakon provedene analize sadržaja, koji opisuju velik broj jedinica sadržaja (enciklopedičkih članaka), omogućit će daljnji razvoj mrežnih enciklopedičkih aplikacija te njihovo bolje pozicioniranje u mrežnome prostoru. S gledišta ovog istraživanja, postojanje strukture proporcionalno je implementaciji metapodataka kao predstavnika sadržaja, te je ključno omogućiti urednicima i autorima enciklopedičkih članaka postavljanje elemenata strukture koji će pospješiti samo vođenje i održavanje projekata, ali i korištenje sadržaja putem mrežnih aplikacija. Stoga je potrebno iskoristiti enciklopedička djela i njihovu organiziranost te faktografsku vrijednost u cilju stvaranja strukture kao vidljivoga zapisa tih enciklopedičkih vrijednosti.

Interoperabilnost je proporcionalna strukturiranosti, te samo bogato strukturirana mrežna enciklopedička izdanja mogu postići interoperabilnost u znatnoj mjeri u mrežnome prostoru. S gledišta interoperabilnosti kao suvremene enciklopedičke značajke, ali i svojstva informacijskih sustava općenito, potrebno je obaviti kodiranje sadržaja, odnosno implementaciju elemenata strukture ili metapodataka sukladno propisanim standardima ili shemama koje određuju sintaktička i semantička obilježja kodiranja. Primjer toga su navedeni projekti DBpedije ili Wikidate koji svoju strukturu zapisuju u RDF metapodatkovnom standardu te ostvaruju interoperabilnost u okviru semantičkoga weba.

Na primjeru DBpedije vidljivo je kako ontologijska baza podataka proizašla iz strukturiranoga enciklopedičkoga sadržaja Wikipedije može pružati snažnu potporu radu alata namijenjenih automatskom prepoznavanju nazivlja. Povezanošću s drugim skupovima podataka pokriva širok raspon pojmova koji se koriste pri označavanju naziva u tekstovima s jakom kontekstualnom pozadinom. Ta saznanja upućuju na to kako je potrebmo razmotriti enciklopedički sadržaj i drugih profesionalno izvedenih izdanja pri izgradnji korpusa namijenjenih radu alata za automatsko prepoznavanje nazivlja, a tome u prilog ide i ovo istraživanje alata, koji su na enciklopedičkim tekstovima pokazali slabije rezultate budući da na njima nisu razvijani, ali bi sveobuhvatni enciklopedički sadržaj sigurno na raspolaganje stavio velik broj pojavnica teksta iz kojih bi se proširio broj i vrste naziva kojima bi sustavi raspolagali.

S druge strane, moguće je pristupiti mapiranju strukturnih polja elemenata strukture među više sustava. Ovaj pristup gubi na važnosti povećanjem skupova (baza) podataka, ali omogućuje nadvladavanje razlika određenog broja sustava neovisno o karakteristikama zapisa

strukturnih elemenata. Moguće je i prevođenje kodiranoga sadržaja s jedne sheme na drugu. Međutim, ako se usklađuje shema s većim i raznolikijim brojem strukturnih elemenata na neku jednostavniju, moguć je gubitak sadržaja, odnosno nemogućnost opisivanja tih elemenata jednostavijim zapisom elemenata. Za povezivanje strukture određenoga broja mrežnih enciklopedija, odnosno izgradnju ograničenih sustava sa specifičnim potrebama, ovaj pristup pokazao se učinkovitim na slučaju izrade modela interoperabilnosti u sklopu ovog istraživanja.

7.3. ANALIZA SADRŽAJA MREŽNOGA ENCIKLOPEDIČKOG SADRŽAJA IZ TEHNIČKOGA PODRUČJA

Analiza sadržaja provedena je kako bi se dobio uvid u faktografsko stanje enciklopedičkoga sadržaja u cilju potencijalnog kodiranja, odnosno strukturiranja. Kako bi se upozorilo na faktografski standard enciklopedičkih djela općenito, u istraživanje je bilo uključeno ukupno osam mrežnih enciklopedičkih projekata. Kao djela hrvatske enciklopedike uključeni su projekti Hrvatske enciklopedije, Hrvatskoga biografskog leksikona, Hrvatske tehničke enciklopedije, Proleksis enciklopedije i Istarske enciklopedije kao proizvodi Leksikografskog zavoda Miroslav Krleža, te hrvatska inačica Wikipedije. Od izdanja na engleskom jeziku uključene su Encyclopaedia Britannica i engleska inačica Wikipedije, a na njemačkom jeziku izdanja Der Brockhaus multimedial 2008 premium i njemačka inačica Wikipedije. Ukupno je, dakle, bilo uključeno deset enciklopedičkih izdanja (tri inačice Wikipedije koje predstavljaju zasebna izdanja) s 455 enciklopedičkih članaka podijeljenih u tri osnovne kategorije.

Pogledom na rezultate zasebnih izdanja (poglavlje 4. 3. 1.) vidljivo je kako zastupljenosti pojedinh naziva variraju od izdanja do izdanja pod utjecajem poglavito postavki tih djela u smislu opsežnosti članaka u vezi s područjem tehnike (opća izdanja ili specijalizirana) i tipom projekta (enciklopedija ili leksikon). Međutim, znatan broj naziva prisutan je u većini članaka iste kategorije neovisno o ispitivanom izdanju, i oni predstavljaju enciklopedički faktografski standard. Tu spadaju podatci poput mjesta, datuma ili godine rođenja i smrti osoba te sjedišta ili godina osnutka organizacija kao najistaknutijih (poglavlje 4. 3. 2.).

Time je potvrđena prva hipoteza ovog istraživanja i postojanje svojevrsnog enciklopedičkoga standarda s gledišta informativnosti, odnosno sustavnog donošenja

istovjetnih vrsta i tipova podataka za određenu kategoriju enciklopedičkoga članka. Rezultati ispitivanja najbolje su prikazani tablicama 11, 12 i 13.

Takvi visokozastupljeni podatci koji opisuju velik broj jedinica sadržaja ključni su za stvaranje opsežnih organiziranih skupova podataka kao temelja bogatih baza podataka u službi enciklopedičkoga sadržaja, ali i kao strukturirani enciklopedički sadržaj koji se može koristiti u druge, neenciklopedičke svrhe. S druge strane, ako promatramo uniju elemenata strukture dobivenih analizom sadržaja, broj naziva za stvaranje strukture vezanih uz kategoriju natuknica osoba iznosi 33, organizacija 23, a općih pojmova sedam. Neki podatci karakteristični su za pojedina enciklopedička izdanja i nisu zabilježeni u drugim izdanjima ovim istraživanjem. Najbolji primjer je izdanje Hrvatskoga biografskog leksikona koji donosi podatke vremenskih obilježja vezanih uz srednješkolsko obrazovanje ili organizacija kao časopisa u kojima su kadidati ostvarivali svoj opus. U takvim slučajevima vidljiva je potreba za uspostavljanjem interoperabilnosti enciklopedičkoga sadržaja među enciklopedičkim izdanjima kako bi ti podatci postali dio svih međusobno interoperabilnih izdanja, odnosno bili stavljeni u uporabu putem zajedničkoga interoperabilnog sustava.

7.4. ISPITIVANJE ALATA ZA PREPOZNAVANJE NAZIVLJA NA ENCIKLOPEDIČKIM TEKSTOVA

Ispitivanje evaluacijskih mjera alata namijenjenih automatskom prepoznavanju nazivlja provedeno je kako bi se mogli donijeti zaključci o automatizaciji ili poluautomatizaciji prigodom prepoznavanja unificiranih faktografskih podataka u cilju njihove preinake iz teksta prirodnog jezika u strukturirane podatke.

Ipak, enciklopedički tekstovi sadržavaju jezične značajke koje otežavaju rad postojećih alata i smanjuju vrijednosti evaluacijskih mjera do te mjere da se u postojećoj izvedbi ne mogu pouzdano koristiti bez većeg nadzora pri prepoznavanju nazivlja u tekstu. Alati ili algoritmi razvijani za određenu domenu otežano mogu provoditi prepoznavanje nazivlja na tekstovima drugih domena ili žanrova. To se odnosi na razlikovanje vokabulara, rečenične sinatkse i semantike ili konteksta u kojem se nazivi pojavljuju. Upravo takav slučaj bio je prisutan prigodom ovog ispitivanja evaluacijskih mjera alata, budući da ni jedan nije u prvom redu namijenjen području enciklopedike ili enciklopedičkim tekstovima zahtjevnima po rečeničnoj strukturi, povećanom broju naziva i njihovim kontekstualnim obilježjima koje treba prepoznati.

Stoga je druga hipoteza kao dio ovog istraživanja u velikoj mjeri opovrgnuta, odnosno alati namijenjeni automatskom prepoznavanju nazivlja prigodom obrade tekstova enciklopedičkih članaka ne ostvaruju vrijednosti evaluacijskih mjera u onoj mjeri u kojoj to čine na tekstovima općega sadržaja na kojima su i razvijani. Unatoč tomu, istraživanje je dokazalo vrijednost evaluacijskih mjera korištenih modela pri ispitivanju određenih vrsta naziva (tablica 20) na zadovoljavajućoj razini, odnosno neke vrijednosti korištene za vrednovanje automatskog prepoznavanja nazivlja dobivene obradom enciklopedičkih tekstova nisu zaostajale za navedenim vrijednostima tih alata koje ostvaruju na tekstovima na kojima su razvijani. Sukladno tomu, moguća je određena primjena automatskoga prepoznavanja nazivlja u enciklopedici uz odabir pojedine kategorije enciklopedičkih članaka i prepoznavanje pojedinih tipova naziva.

Unapređenje automatizacije u procesu označavanja naziva u području enciklopedike može se postići iz dva gledišta. Jedno se odnosi na izgradnju ili razvijanje alata na enciklopedičkim tekstovima ili nadogradnju postojećih alata s enciklopedičkom domenom. Drugi smjer predstavlja prilagodbu enciklopedičkih tekstova općejezičnoj razini kojoj su postojeće aplikacije za prepoznavanje nazivlja namijenjene (poglavlje 5. 2. 3. 2.). To ne znači odstupanje od sastavnica enciklopedičkoga koncepta, već odmak od pojedinih uvriježenih enciklopedičkih formulacija i izraza, kojih promjena ne bi ugrozila enciklopedički standard u cjelini, a pospješila bi strojnu obradu enciklopedičkih tekstova kao nositelja sadržaja, odnosno potencijalne strukture. Tomu pogotovo može pridonijeti sve veća zastupljenost enciklopedičkih izdanja na novim medijima, pogovo u mrežnom okruženju, koja pruža neograničenost opsega pojedinih članaka i kontinuiranost u razvoju takvih projekata.

Treba napomenuti kako je ovo ispitivanje alata namijenjenih automatskom prepoznavanju nazivlja provedeno na enciklopedičkim tekstovima iz područja tehnike. Sukladno tomu, ispitivanje postojećih alata na enciklopedičkim tekstovima drugih područja enciklopedičkoga djelovanja moglo bi polučiti razlikovanje u konačnim rezultatima. Međutim, razmatranjem ovih rezultata, zaključak je kako bi na njihove vrijednosti evaluacijskih mjera djelovala razlika u omjerima pojedinih vrsta i tipova naziva, ali ne i karakteristike tekstova proizašle iz temeljnih enciklopedičkih postavki.

7.5. USPOSTAVA MODELA INTEROPERABILNOSTI MREŽNOGA ENCIKLOPEDIČKOG SADRŽAJA IZ TEHNIČKOGA PODRUČJA

Nakon analize sadržaja, te na njenim spoznajama, strukturiranja sadržaja Hrvatske tehničke enciklopedije, Proleksis enciklopedije i Encyclopaedije Britannice, ta struktura korištena je kao temelj rada modela koji omogućuje njenu razmjenu (interoperabilnost). Također, njegova implementacija dovodi u vezu sva istraživanja provedena u ovoj disertaciji koja se odnose na dosadašnji stupanj strukturiranosti mrežnih enciklopedičkih izdanja, karakteristike interoperabilnoga sadržaja s gledišta tehničke izvedbe, potrebe za provođenjem analize sadržaja u cilju stvaranja novog strukturiranog enciklopedičkoga sadržaja te razmatranja što veće automatizacije tog procesa uključivanjem računalnih algoritama, odnosno automatskoga prepoznavanja nazivlja radi pospješivanja učinkovitosti tih procesa.

Implementacijom i testiranjem mrežnoga modela enciklopedičke interoperabilnosti skrenula se pozornost na vezu između strukturiranosti i interoperabilnosti, odnosno postizanja interoperabilnosti mrežnih enciklopedičkih djela na temelju unaprijed strukturiranoga sadržaja. Time je dokazana treća hipoteza, odnosno mogućnost postizanja enciklopedičke interoperabilnosti kao razmjene vrijednosti elemenata strukture više enciklopedičkih izdanja u mrežnome prostoru.

Pojedina mrežna enciklopedička izdanja bogato su strukturirana te predstavljaju potencijal s gledišta interoperabilnosti. Međutim, kvantitativna i kvalitativna analiza sadržaja enciklopedičkih tekstova upućuje na dodani potencijal temeljnoga sadržaja mrežnih enciklopedičkih projekata, odnosno mogućnost dodatnoga strukturiranja i proporcionalno tomu postizanja interoperabilnosti na višoj razini. Taj proces može se pospješiti uključivanjem račnalnih algoritama ili alata namijenjenih strojnoj obradi sadržaja. Automatskim prepoznavanjem nazivlja otvaraju se mogućnosti učinkovitoga strukturiranja mrežnoga enciklopedičkoga sadržaja razvojem alata za njegovu obradu, ali i prilagodbom enciklopedičkih tekstova općejezičnoj razini za koju su se razvijale dosadašnje aplikacije.

Kao što je navedeno u poglavlju 3. 4., za postizanje interoperabilnosti ključno je zadovoljiti uvjete sintaktičke i semantičke interoperabilnosti. Jedan način je prihvaćanjem istoga standarda strukturnih elemenata, a drugi njihovim mapiranjem i nadvladavanjem

razlikovanja u razmatranim metapodatkovnim standardima. Upravo mapiranjem strukturnih elemenata postignuta je interoperabilnost sadržaja povezanih izdanja s modelom koji na sebe veže strukturne elemente. Takav pristup za ograničene sustave sa specifičnom strukturom opravdan je budući da svako enciklopedičko izdanje djeluje zasebno te bi bilo potrebno napraviti kompromise oko prihvaćanja istog standarda s obzirom na strukturne potrebe svakog projekta, odnosno količinu i kontekstualna obilježja podataka. U slučaju povezanih izdanja u ovom istraživanju, njihova struktura nije zasnovana na određenoj uvriježenoj shemi metapodataka.

Model interoperabilnosti implementiran kao središnja baza podataka omogućuje vezanje istovjetih jedinica sadržaja (članaka) više izdanja uz vezanje i omogućavanje razmjene njihove strukture, što model čini čvorištem enciklopedičkoga sadržaja, ali i opsežno enciklopedičko izdanje koje čine sva povezana izdanja. Budući da enciklopedička izdanja sadrže različite popise članaka (abecedarije), stoga i različit sadržaj, model omogućava stvaranje zajedničkoga sustava jedinica sadržaja, odnosno zaokruživanje svih sadržajnih cjelina vezanih uz enciklopedički rad. Na taj način objedinjuje enciklopedičke napore izvedene za izradu svakog enciklopedičkog izdanja, te ujedno i sam dobiva potencijal postati enciklopedički projekt, primjerice kao središnji enciklopedički portal namijenjen korisnicima. Također, ovakav sustav može pridonijeti racionalizaciji leksikografskoga rada i većoj kolaboraciji enciklopedista ili leksikografa, neovisno o projektima na kojima rade. Automatizacija procesa interakcije cjelokupnoga strukturiranog enciklopedičkog sadržaja predstavlja korak u smjeru njegove veće točnosti, pouzdanosti, ujednačenosti, ažurnosti, pretraživosti, adaptivnosti korisniku ili općenito kvalitete.

Kako je ovo istraživanje provedeno na enciklopedičkim tekstovima iz tehničkoga područja, za očekivati je kako bi analize drugih područja enciklopedičkoga djelovanja pokazale potrebu za strukturiranjem drugih strukturnih elemenata. U tom smislu, daljnjim razvojem modela moguća je implementacija enciklopedičke interoperabilnosti i na druge enciklopedičke sadržaje, odnosno on postaje proširiva platforma za pohranu i razmjenu strukturiranoga enciklopedičkoga sadržaja uopće.

Također, kao računalni sustav može obuhvatiti strukturne elemente i drugih, neenciklopedičkih, projekata koji bi bili korisni s gledišta enciklopedičke strukturiranosti i izgradnje mrežnih enciklopedičkih aplikacija. Primjeri takvih projekata prožetih

metapodatcima su npr. portali Znameniti.hr [225] koji pruža pristup digitaliziranim djelima velikana hrvatske kulture, znanosti, umjetnosti i javnoga života, te Europeana[226], platforma za slobodno pretraživanje građe digitalnih knjižnica.

Općenito, postizanjem interoperabilnosti strukturiranoga enciklopedičkoga sadržaja omogućuje se nadogradnja svakog zasebnog enciklopedičkog izdanja kao mrežne aplikacije prema potrebama, a time ujedno i pospješuju funkcionalnosti prema krajnjem korisniku.

[225] http://znameniti.hr/ (31. 5. 2020.)

[226] https://www.europeana.eu/hr (31. 5. 2020.)

8. LITERATURA

Allen, C., Jagodzinski, C.: From SEP to SEPIA: How and why Indiana University is helping the Stanford Encyclopedia of Philosophy. Against the Grain, 18(2006) 4, str. 42–43.

Allen, C., Nodelman, U., Zalta, Edward N.: The Stanford Encyclopedia of Philosophy: A Developed Dynamic Reference Work. Metaphilosophy, 33(2002) 1–2, str. 210–228.

Auer S., Bizer C., Kobilarov, G.: DBpedia: A Nucleus for a Web of Open Data. U: The Semantic Web. 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007. Busan, 2007., str. 722–735.

Baca, M.: Introduction to metadata. Los Angeles, 2008.

Baker, T., Dekkers, M., Heery, R.: What Terms Does Your Metadata Use?. Application Profiles as Machine-Understandable Narratives. U: International Conference on Dublin Core and Metadata Applications. Tokyo, 2001., str. 151–159.

Bizer, C., Heath, T., Berners-Lee, T.: Linked Data – The Story So Far. (2009) http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf (5. 5. 2017.)

Boguraev, B., Briscoe, T.: Computational Lexicography for Natural Language Processing. New York, 1989.

Bokka, K. R. et al.: Deep Learning for Natural Language Processing: Solve your natural language processing problems with smart deep neural networks. Birmingham, 2019.

Casebourne, I. et al.: Assessing the accuracy and quality of Wikipedia entries compared to popular online encyclopaedias: A comparative preliminary study across disciplines in English, Spanish and Arabic. Brighton, 2012.

Chan, L. M., Zeng, M. L.: Metadata Interoperability and Standardization – A Study of Methodology (Part I and II). D – Lib Magazine, 12(2006) 6.

Čepo, D.: Društveno-politička uvjetovanost leksikografskih članaka: studija slučaja članaka Europska unija. Studia lexicographica, 2(2008) 1(2), str. 9–24.

DBpedia. https://wiki.dbpedia.org/about (24. 9. 2019.)

Domas White, M., Marsh E.: Content Analysis: A Flexible Methodology. Library Trends, 55(2006) 1, str. 22–45.

Encyclopaedia Britannica. https://www.britannica.com/topic/Encyclopaedia-Britannica-English-language-reference-work (17. 9. 2019.)

Erxleben, F. et al.: Introducing Wikidata to the Linked Dana Web. U: Proceedings of the 13th International Semantic Web Conference - Part I. Riva del Garda, 2014., str. 50–65.

Exner, P., Nugues, P.: Entity Extraction: From Unstructured Text to DBpedia RDF Triples. U: Proceedings of the Web of Linked Entities Workshop in conjuction with the 11th International Semantic Web Conference (ISWC 2012). Boston, 2012., str. 58–69.

Fallis, D.: Toward an Epistemology of Wikipedia. Journal of the American Society for Information Science and Technology, 59(2008) 10, str. 1662–1674.

Fallis, D.: On verifying the Accuracy of Information: Philosophical Perspectives. Library Trends, 52(2004) 3, str. 463–487.

Fallis, D.: Epistemic Value Theory and Information Ethics. Minds and Machines, 14(2004) 1, str. 101–117.

Fatally flawed. Refuting the recent study on encyclopedic accuracy by the journal Nature. Encyclopaedia Britannica (2006) https://corporate.britannica.com/britannica_nature_response.pdf (17. 9. 2019.)

Featherstone, M., Venn, C.: Problematizing Global Knowledge and the New Encyclopedia Project. Theory, Culture & Society, 23(2006) 2–3, str. 1–20.

Filipić L., Jurić T., Stupar M.: Strojno prepoznavanje naziva u tekstovima pisanima hrvatskim jezikom. Studentski znanstveni rad, Rektorova nagrada, Filozofski fakultet, Sveučilište u Zagrebu, 2012.

Finkel, J, Grenager T., Manning, C.: Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. U: Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005). Stroudsburg, 2005., str. 363–370.

Francisco, J., Garcia, M.: The evolution of thesauri and the history of knowledge organization: Between the sword of mapping knowledge and the wall of keeping it simple. Brazilian Journal of Information Studies: Research Trends, 10(2016) 1., str. 1–11.

Garrido P., Tramullas, J., Martinez, F. J.: Application of Semantic Tagging to Generate Superimposed Information on a Digital Encyclopedia. U: Metadata and Semantic Research, 4th International Conference. Alcalà de Henares, 2010., str. 84–94.

Giedymin, J.: Reliability of informations. British Journal for the Philosophy of Science, 13(1963) 52, str. 287–302.

Gilchrist A., Mahon, B.: Information Arhitecture. Designing iformation environments for purpose. London, 2004.

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

Glavaš, G. et al.: CroNER: A State-of-the-Art Named Entity Recognition and Classification for Croatian. U: Proceedings of the Eighth Language Technologies Conference. Ljubljana, 2012., str. 73–78.

Hamasaki, M., Goto, M., Takeda, H.: Social Infobox: Collaborative Knowledge Construction by Social Property Tagging. U: Proceedings of the 2011 ACM Conference on Computer Supported Cooperative Work. Hangzhou, 2011., str. 641–644.

Hammer, E. M.., Zalta, E. N.: A Solution to the Problem of Updating Encyclopedias. Computers and the Humanities, 31(1997) 1, str. 47–60.

Haslhofer, B., Klas, W.: A Survey of Techniques for Achieving Metadata Interoperability. ACM Computing Surveys (CSUR), 42(2010) 2.

Haynes, D.: Metadata for information management and retrieval. London, 2004.

Heery, R., Patel, M.: Application profiles: Mixing and Matching Metadata Schemas. Ariadne - Web magazine for information professionals, (2000) 25.

Hjørland, B.: The foundation of the concept of relevance. Journal of the American Society for Information Science and Technology, 61(2010) 2, str. 217–237.

Ide, N., Pustejovsky, J.: What Does Interoperability Mean, Anyway? Toward an Operational Definition of Interoperability for Language Technology. U: Proceedings of the Second International Conference on Global Interoperability for Language Resources. Hong Kong, 2010.

Jain, A., Kulkarni, G., Shah, V.: Natural Language Processing. International Journal of Computer Sciences and Engineering. 6(2018) 1, str. 161−167.

Jandrić, P., Boras, D.: Kritičko e-obrazovanje. Zagreb, 2012.

Jecić, Z.: Enciklopedički koncept u mrežnom okruženju. Studia lexicographica, 7(2013) 2(13), str. 99−115.

Jecić, Z: Od Tehničke enciklopedije do Hrvatske tehničke enciklopedije. Studia lexicographica, 12(2018) 23, str. 53–79.

Jecić, Z., Boras, D., Domijan, D.: Prilog definiranju pojma virtualna enciklopedija. Studia lexicographica, 2(2008) 1(2), str. 115−126.

Jurafsky D., Martin, J. H.: Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Second Edition. New Jersey, 2009.

Kapustić, S.: Metodika organizacijskog projektiranja. Samobor, 1984.

Kane, G. C., Ransbotham, S.: Collaborative development in Wikipedia. (2012) https://arxiv.org/pdf/1204.3352.pdf (21. 3. 2017.)

Katz, W. A.: Introduction to Reference Work, Volume I. Basic Information Sources. New York, 1978.

Katz, W. A.: Introduction to Reference Work, Volume II. Reference Services and Reference Processes. New York, London, 1978.

Khurana et al.: Natural Language Processing: State of The Art, Current Trend and Challenges (2006) https://arxiv.org/abs/1708.05148v1 (29. 5. 2020.)

Klobučar Srbić, I., Smolčić, I., Tolj, J.: Bibliografija Tehničke enciklopedije (1963-97). Studia lexicographica, 12(2018) 23, str. 199–257.

Kolbitsch, J., Safran, C., Maurer, H.: Dynamic Adaptation of Content and Structure in Electronic Encyclopaedias. Journal of Digital Information, 8(2007) 3.

Kräenbring J. et al.: Accuracy and Completeness of Drug Information in Wikipedia: A Comparison with Standard Textbooks of Pharmacology. PLoS ONE, 9(2014) 9.

Krippendorff, K.: Content analysis: An introduction to its methodology. Thousands Oaks, 2004.

Kubelka, O., Šoštarić, P.: Wikipedija nasuprot Hrvatskoj enciklopediji, kvalitativan odnos slobodno i tradicionalno uređenoga enciklopedijskoga sadržaja a hrvatskome jeziku. Studia lexicographica, 5(2001) 2(9), str. 119–134.

Kuhn, T. S.: Struktura znanstvenih revolucija. Zagreb, 2003.

Laffetry, J., McCallun, A., Pereira F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Dana. U: Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001). San Francisco, 2001., str. 282–289.

Lange, D., Böhm, C., Naumann, F.: Extracting structured information from Wikipedia articles to populate infoboxes. U: Proceedings of the 19th ACM international conference on Information and knowledge management. New York, 2010., str. 1661–1664.

Lehmann, J., Isele, R., Jakob, M.: DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. Semantic Web, 6(2015) 2, str. 167–195.

Li, J., Gao, J., Dong, J.: A metadata registry for metadata interoperability. Data science Journal, 6(2007), str. 379–384.

Ljubešić, N. et al.: Combining Available Datasets for Building Named Entity Recognition Models of Croatian and Slovene. Slovenščina 2.0, 2(2013) 1, str. 35–57.

Magnus, P. D.: Epistemology and the Wikipedia. North American Computing and Philosophy Conference. New York, 2006.

Manning, C. D. et al.: The Stanford CoreNLP Natural Language Processing Toolkit. U: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Stroudsburg, 2014., str. 55–60.

Matešić, M., Vučković K., Dovedan, Z.: Social Software: Teaching Tool or Not?. U: INFuture2009: Digital Resources and Knowledge Sharing. Zagreb, 2009., str. 433−422.

Mendes, P. N. et al.: DBpedia Spotlight: Shedding Light on the Web of Documents. U: Proceedings of the 7th International Conference on Semantic Systems (I-Semantics 2011). Graz, 2011., str. 1–8.

Mika, P. et al.: Learning to tag and tagging to learn: A case study on Wikipedia. IEEE Intelligent Systems, 23(2008) 5, str. 26–34.

Morsey, M. et al.: DBpedia and the live extraction of structured data from Wikipedia. Program: electronic library and information systems, 46(2012) 2, str. 157–181.

Nadeau, D., Turney, P. D., Matwin, S.: Unsupervised Named-Entity Recognition: Generating Gazetteers and Resolving Ambiguity. U: Proceedings of the 19th international conference on Advances in Atificial Inteligence: Canadian Society of Computational Studies of Inteligence. Quebec City, 2006., str. 266–277.

Nodelman, U. et al.: Stanford Encyclopedia of Philosophy: A Dynamic Reference Work. U: Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries. New York, 2003., str. 380.

Ouksel, A. M., Sheth, A.: Semantic Interoperability in Global Information Systems: A Brief Introduction to the Research Area and the Special Section, 28(1999) 1, str. 5–12.

Pasternack, J., Roth, D.: The Wikipedia Corpus. (2008) http://cogcomp.cs.illinois.edu/papers/PasternackRo08.pdf (7. 3. 2017.)

Pattuelli, C., Rubinow, S.: The knowledge organization of DBpedia: a case study. Journal of Documentation, 69(2013) 6, str. 762–772.

Perry, J., Zalta, E. N.: Why Philosophy Needs a ‘Dynamic’ Encyclopedia. (1997) https://plato.stanford.edu/pubs/why.html (13. 4. 2017.)

Piscopo, A. et al.: What do Wikidata and Wikipedia Have in Common? An Analysis of their Use of External References. U: Proceedings of the 13th International Symposium on Open Collaboration. Galway, 2017.

Qin, J., Prado, J. C.: The semantic and syntactic model of metadata. Alfabetização Digital e Accesso ao Conhecimento, 4(2006), str. 143–156.

Rahm, E., Bernstein, P. A.: A survey of approaches to automatic schema matching. The VLDB Journal, 10(2001) 4, str. 334–350.

Rector, L.: Comparison of Wikipedia and Other Encyclopaedias for Accuracy, Breadth, and Depth in Historical Articles. Reference Services Review, 36(2008) 1, str. 7–22.

Riley, J.: Understanding metadata: What is metadata, and what is it for?. Baltimore, 2017.

Reiss, J., Sprenger, J.: Scientific Objectivity. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/sum2016/entries/scientific-objectivity/ (17. 9. 2019.)

Ruth, W., Shortreed, J., Landoni, M.: Browsing and Searching E-encyclopaedias. U: Recherche d'Information Assistee par Ordinateur. Avignon, 2004., str. 763−772.

Saracevic, T.: The Notion of Relevance in Information Science. Everybody knows what relevance is. But what is it really?. Morgan & Claypool, 2017.

Saracevic, T.: Relevantnost i kako se istraživala. Vjesnik bibliotekara Hrvatske, 50(2007) 1/2, str. 1–26.

Smolčić, I., Jermen, N., Jecić, Z.: Analiza enciklopedičke relevantnosti na primjeru hrvatske tehničke baštine. Studia lexicographica, 9(2015) 1(16), str. 133–160.

Smolčić, I., Tolj, J., Jecić, Z.: Epistemological Value of Contemporary Encyclopedic Projects. U: INFuture2017: Integrating ICT in Society. Zagreb, 2017., str. 141−149.

Starčević Stančić, I., Kraus, C.: Hrvatska enciklopedija – od tiskanoga do mrežnoga izdanja. Studia lexicographica, 8(2014) 1(14), str. 99–116.

Sultana, A. et al.: Infobox suggestion for Wikipedia entities. U: Proceedings of the 21st ACM international conference on Information and knowledge management. Maui, 2012., str. 2307–2310.

Sutton, C., McCallum, A: An Introduction to Conditional Random Fields. Foundations and Trend sin Machine Learning, 4(2012) 4, str. 267–373.

Swanson, D. R.: Undiscovered public knowledge. Library Quarterly, 56(1986) 2, str. 103–118.

Swanson, D. R., Smalheiser, N. R.: Undiscovered public knowledge: a ten year update. U: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. Portland, 1996., str. 295–298.

Špiranec, S., Babić, A., Lešković, A.: New Access Structures to Scientific Information: The Case of Science 2.0. U: INFuture2009: Digital Resources and Knowledge Sharing. Zagreb, 2009., str. 451−460.

Tananbaum, G.: I Hear the Train A Comin'. Against the Grain, 18(2006) 1, str. 84–85.

Thagard, P.: Internet epistemology: Contributions of new information technologies to scientific research. (1997) http://cogsci.uwaterloo.ca/Articles/Pages/Epistemology.html (24. 4. 2017.)

Tomović, S.: Ideološke implikacije u enciklopedistici. U: Radovi Leksikografskoga zavoda „Miroslav Krleža“, sv. 1. Zagreb, 1991., str. 45–51.

Torres, D. et al.: Improving Wikipedia with DBpedia. U: Proceedings of the 21st international conference companion on World Wide Web. Lyon, 2012., str. 1107−1112.

Tuđman, M.: Obavijest i znanje. Zavod za informacijske studije, Zagreb, 1990.

Tuđman, M.: Prikazalište znanja. Hrvatska sveučilišna naklada, 2003.

Tuđman, M., Boras, D., Dovedan, Z.: Uvod u informacijsku znanost. Školska knjiga, Zagreb, 1993.

Voss, J.: Measuring Wikipedia. U: Proceedings of the 10th International Conference on Scientometrics and Informetrics. Stockholm, 2005., str. 221–231.

Voss, J.: Tagging, Folksonomy & Co - Renaissance of Manual Indexing?. U: 10th international Symposium for Information Science. Cologne, 2007.

Voss, J.: Collaborative thesaurus tagging the Wikipedia way, 2006. https://arxiv.org/abs/cs.IR/0604036 (26. 7. 2018.)

Vujić, A.: Utemeljenje enciklopedijske leksikografije kao informacijske znanosti. U: Radovi zavoda za informacijske studije, knj. 1. Zagreb, 1990., str. 141−146.

Vujić, A.: Acta Lexicographica. Prema znanstvenom utemeljenju enciklopedijske leksikografije. Zagreb, 2015.

Zalta, E. N.: The Stanford Encyclopedia of Philosophy: A University/Library Partnership in Support of Scholarly Communications and Open Access. College & Research Libraries News, 67(2006) 8, str. 502–504.

Zalta, E. N.: Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/prc-oct99.html (10. 3. 2017.)

9. BIBLIOGRAFSKI ZAPISI ENCIKLOPEDIČKIH DJELA UKLJUČENIH U ISTRAŽIVANJE

Hrvatska tehnička enciklopedija. Leksikografski zavod Miroslav Krleža, Zagreb, 2018–, 1 sv.
Glavni urednik: Zdenko Jecić
Mrežna inačica: https://tehnika.lzmk.hr/

Hrvatska enciklopedija. Leksikografski zavod Miroslav Krleža, Zagreb, 1999–2009, 11 sv.
Glavni urednici: Dalibor Brozović, August Kovačec, Slaven Ravlić
Mrežna inačica: http://www.enciklopedija.hr/

Opća i nacionalna enciklopedija u 20 knjiga. PRO LEKSIS d. o. o., Večernji list d. d., Zagreb, 2005–2007, 21 knj.
Glavni urednik: Antun Vujić
Mrežna inačica: Proleksis enciklopedija; https://proleksis.lzmk.hr/

Istarska enciklopedija. Leksikografski zavod Miroslav Krleža, Zagreb, 2005.
Glavni urednik: Miroslav Bertoša, Robert Matijašić
Mrežna inačica: http://istra.lzmk.hr/

Hrvatski biografski leksikon. Leksikografski zavod Miroslav Krleža, Zagreb, 1983–, 8. sv.
Glavni urednici: Nikica Kolumbić, Aleksandar Stipčević, Trpimir Macan, Nikša Lučić
Mrežna inačica: http://hbl.lzmk.hr/

Encyclopaedia Britannica. Encyclopaedia Britannica Inc., Chicago, 1768–2010, 15. izdanje, 32 sv.
Glavni urednici: Philip W. Goetz, Robert McHenry, Dale Hoiberg
Mrežna inačica: https://www.britannica.com/

Brockhaus Enzyklopädie. F. A. Brockhaus AG, Leipzig, 1796–2006, 21. izdanje, 30 sv.
Glavna urednica: Annette Zwahr
DVD izdanje: Der Brockhaus multimedial 2008 premium

Hrvatska Wikipedija: https://hr.wikipedia.org/wiki/Glavna_stranica
Engleska Wikipedija: https://en.wikipedia.org/wiki/Main_Page
Njemačka Wikipedija: https://de.wikipedia.org/wiki/Wikipedia:Hauptseite

10. PRILOZI

Prilog 1. Rezultati kvantitativne analize sadržaja uzorka Hrvatske tehničke enciklopedije

Hrvatska tehnička enciklopedija							
Kategorije natuknica	Natuknica	Opseg (redaka)	Broj naziva				
			OS	ORG	LOK	VO	Σ
osobe	Alfirević, Ivo	39	3	11	3	34	51
	Andrijanić, Branko	14	1	9	5	12	27
	Bačić, Nikola	15	1	5	6	6	18
	Belamarić, Igor	31	2	12	3	17	34
	Bošnjaković, Fran	64	2	21	16	27	66
	Deželić, Đuro Stjepan	60	9	6	10	29	54
	Fizir, Rudolf	92	3	21	23	44	91
	Gabelich, Gary	32	1	3	5	17	26
	Jecić, Stjepan	48	4	18	9	32	63
	Miler, Josip	21	2	10	9	13	34
	Penkala, Slavoljub Eduard	238	17	9	35	38	99
	Prpić-Oršić, Jasna	21	3	7	2	21	33
	Ressel, Josip	65	7	3	32	20	62
	Rožanić, Igor	26	3	14	9	23	49
	Senjanović, Ivo	38	2	17	6	32	57
	Vrančić, Faust	268	37	5	40	48	130
	Σ	1072	97	171	213	413	894
organizacije	Brodarski institut	241	49	48	16	73	186
	Brodogradnja	43	11	11	5	29	56
	Croatia Airlines d. d.	69	0	24	24	23	71
	Dok-Ing d. o. o.	67	1	14	29	8	52
	Društvo za plastiku i gumu	27	1	10	1	5	17
	Fakultet strojarstva i brodogradnje	256	57	97	31	144	329
	Hrvatska komora inženjera tehnologije prometa i transporta	25	1	1	1	2	5
	Hrvatski automodelarski savez	11	1	6	5	3	15
	Hrvatski strojarski i brodograđevni inženjerski savez	27	1	11	1	11	24
	Hrvatsko vojno učilište Dr. Franjo Tuđman	152	7	81	21	34	143
	LTH metalni lijev d. o. o.	18	0	9	8	7	24

	Ljevarstvo	17	9	11	0	34	54
	Memorijalni centar Domovinskog rata Vukovar	20	1	6	1	2	10
	Metalac, Čakovec	18	0	8	1	6	15
	Polimeri	33	9	2	2	31	44
	Pomorski i povijesni muzej Hrvatskoga primorja Rijeka	86	6	27	9	28	70
	Prvomajska	261	14	71	106	71	262
	Ribarska zbirka Komiža	24	3	2	12	5	22
	Strojarstvo	38	6	9	2	32	49
	Tehnički fakultet u Rijeci	201	36	54	6	85	181
	Tehnički muzej Nikola Tesla	297	29	35	21	90	175
	Torpedo	147	7	20	17	38	82
	Tvornica motora Zagreb	53	1	7	5	9	22
	Uljanik d. d.	228	0	42	9	43	94
	Vatrogasni vjesnik	32	13	9	3	42	67
	Σ	2391	263	615	336	855	2069
opći pojmovi	aditivna proizvodnja	43	3	9	5	5	22
	automobil	271	21	36	69	88	214
	galvanotehnika	42	1	1	3	6	11
	parobrod	323	23	53	192	102	370
	pučinsko inženjerstvo	174	4	29	28	30	91
	raketno oružje	152	2	38	22	35	97
	strojarstvo	687	116	161	125	244	646
	trabakul	41	0	0	2	3	5
	žičara	149	3	19	78	42	142
	Σ	1882	173	346	524	555	1598
Σ		5345	533	1132	1073	1823	4561

Prilog 2. Rezultati kvantitativne analize sadržaja uzorka Hrvatske enciklopedije

Hrvatska enciklopedija							
Kategorije natuknica	Natuknica	Opseg (redaka)	Broj naziva				
			OS	ORG	LOK	VO	Σ
osobe	Bošnjaković, Fran	27	3	16	11	26	56
	Brnić, Josip	9	1	6	4	9	20
	Bukl, Stjepan	10	2	0	7	6	15
	Dolenc, Anton	21	1	6	8	15	30
	Fizir, Rudolf	14	1	4	6	12	23
	Hraste, Marin	11	1	3	2	9	15
	Ožegović, Franjo	19	1	10	10	11	32
	Penkala, Slavoljub Eduard	51	10	2	13	13	38
	Perić, Nedjeljko	11	1	6	4	16	27
	Prikril, Boris	21	1	6	6	25	38
	Tkalić, Mladen	13	1	5	4	12	22
	Vrančić, Faust	146	22	5	33	38	98
	Zdenković, Rudolf	15	1	9	9	18	37
	Zglav, Milan	13	1	6	7	13	27
	Žerdik, Mladen	15	1	8	3	12	24
	Žugaj, Mladen	16	1	10	6	13	30
	Σ	412	49	102	133	248	532
organizacije	Adris grupa d.d.	23	0	18	6	11	35
	Atlantska plovidba d.d.	8	0	2	1	3	6
	Brodarski institut	18	0	6	2	5	13
	Brodogradnja	10	0	4	2	5	11
	Čovjek i svemir	9	0	2	0	4	6
	Đuro Đaković Holding d.d.	13	0	4	1	5	10
	Građevinar	11	0	7	0	3	10
	Hrvatska zajednica tehničke kulture	23	0	9	1	5	15
	Hrvatske željeznice d.o.o.	10	0	5	1	4	10
	Hrvatsko metalurško društvo	8	0	5	0	5	10
	INA - Industrija nafte d.d.	39	0	28	7	13	48
	Industrogradnja d.d.	8	0	1	1	2	4
	Institut građevinarstva Hrvatske	15	0	9	8	7	24
	Jadranska straža	20	1	3	5	8	17
	Jugoturbina	8	0	6	1	10	17

	Luka Rijeka d.d.	6	0	1	0	4	5
	Metallgesellschaft AG	11	1	2	7	5	15
	Nothrop Grumman Corporation	13	1	2	1	5	9
	Pomorski zbornik	14	3	3	12	6	24
	Rimac Automobili	7	1	1	2	1	5
	Strojarstvo	11	0	4	0	2	6
	Tehnički muzej Nikola Tesla	34	5	18	4	15	42
	TLM d.d.	9	0	8	1	4	13
	TŽV Gredelj d.o.o.	10	0	5	1	3	9
	Uljanik brodogradilište d.d.	10	0	2	2	5	9
	Σ	348	12	155	66	140	373
opći pojmovi	aditivna proizvodnja	112	1	4	5	5	15
	automobil	434	20	17	24	39	100
	cestovni promet	94	0	0	13	14	27
	površinski kop	29	0	0	0	0	0
	prehrambena tehnologija	110	40	15	21	26	102
	računalo	100	16	0	2	21	39
	valjanje	74	0	9	7	13	29
	zrakoplovstvo	306	67	16	59	78	220
	žičara	118	1	1	32	16	50
	Σ	1377	145	62	163	212	582
Σ		2137	206	319	362	600	1487

Prilog 3. Rezultati kvantitativne analize sadržaja uzorka Proleksis enciklopedije

Proleksis enciklopedija							
Kategorije natuknica	Natuknica	Opseg (redaka)	Broj naziva				
			OS	ORG	LOK	VO	Σ
osobe	Bauersfeld, Walther	4	1	1	4	4	10
	Bazjanac, Davorin	10	1	1	5	7	14
	Bego, Vojislav	4	1	3	3	5	12
	Bošnjaković, Fran	5	1	1	5	11	18
	Colt, Samuel	5	1	0	5	6	12
	Ericsson, John	6	1	0	5	4	10
	Fizir, Rudolf	4	1	0	2	6	9
	Hanaman, Franjo	8	2	4	7	9	22
	Linde, Carl von	6	1	0	2	8	11
	Lupis, Ivan Blaž	9	2	1	5	6	14
	Penkala, Slavoljub Eduard	13	2	0	8	10	20
	Puh, Janez	8	1	1	4	7	13
	Puretić, Mario	6	2	0	4	6	12
	Stephenson, George	13	5	0	7	9	21
	Tainter, Charles Summer	9	3	0	3	4	10
	Telford, Thomas	5	1	0	8	4	13
	Tesla, Nikola	13	2	0	7	6	15
	Tonković, Kruno	5	1	2	5	5	13
	Vrančić, Faust	13	2	0	9	9	20
	Σ	146	31	14	98	126	269
organizacije	ADAC	4	0	2	1	1	4
	Battelle Memorial Institute	5	1	1	3	1	6
	Caterpillar Inc.	13	1	4	2	4	11
	Citroen SA	15	3	6	1	10	20
	Daum	3	1	1	1	1	4
	Empire State Building	3	1	1	1	2	5
	EURECA	6	0	2	3	1	6
	Intel Corporation	8	0	1	2	3	6
	NASA	9	0	3	7	2	12
	Polaroid Corporation	4	2	1	0	1	4
	Sears Tower	6	0	2	3	3	8
	Σ	76	9	24	24	29	86

opći pojmovi	automobil	28	7	0	8	7	22
	cement	15	0	0	3	1	4
	cestovni promet	22	2	0	13	9	24
	čišćenje	12	0	0	0	0	0
	elektrostrojarstvo	2	0	0	0	0	0
	lokomotiva	12	2	0	0	2	4
	Ottov motor	7	1	0	0	2	3
	PASCAL	6	2	0	1	1	4
	periskop	7	1	0	0	1	2
	reverberacija	7	0	0	0	0	0
	salonit	7	0	0	1	0	1
	serklaža	4	0	0	0	0	0
	stroj	8	0	0	0	0	0
	strojarstvo	11	0	0	0	0	0
	teksilni strojevi	8	0	0	0	0	0
	tinjalica	6	0	0	0	0	0
	trabakul	5	0	0	0	0	0
	turbokompresor	9	0	0	0	0	0
	zrakoplov	8	0	0	0	0	0
	žičara	7	0	0	0	0	0
	Σ	191	15	0	26	23	64
Σ		413	55	38	148	178	419

Prilog 4. Rezultati kvantitativne analize sadržaja uzorka Istarske enciklopedije

Istarska enciklopedija							
Kategorije natuknica	Natuknica	Opseg (redaka)	Broj naziva				
			OS	ORG	LOK	VO	Σ
osobe	Armanda, Adam	16	2	6	10	12	30
	Fakin, Milan	14	1	4	11	8	24
	Geršić, Ivo	14	1	6	11	18	36
	Heininger, Ottomaro	7	2	1	5	5	13
	Lugnani, Giuseppe de	24	1	10	10	17	38
	Mazzoni, Angiolo	14	3	2	17	13	35
	Popper, Siegfried	11	1	2	7	8	18
	Stipetić, Đuro	14	3	6	10	11	30
	Stipetić, Gjuro	15	1	7	8	12	28
	Tommasi, Natale	24	2	5	17	15	39
	Turina, Alberto	13	1	2	10	15	28
	Σ	166	18	51	116	134	319
organizacije	Brionka d. d. Pula	16	1	2	8	4	15
	Društvo južnih željeznica	22	1	9	18	9	37
	Fakulteta za pomorstvo in promet (FPP)	21	0	14	10	15	39
	Gospodarstvo Istre	21	3	5	8	3	19
	Hidrografski zavod u Puli	35	1	13	9	30	53
	Istarska tvornica vapna d. o. o. (ITV)	18	2	5	8	5	20
	Istarski ugljenokopi Raša	64	3	7	31	25	66
	Narodni muzej Labin	46	5	8	5	16	34
	Pomorski arsenal u Puli	101	6	10	27	89	132
	Studijski centar za arhitekturu Arhitektonskog fakulteta Sveučilišta u Zagrebu	8	1	2	1	3	7
	Tvornica duhana Rovinj d. o. o.	40	1	9	4	13	27
	Uljanik Brodogradilište d. d.	101	4	31	8	28	71
	Uljanik Strojogradnja d. d.	56	0	14	9	10	33
	Visoka tehnička škola u Puli – Politehnički studiji	15	1	3	1	3	8
	Zavičajni muzej Buzet	21	0	1	3	6	10
	Σ	585	29	133	150	259	571

opći pojmovi	automobil	42	0	6	28	18	52
	građevinske tehnike	55	0	0	3	6	9
	industrija	68	0	58	74	11	143
	most	79	0	0	53	26	79
	tuneli, prometni	45	0	1	30	19	50
	urbanizam	92	16	3	57	27	103
	utvrde ili fortifikacije	209	2	0	72	84	158
	vodogradnja	138	17	6	80	55	158
	zavarivanje u brodogradnji	38	0	5	0	1	6
	Σ	766	35	79	397	247	758
Σ		1517	82	263	663	640	1648

Prilog 5. Rezultati kvantitativne analize sadržaja uzorka Hrvatskoga biografskog leksikona

Hrvatski biografski leksikon							
Kategorije natuknica	Natuknica	Opseg (redaka)	Broj naziva				
			OS	ORG	LOK	VO	Σ
osobe	Amramović, Vladimir	40	3	18	13	25	59
	Bakrač, Boris	45	2	42	8	39	91
	Bošnjaković, Fran	45	7	7	18	32	64
	Brlek, Veljko	27	2	12	6	24	44
	Celmić, Ivan	44	3	17	17	43	80
	Cvjetković, Anton	39	2	9	15	12	38
	Čatić, Igor	35	5	21	12	64	102
	Đurađević, Aleksandar	39	1	15	18	29	63
	Fizir, Rudolf	25	1	6	11	21	39
	Franković, Ante	41	1	16	26	37	80
	Franotović, Damir	12	1	5	12	18	36
	Grubišić, Ante	22	4	2	5	8	19
	Hlavka, Jakob	24	3	7	13	10	33
	Ibrahimpašić, Ifet	24	1	9	6	22	38
	Ivan Krstitelj Rabljanin	82	29	6	29	32	96
	Jecić, Stjepan	44	3	23	19	70	115
	Katavić, Ivan	27	2	23	15	45	85
	Kolin, Ivan	45	6	25	21	49	101
	Tesla, Nikola	505	109	80	177	371	737
	Vrančić, Faust	273	94	17	60	78	249
Σ		1438	279	360	501	1029	2169

Prilog 6. Rezultati kvantitativne analize sadržaja uzorka Encyclopaedije Britannice

Encyclopaedia Britannica							
Kategorije natuknica	Natuknica	Opseg (redaka)	Broj naziva				
			OS	ORG	LOK	VO	Σ
osobe	Cerf, Vinton	71	26	27	7	21	81
	Colt, Samuel	39	3	4	9	14	30
	Eads, James B.	111	22	2	35	10	69
	Ford, Henry	305	84	31	38	47	200
	Goddard, Robert	103	22	7	15	30	74
	Gray, James Nicolas	45	9	14	8	33	64
	McKay, Donald	20	2	1	9	9	21
	Miller, Oskar von	23	4	5	9	7	25
	Moore, Gordon	61	26	13	10	17	66
	Noyce, Robert	80	31	18	11	16	76
	Shaun, Donovan	16	16	12	6	16	50
	Sikorsky, Igor	130	17	6	23	32	78
	Stevens, John Frank	46	11	3	8	9	31
	Tesla, Nikola	109	36	6	24	19	85
	Torvalds, Linus	48	4	17	4	9	34
	Wankel, Felix	49	5	10	13	11	39
	Wozniak, Steve	76	19	16	11	17	63
	Σ	1332	337	192	240	317	1086
organizacije	Aerial Experiment Association	62	24	12	5	12	53
	Apple Inc.	286	30	96	4	41	171
	Association for Computing Machinery	22	1	7	3	6	17
	California Institute of Technology	30	5	8	12	5	30
	Cisco Systems	48	10	22	6	14	52
	Dell Inc.	39	1	18	4	14	37
	Franklin Institute	16	2	4	4	6	16
	General Electric	40	2	20	4	10	36
	Hitachi, Ltd.	80	3	42	17	22	84
	Institute of Electrical and Electronics Engineers	7	0	3	1	3	7
	Intel	158	25	81	6	32	144
	Johns Hopkins University	28	1	11	12	6	30
	Leonardo da Vinci Museum of Science and Technology	14	4	1	2	3	10

	Museum of Science and Industry	19	1	4	6	4	15
	National Science Museum	5	0	0	1	1	2
	Ontario Science Centre	11	0	2	3	2	7
	Scientific American	25	6	0	3	6	15
	Saab AB	38	0	26	1	11	38
	Texas Tech University	32	0	10	8	10	28
	TRW Inc.	74	11	32	9	23	75
	Xerox Corporation	74	11	29	5	22	67
	Σ	1108	137	428	116	253	934
opći pojmovi	aqueduct	89	0	0	22	9	31
	automobile	1676	122	143	111	212	588
	bionics	117	0	0	0	1	1
	forging	26	0	0	0	0	0
	radio technology	243	19	4	14	18	55
	revolver	29	1	3	1	6	11
	robot	216	11	28	11	36	86
	security and protection system	193	3	0	5	2	10
	ship	2357	82	55	260	190	587
	telecomunication	633	9	0	2	6	17
	telephone	964	47	31	55	113	246
	Wi-Fi	50	1	4	0	4	9
	Σ	6593	295	268	481	597	1641
Σ		9033	769	888	837	1167	3661

Prilog 7. Rezultati kvantitativne analize sadržaja uzorka digitalnoga izdanja Der Brockhaus multimedial 2008 premium

Der Brockhaus multimedial 2008 premium							
Kategorije natuknica	Natuknica	Opseg (redaka)	Broj naziva				
			OS	ORG	LOK	VO	Σ
osobe	Benz, Carl Friedrich	7	1	4	3	6	14
	Colt, Samuel	3	1	0	3	2	6
	Daimler, Gottlieb Wilhelm	14	5	2	3	12	22
	Diesel, Rudolf	8	2	3	2	5	12
	Einstein, Albert	154	37	9	19	63	128
	Grundig, Max	10	2	3	3	10	18
	Hughes, David Edward	6	2	0	2	4	8
	Humann, Carl	5	0	0	4	7	11
	Martens, Adolf	4	1	0	4	2	7
	Morse, Samuel Finley Breese	10	2	1	9	7	19
	Parsons, Sir Charles Algernon	6	1	0	2	4	7
	Schwarz, David	4	2	0	2	4	8
	Siemens, Ernst Werner von	48	12	3	10	18	43
	Tesla, Nikola	16	5	1	8	7	21
	Watt, James	8	0	0	3	6	9
	Wright, Frank Lloyd	50	9	0	34	40	83
	Zeiß, Carl	4	3	2	2	4	11
	Zeppelin, Ferdinand Graf von	12	3	1	4	4	12
	Σ	369	88	29	117	205	439
organizacije	Daimler-Benz	167	18	50	21	63	152
	Audi AG	7	0	7	1	4	12
	Deutsche Telekom AG	30	0	17	1	7	25
	Deutsches Museum von Meisterwerken der Naturwissenschaft und Tęchnik	8	1	2	1	2	6
	General Motors Corporation	14	0	7	6	2	15
	Institut für Radioastronomie im Millimeterbereich	8	0	5	6	2	13
	International Organization for Standardization	7	0	2	1	1	4
	Max Planck Gesellschaft zur Förderung der Wissenschaften e. V.	17	1	5	2	2	10
	Microsoft	162	5	37	14	33	89

	OPEC	38	0	15	6	15	36
	Opel	124	10	33	16	53	112
	Siemens AG	22	2	18	4	6	30
	Bauholding Strabag SE	7	0	8	2	4	14
	VDE	10	0	3	1	1	5
	VDI Verein Deutscher Ingenieure e. V.	29	0	15	2	3	20
	Vereinigte Stahlwerke AG	5	0	2	1	2	5
	Σ	655	37	226	85	200	548
općí pojmovi	Abbrand	9	0	0	0	0	0
	Anker	43	0	0	0	0	0
	Ingenieur	17	0	0	0	0	0
	Dampfmaschine	34	4	0	0	5	9
	Internet	172	0	5	3	10	18
	Kältemaschine	40	0	0	0	0	0
	Kulturtechnik	6	0	0	0	0	0
	Minensuchboot	3	0	0	0	0	0
	Ottomotor	30	1	0	0	0	1
	Programmiersprache	158	0	0	0	3	3
	Solarheizung	13	0	0	0	0	0
	Tachometer	5	0	0	0	0	0
	Technik	127	5	0	9	12	26
	Torpedo	17	0	0	0	0	0
	Wasserbombe	4	0	0	0	0	0
	Rakete	58	4	1	5	10	20
	Σ	736	14	6	17	40	77
Σ		1760	139	261	219	445	1064

Prilog 8. Rezultati kvantitativne analize sadržaja uzorka Wikipedije na hrvatskom jeziku

Wikipedija (Hrvatski jezik)							
Kategorije natuknica	Natuknica	Opseg (redaka)	Broj naziva				
			OS	ORG	LOK	VO	Σ
osobe	Antun Lučić (mineralog)	27	8	2	14	7	31
	Antun Vrančić	40	17	0	15	13	45
	Milan Emil Uzelac	74	13	9	27	27	76
	Faust Vrančić	338	55	15	55	61	186
	Fran Bošnjaković	67	8	16	22	26	72
	Franjo Hanaman	42	9	15	10	11	45
	Ivan Vukić	14	10	4	4	7	25
	Hermann Bollé	121	24	12	3	105	144
	Janko Vuković Podkapelski	289	98	14	46	69	227
	Max Planck	154	57	7	23	45	132
	Nikola Tesla	510	90	24	72	70	256
	Robert Fulton	28	14	0	15	8	37
	Robert Whitehead	40	15	5	14	12	46
	Rudolf Fizir	115	21	15	27	34	97
	Rudolf Perešin	103	9	10	38	22	79
	Eduard Slavoljub Penkala	137	24	10	38	41	113
	Werner von Siemens	13	2	2	4	6	14
	Σ	2112	474	160	427	564	1625
organizacije	Association for Computing Machinery	4	0	3	1	1	5
	Brodosplit	141	2	27	23	70	122
	Hrvatska akademska i istraživačka mreža	199	2	96	19	15	132
	Đuro Đaković Holding	73	0	43	28	4	75
	General Electric	151	19	96	26	47	188
	Hrvatske željeznice	143	0	13	129	9	151
	INA	54	0	37	23	32	92
	Međunarodno društvo za organizaciju znanja	22	0	3	20	2	25
	Jadrolinija	617	19	67	426	381	893
	John Deere	74	11	10	9	16	46
	Končar Elektroindustrija	121	0	44	9	8	61
	Hrvatski željeznički muzej	14	0	5	1	2	8
	Massachusetts Institute of Technology	9	1	4	4	1	10

	Institut Max Planck	8	1	1	2	1	5
	Microsoft	68	2	11	1	6	20
	MOS Technology	35	0	12	0	5	17
	Nacionalni centar za biotehnološke informacije	22	8	8	3	1	20
	Northrop Grumman	20	0	10	2	2	14
	Rimac Automobili	22	1	9	3	6	19
	Science (znanstveni časopis)	10	2	5	0	3	10
	Muzej znanosti	25	2	15	4	7	28
	Brodogradilište Uljanik	46	1	6	6	16	29
	Σ	1878	71	525	739	635	1970
opći pojmovi	Automobil	31	5	0	19	9	33
	Građevinarstvo	103	0	10	10	0	20
	Mikroprocesor	90	2	5	1	8	16
	Minolovac	14	0	0	0	0	0
	Pneumatika	394	0	0	0	0	0
	Sirovine	62	0	0	0	0	0
	Žičara	25	0	0	0	0	0
	Kompresor	172	0	0	0	0	0
	Tkalački stan	15	0	0	0	0	0
	Cement	97	0	4	1	0	5
	Vlak	28	0	0	0	0	0
	Σ	1031	7	19	31	17	74
Σ		5021	552	704	1197	1216	3669

Prilog 9. Rezultati kvantitativne analize sadržaja uzorka Wikipedije na engleskom jeziku

Wikipedija (Engleski jezik)							
Kategorije natuknica	Natuknica	Opseg (redaka)	Broj naziva				
			OS	ORG	LOK	VO	Σ
osobe	Hermann Bollé	24	6	3	12	8	29
	Fran Bošnjaković	45	3	10	14	11	38
	Rudolf Fizir	39	5	6	11	10	32
	Robert Fulton	274	90	28	207	53	378
	Franjo Hanaman	12	4	0	9	4	17
	Anthony Francis Lucas	90	27	8	44	13	92
	Giovanni Luppis	89	29	3	10	12	54
	Slavoljub Eduard Penkala	38	11	5	4	6	26
	Rudolf Perešin	39	5	8	18	13	44
	Max Planck	476	198	40	36	116	390
	Werner von Siemens	99	36	7	21	36	100
	Nikola Tesla	1292	476	86	114	169	845
	Emil Uzelac	75	15	13	19	20	67
	Antun Vrančić	90	53	2	40	27	122
	Fausto Veranzio	195	57	6	45	37	145
	Janko Vuković	60	24	6	4	10	44
	Robert Whitehead	155	69	18	36	38	161
	Σ	3092	1108	249	644	583	2584
organizacije	Association for Computing Machinery	293	32	50	10	48	140
	Brodosplit	148	4	10	22	56	92
	CARNet	190	0	48	26	13	87
	Croatian Railway Museum	12	0	3	6	3	12
	Croatian Railways	361	0	24	239	34	297
	Đuro Đaković (company)	152	4	18	9	42	73
	General Electric	647	40	375	94	163	672
	INA d.d.	116	0	38	33	30	101
	International Society for Knowledge Organization	47	1	19	20	4	44
	Jadrolinija	67	1	4	129	92	226
	John Deere	584	69	200	248	76	593
	KONČAR Group	35	1	9	3	3	16
	Hrvatski Telekom	21	0	17	0	5	22
	Max Planck Society	300	58	149	53	111	371

	Microsoft	801	55	309	56	187	607
	MOS Technology	231	7	46	11	18	82
	National Center for Biotechnology Information	115	9	31	2	4	46
	Northrop Grumman	428	31	219	82	111	443
	Rimac Automobili	165	17	34	14	24	89
	Science (journal)	107	25	18	7	25	75
	Science Museum, London	143	45	54	20	81	200
	Uljanik	48	1	6	6	15	28
	Σ	5011	400	1681	1090	1145	4316
opći pojmovi	Cable car	9	0	0	0	2	2
	Car	648	67	95	74	116	352
	Cement	643	30	8	76	57	171
	Civil engineering	407	11	11	25	17	64
	Compressor	526	0	3	4	2	9
	Loom	289	12	3	15	20	50
	Microprocessor	761	38	155	4	108	305
	Minesweeper	103	6	9	12	5	32
	Natural resource	220	3	9	15	3	30
	Pneumatics	104	1	0	1	3	5
	Train	454	1	18	46	13	78
	Σ	4164	169	311	272	346	1098
Σ		12267	1677	2241	2006	2074	7998

Prilog 10. Rezultati kvantitativne analize sadržaja uzorka Wikipedije na njemačkom jeziku

Wikipedija (Njemački jezik)							
Kategorije natuknica	Natuknica	Opseg (redaka)	Broj naziva				
			OS	ORG	LOK	VO	Σ
osobe	Anthony Francis Lucas	84	22	5	36	15	78
	Antun Vrančić	37	24	0	26	10	60
	Emil Uzelac	49	6	5	12	13	36
	Fausto Veranzio	108	21	0	38	26	85
	Fran Bošnjaković	78	25	26	17	52	120
	Franjo Hanaman	39	6	8	11	28	53
	Giovanni Luppis	46	13	0	9	7	29
	Hermann Bollé	60	13	3	25	13	54
	Janko Vuković-Podkapelski	31	6	2	10	6	24
	Max Planck	1040	341	96	93	215	745
	Nikola Tesla	484	109	49	94	119	371
	Robert Fulton (Ingenieur)	29	11	3	13	10	37
	Robert Whitehead (Ingenieur)	65	36	5	19	20	80
	Rudolf Fizir	47	7	5	23	14	49
	Rudolf Perešin	109	12	9	22	30	73
	Slavoljub Eduard Penkala	118	15	6	26	16	63
	Werner von Siemens	416	116	44	105	113	378
	Σ	2840	783	266	579	707	2335
organizacije	Association for Computing Machinery	100	10	6	1	12	29
	Brodosplit	40	0	5	6	17	28
	CARNet	20	0	9	8	4	21
	Đuro Đaković (Unternehmen)	22	2	8	6	4	20
	General Electric	216	33	97	51	91	272
	Hrvatske željeznice	162	0	21	81	19	121
	Industrija nafte	115	2	48	46	41	137
	International Society for Knowledge Organization	20	1	4	0	2	7
	Jadrolinija	47	0	6	148	104	258
	John Deere	403	11	130	186	32	359
	Končar Group	17	0	7	2	4	13
	Kroatisches Eisenbahnmuseum	16	0	5	2	2	9
	Massachusetts Institute of Technology	394	182	41	55	303	581

	Max-Planck-Gesellschaft	605	59	192	45	141	437
	Microsoft	1074	67	252	46	159	524
	MOS Technology	91	6	30	8	12	56
	National Center for Biotechnology Information	17	0	4	3	1	8
	Northrop Grumman	48	3	23	4	17	47
	Rimac Automobili	117	11	30	10	14	65
	Science	86	16	15	4	22	57
	Science Museum	51	3	14	8	7	32
	Uljanik	38	0	9	4	12	25
	Σ	3699	406	956	724	1020	3106
opći pojmovi	Automobil	381	7	19	33	37	96
	Bauingenieurwesen	290	5	3	21	23	52
	Mikroprozessor	586	4	55	1	31	91
	Minenabwehrfahrzeug	220	0	2	13	12	27
	Pneumatik	509	0	2	2	5	9
	Rohstoff	294	0	29	17	10	56
	Seilbahn	327	26	32	42	13	113
	Verdichter	217	1	0	0	3	4
	Webstuhl	84	2	1	12	10	25
	Zement	455	8	3	53	26	90
	Zug (Schienenverkehr)	90	0	3	6	0	9
	Σ	3453	53	149	200	170	572
Σ		9992	1242	1371	1503	1897	6013

11. POPIS OZNAKA

A – točnost

BRO – Der Brockhaus multimadial 2008 premium

EB – Encyclopaedia Britannica

F1 – F mjera

HBL – Hrvatski biografski leksikon

HE – Hrvatska enciklopedija

HTE – Hrvatska tehnička enciklopedija

IE – Istarska enciklopedija

LOK – lokacije

ORG – organizacije

OS – osobe

P – preciznost

PE – Proleksis enciklopedija

R – odziv

VO – vremenska obilježja

W_{de} – njemačka Wikipedija

W_{en} – engleska Wikipedija

W_{hr} – hrvatska Wikipedija

x – aritmetička sredina

12. POPIS SLIKA

Slika 1. Udio pogrešaka u uzorku hrvatske Wikipedije (Kubelka, Šoštarić, 2011)..... 16

Slika 2. Udio pogešaka u uzorku Hrvatske enciklopedije (Kubelka, Šoštarić, 2011) 16

Slika 3. Shematski prikaz rada na projektu Hrvatske tehničke enciklopedije 25

Slika 4. Rast opsega Hrvatske enciklopedije za razdoblje 2014–19..... 32

Slika 5. Grafički prikaz opsega engleske inačice Wikipedije izvedenoga po enciklopedijskim svescima (Izvor: https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia)..... 33

Slika 6. Omjer prisutnosti pojedine vrste uređaja prigodom pristupa mrežnom sadržaju Hrvatske tehničke enciklopedije za razdoblje od 31. 5. 2019. do 28. 8. 2019. s pokazateljima tendencija rasta ili pada u odnosu na istovjetno prethodno razdoblje prikazan aplikacijom Google Analytics 35

Slika 7. Gledišta postizanja funkcionalne strukturiranosti 40

Slika 8. Početna mrežna stranica Hrvatske tehničke enciklopedije i Portala hrvatske tehničke baštine (Izvor: https://tehnika.lzmk.hr/) 42

Slika 9. Prikaz korisničkoga sučelja i digitaliziranoga primjerka Tehničke enciklopedije kao dijela Portala hrvatske tehničke baštine (Izvor: https://tehnika.lzmk.hr/tehnicka-enciklopedija-1966-1997/) 43

Slika 10. Grafički prikaz članka Armanda, Adam Hrvatske tehničke enciklopedije s pripadajućim infookvirom (Izvor: https://tehnika.lzmk.hr/armanda-adam/)..... 45

Slika 11. Rubrika Države svijeta Proleksis enciklopedije Leksikografskog zavoda Miroslav Krleža (Izvor: https://proleksis.lzmk.hr/drzave-svijeta/) 47

Slika 12. Primjer strukturiranja korisničkoga sučelja namijenjenoga prikazivanju biografskih članaka Encyclopaedije Britannice Online (Izvor: https://www.britannica.com/).. 48

Slika 13. Organizacijska struktura sadržaja Wikipedije na temelju implementiranoga sustava kategorija (Voss, 2006)..... 50

Slika 14. Primjeri infookvira natuknica Nikola Tesla i Andrija Mohorovičić hrvatske Wikipedije (Izvor: https://hr.wikipedia.org/) 52

Slika 15. Shema mapiranja strukturnih elemenata dvaju sustava 58

Slika 16. Skupovi podataka objavljeni u okviru semantičkoga weba s međusobnim vezama (Izvor: https://en.wikipedia.org/wiki/Linked_data) 60

Slika 17. Shematski prikaz RDF trojke (tripleta) 62

Slika 18. Arhitektura semantičkoga weba (Izvor: https://www.obitko.com/tutorials/ontologies-semantic-web/semantic-web-architecture.html) 63

Slika 19. Sastavnice strukturiranoga sadržaja Wikidate (Izvor: https://en.wikipedia.org/wiki/Wikidata) 65

Slika 20. Prikaz temeljnih koncepata tematskih mapa (Izvor: https://en.wikipedia.org/wiki/Topic_map) 66

Slika 21. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Hrvatske tehničke enciklopedije 83

Slika 22. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija Hrvatske tehničke enciklopedije 85

Slika 23. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka općih pojmova Hrvatske tehničke enciklopedije 86

Slika 24. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Hrvatske enciklopedije 88

Slika 25. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija Hrvatske enciklopedije 89

Slika 26. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka općih pojmova Hrvatske enciklopedije 90

Slika 27. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Proleksis enciklopedije 92

Slika 28. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija Proleksis enciklopedije 93

Slika 29. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Istarske enciklopedije 95

Slika 30. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija Istarske enciklopedije 96

Slika 31. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Hrvatskoga biografskog leksikona 99

Slika 32. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba Encyclopaedije Britannice 101

Slika 33. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija Encyclopaedije Britannice 102

Slika 34. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka općih pojmova Encyclopaedije Britannice 103

Slika 35. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba izdanja Der Brockhaus multimedial 2008 premium 104

Slika 36. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija izdanja Der Brockhaus multimedial 2008 premium 105

Slika 37. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba izdanja hrvatske, engleske i njemačke inačice Wikipedije 108

Slika 38. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija izdanja hrvatske, engleske i njemačke inačice Wikipedije .. 110

Slika 39. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka općih pojmova izdanja hrvatske, engleske i njemačke inačice Wikipedije 111

Slika 40. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka osoba cjelokupnog uzorka 113

Slika 41. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka organizacija cjelokupnog uzorka 115

Slika 42. Grafički prikaz zastupljenosti vrsta i tipova naziva promatranih analizom sadržaja članaka općih pojmova cjelokupnog uzorka..... 117

Slika 43. Primjer označenoga dijela teksta enciklopedijskoga članka Jecić, Stjepan nakon procesa automatskog prepoznavanja nazivlja s pomoću alata CroNER..... 125

Slika 44. Primjer označenoga dijela teksta enciklopedijskoga članka Bazjanac, Davorin nakon procesa automatskog prepoznavanja nazivlja s pomoću alata ReLDI..... 127

Slika 45. Primjer označenoga teksta dijela enciklopedijskoga članka Apple Inc. nakon procesa automatskog prepoznavanja nazivlja s pomoću alata Stanford CoreNLP 129

Slika 46. Grafički prikaz odnosa svih elemenata uzorka koji utječu na izvedbu sustava za automatsko prepoznavanje nazivlja (Izvor: https://en.wikipedia.org/wiki/Precision _and_recall) 130

Slika 47. Shemtski prikaz i značajke modela enciklopedičke interoperabilnosti 145

Slika 48. Obrazac za unos podataka (engl. *Airtable Form*) Hrvatske tehničke enciklopedije..... 148

Slika 49. Prikaz dijela sučelja modela enciklopedičke interoperabilnosti 149

13. POPIS TABLICA

Tablica 1. Usporedba opsega sadržaja hrvatske Wikipedije i Hrvatske enciklopedije (Kubelka, Šoštarić, 2011) 15

Tablica 2. Omjer broja naziva i opsega uzorka dijela mrežnih enciklopedičkih izdanja zastupljenih u istraživanju 26

Tablica 3. Broj ažuriranih i novih objavljenih članaka Hrvatske enciklopedije Leksikografskog zavoda Miroslav Krleža 30

Tablica 4. Svojstva i uporaba tipova metapodataka (Riley, 2017) 37

Tablica 5. Raspodjela glavnih područja Hrvatske tehničke enciklopedije po svescima (Jecić, 2018) 41

Tablica 6. Odabrane sheme metapodataka u informacijskim znanostima 56

Tablica 7. Prikaz kategorija članaka s pripadajućim faktografskim podatcima struke metalurgija Hrvatske enciklopedije 77

Tablica 8. Vrste i tipovi naziva zabilježeni ispitivanjem s pripadajućim značenjima kategorije članaka osoba 80

Tablica 9. Vrste i tipovi naziva zabilježeni ispitivanjem s pripadajućim značenjima kategorije članaka organizacija 81

Tablica 10. Vrste i tipovi naziva zabilježeni ispitivanjem s pripadajućim značenjima kategorije članaka organizacija 82

Tablica 11. Zastupljenost tipova naziva kategorije članaka osoba svih zastupljenih izdanja uzorka uz navedene srednje vrijednosti 114

Tablica 12. Zastupljenost tipova naziva kategorije članaka organizacija svih zastupljenih izdanja uzorka uz navedene srednje vrijednosti 116

Tablica 13. Zastupljenost tipova naziva kategorije članaka općih pojmova svih zastupljenih izdanja uzorka uz navedene srednje vrijednosti 118

Tablica 14. Kategorije naziva s pripadajućim primjerima relevantne za ispitivanje označene alatom naziva CroNER 121

Tablica 15. Broj ručno označenih naziva prema kategorijama i jedinicama sadržaja enciklopedičkih članaka, vrstama i tipovima naziva i postotak njihove zastupljenosti u tekstu ovisno o broju pojavnica za svaku kategoriju i jedinicu sadržaja uzorka na hrvatskom jeziku 133

Tablica 16. Broj ručno označenih naziva prema kategorijama i jedinicama sadržaja enciklopedičkih članaka, vrstama i tipovima naziva i postotak njihove zastupljenosti u tekstu ovisno o broju pojavnica za svaku kategoriju i jedinicu sadržaja uzorka na engleskom jeziku 134

Tablica 17. Rezultati ispitivanja evaluacijskih mjera CroNER alata na enciklopedičkome tekstu 136

Tablica 18. Rezultati ispitivanja evaluacijskih mjera ReLDI alata na enciklopedičkome tekstu 137

Tablica 19. Rezultati ispitivanja evaluacijskih mjera Stanford CoreNLP alata na enciklopedičkome tekstu 138

Tablica 20. Rezultati ispitivanja evaluacijskih mjera za određene vrste naziva alata CroNER, ReLDI i Stanford CoreNLP 140

Tablica 21. Kvantitativni pokazatelji omjera naziva i strukturnih vrijednosti Proleksis enciklopedije implementiranih u model enciklopedičke interoperabilnosti 152

Tablica 22. Kvantitativni pokazatelji omjera naziva i strukturnih vrijednosti Hrvatske tehničke enciklopedije implementiranih u model enciklopedičke interoperabilnosti 153

Tablica 23. Kvantitativni pokazatelji omjera naziva i strukturnih vrijednosti Encyclopaedije Britannice implementiranih u model enciklopedičke interoperabilnosti 154

14. POPIS PRILOGA

Prilog 1. Rezultati kvantitativne analize sadržaja uzorka Hrvatske tehničke enciklopedije 174

Prilog 2. Rezultati kvantitativne analize sadržaja uzorka Hrvatske enciklopedije 176

Prilog 3. Rezultati kvantitativne analize sadržaja uzorka Proleksis enciklopedije 178

Prilog 4. Rezultati kvantitativne analize sadržaja uzorka Istarske enciklopedije 180

Prilog 5. Rezultati kvantitativne analize sadržaja uzorka Hrvatskoga biografskog leksikona 182

Prilog 6. Rezultati kvantitativne analize sadržaja uzorka Encyclopaedije Britannice 183

Prilog 7. Rezultati kvantitativne analize sadržaja uzorka digitalnoga izdanja Der Brockhaus multimedial 2008 premium 185

Prilog 8. Rezultati kvantitativne analize sadržaja uzorka Wikipedije na hrvatskom jeziku 187

Prilog 9. Rezultati kvantitativne analize sadržaja uzorka Wikipedije na engleskom jeziku 189

Prilog 10. Rezultati kvantitativne analize sadržaja uzorka Wikipedije na njemačkom jeziku 191

15. ŽIVOTOPIS

Ivan Smolčić rođen je 15. studenoga 1987. godine u Zagrebu. Nakon završene osnovne škole, u Velikoj Gorici završio je prirodoslovno-matematičku gimnaziju. Studij rudarstva na Rudarsko-geološko-naftnome fakultetu u Zagrebu upisao je 2008. godine, a diplomirao je 2013. godine obranivši diplomski rad naslova *Utjecaj energije iniciranja na brzinu detonacije ANFO eksploziva* te stekao akademski naziv magistra inženjera rudarstva.

U Zagrebu je Poslijediplomski specijalistički studij inženjerstva materijala na Fakultetu strojarstva i brodogradnje upisao 2013. godine, a naslov sveučilišnoga specijalista strojarstva i brodogradnje postigao je 2016. godine obranivši završni rad naslova *Utjecaj načina šaržiranja obradaka na svojstva PACVD prevlaka*.

Poslijediplomski doktorski studij informacijskih i komunikacijski znanosti na Filozofskome fakultetu upisao je 2015. godine. Od 2014. godine zaposlenik je Leksikografskog zavoda Miroslav Krleža u redakciji Hrvatske tehničke enciklopedije.

Popis objavljenih radova:

Smolčić, I., Jermen, N., Jecić Z.: Analiza enciklopedičke relevantnosti na primjeru hrvatske tehničke baštine. Studia lexicographica, 9(2015) 1(16), str. 133–160.

Smolćić, I., Jermen, N., Jecić, Z.: Zastupljenost riječke torpedne baštine u enciklopedičkim izdanjima. VII. Međunarodna konferencija o torpednoj baštini: Torpedo – povijest i baština – Zbornik sažetaka. Ur. Smokvina, M. Rijeka: Pro Torpedo Rijeka, 2016., str. 68–68.

Smolčić, I., Tolj, J., Jecić, Z.: Epistemological Value of Contemporary Encyclopedic Projects. INFuture 2017: Integrating ICT in Society. Urednici: Atanassova, J., Zaghouani, W., Kragić, B., Aas, K., Stančić, H., Seljan, S. Zagreb: Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2017., str. 141–149.

Jecić, S., Smolčić, I.: Tehnička visoka škola (1919–26) i Tehnički fakultet (1926–56) u Zagrebu – temelji suvremenoga razvoja tehnike u Hrvatskoj. Studia lexicographica, 12(2018) 23, str. 81–100.

Klobučar Srbić, I., Smolčić, I., Tolj, J.: Bibliogtafija Tehničke enciklopedije (1963–97). Studia lexicographica, 12(2018) 23, str. 199–257.

Tolj, J., Smolčić, I., Jecić, Z.: Enhancing encyclopedic characteristics using geotagging: why it matters?. INFuture 2019: Knowledge in the Digital Age. Urednici: Bago, P., Hebrang Grgić, I., Ivanjko, T., Juričić, V., Miklošević, Ž., Stublić, H. Zagreb: Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2019., str. 33–38.

Smolčić, I., Jecić, Z., Bago, P.: Tablični prikazi kao prilozi strukturiranosti enciklopedičkih djela. Rasprave, 46(2020) 2, str. 1019–1038.

16. CURRICULUM VITAE

Ivan Smolčić was born on November 15th 1987 in Zagreb. After Elementary school, he graduated natural sciences and mathematics high school in Velika Gorica. In 2008 he enrolled at the Faculty of Mining, Geology and Petroleum Engineering in Zagreb and graduated in 2013. with master's thesis titled *Initiation energy impact on velocity of detonation of ANFO explosive*, acquiering academic title of Master of Mining Engineering.

The same year he enrolled in the Postgraduate specialist studies of materials engineering at the Faculty of Mechanical Engineering and Naval Architecture in Zagreb where he graduated in 2016. with tesis *The influence of sample batching method on the properties of PACVD coatings*.

He enrolled in Postgraduate doctoral studies of information and communication sciences at the Faculty of Humanities and Social Sciences in Zagreb in 2015. Since 2014 he has been an employee at the Miroslav Krleža Institute of Lexicography as a redaction member of the Croatian encyclopaedia of technology.

List of published papers:

Smolčić, I., Jermen, N., Jecić, Z.: Analiza enciklopedičke relevantnosti na primjeru hrvatske tehničke baštine. Studia lexicographica, 9(2015) 1(16), pg. 133–160.

Smolćić, I., Jermen, N., Jecić, Z.: Zastupljenost riječke torpedne baštine u enciklopedičkim izdanjima. 7th International conference on industrial heritage: Torpedo – history and heritage – Collection of summaries. Ur. Smokvina, M. Rijeka: Pro Torpedo Rijeka, 2016., pg. 68–68.

Smolčić, I., Tolj, J., Jecić, Z.: Epistemological Value of Contemporary Encyclopedic Projects. INFuture 2017: Integrating ICT in Society. Editors: Atanassova, J., Zaghouani, W., Kragić, B., Aas, K., Stančić, H., Seljan, S. Zagreb: Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2017., pg. 141–149.

Jecić, S., Smolčić, I.: Tehnička visoka škola (1919–26) i Tehnički fakultet (1926–56) u Zagrebu – temelji suvremenoga razvoja tehnike u Hrvatskoj. Studia lexicographica, 12(2018) 23, pg. 81–100.

Klobučar Srbić, I., Smolčić, I., Tolj, J.: Bibliogtafija Tehničke enciklopedije (1963–97). Studia lexicographica, 12(2018) 23, pg. 199–257.

Tolj, J., Smolčić, I., Jecić, Z.: Enhancing encyclopedic characteristics using geotagging: why it matters?. INFuture 2019: Knowledge in the Digital Age. Editors: Bago, P., Hebrang Grgić, I., Ivanjko, T., Juričić, V., Miklošević, Ž., Stublić, H. Zagreb: Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2019., pg. 33–38.

Smolčić, I, Jecić, Z., Bago, P.: Tablični prikazi kao prilozi strukturiranosti enciklopedičkih djela. Rasprave, 46(2020) 2, pg. 1019–1038.