

Propagandos technikų lingvistinių požymių lyginamoji analizė taikant natūralios kalbos apdorojimo metodus

Evelina Vaitkevičiūtė, Gražina Korvel

Vilniaus universitetas, Duomenų mokslo ir skaitmeninių technologijų institutas,
Akademijos g. 4, LT-08412 Vilnius
evelina.vaitkeviciute@mif.vu.lt

Santrauka. Pastaruoju metu dėl besitęsiančių geopolitinių konfliktų ir vis labiau populiarėjančio dirbtinio intelekto, propagandos technikų analizė tapo svarbia tyrimų sritimi. Nors, nemažai propagandos technikų lingvistinės analizės tyrimų atlikta angliškuose tekstuose, tačiau palyginti nedaug jų atlikta su mažiau išteklių turinčiomis kalbomis. Šiame darbe buvo siekiama atlikti propagandos technikų lingvistinių požymių lyginamąją analizę ir taip nustatyti bendrąsias ir unikaliąsias kiekvienos technikos savybes. Gauti rezultatai rodo, kad tarp skirtingų propagandos technikų būdingi ryškūs technikų fragmentų kiekio, skaitinių požymių, kalbos dalių pasiskirstymo ir dažniausiai pasitaikančių žodžių skirtumai.

Raktiniai žodžiai: propagandos technikos; lingvistinė analizė; natūralios kalbos apdorojimas; kalbos dalių žymėjimas; n-gramos; raktinių frazių išskyrimas.

1 Įvadas

Pastaruoju metu internetinėje erdvėje propaganda tapo galinga priemone manipuliuoti visuomenės sąmone [1]. Žiniasklaidoje ir socialiniuose tinkluose sutinkami propagandinio pobūdžio šaltiniai prisideda prie tikrovės iškreipimo, dezinformacijos skleidimo ir pasitikėjimo informacijos šaltiniais mažinimo [2].

Šiuo požiūriu labai svarbus uždavinys yra nuolatinė propagandos technikų savybių analizė ir šių technikų nustatymo tekste priemonių tobulinimas, kas galėtų padidinti informacijos saugumo lygį ir leistų veiksmingai kovoti su dezinformacija. Pasitelkiant natūralios kalbos apdorojimo metodus, atliekama lingvistinių požymių analizė, leidžianti rasti kiekvienai technikai būdingas savybes [2]. Vienas iš tokių metodų – kalbos dalių žymėjimas, kuriu analizuojama leksinė ir sintaksinė informacija [3]. Dažnai lingvistinių

požymių išskyrimui ir analizei taip pat pasitelkiami n -gramų junginiai [1]. Nemažai lingvistinės analizės tyrimų atlikta angliškuose tekstuose, tačiau dėl sudėtingos kalbų morfologijos, šių tyrimų vis dar trūksta mažiau išteklių turinčioms kalboms [3, 4].

Šiame darbe nagrinėjamos 10 propagandos technikų „Emocinė raiška“, „Whataboutism / Red Herring / Straw Man“, „Supaprastinimas“, „Neapibrėžtumas“, „Apeliavimas į autoritetą“, „Mojavimas vėliava“, „Sekimas iš paskos“, „Abejojimas“, „Reductio ad Hitlerum“ ir „Pakartojimas“ [5, 6]. Pateikiama šių propagandos technikų lingvistinių požymių lyginamoji analizė, kurios tikslas – nustatyti dažniausiai naudojamą žodyną bei žodžių junginių struktūras būdingus kiekvienai propagandos technikai ir taip rasti bendrąsias ir unikalias kiekvienos technikos savybes.

2 Duomenys

Šiame darbe naudotą duomenų rinkinį sudaro 682 lingvistikos ir socialinių mokslų tyrėjų anotuoti naujienų straipsniai, surinkti iš lietuviškų naujienų šaltinių, vykdamas Vilniaus universiteto propagandos ir dezinformacijos tyrimų projektą ATSPARA [7]. Naujienų šaltiniai pasirinkti remiantis Lietuvos nacionalinio radijo ir televizijos atliktu tyrimu [8]. Kadangi šio darbo rengimo metu projekto tekstynas vis dar yra jo sudarymo procese, todėl naudota paskutinė naujausia iki 2024 m. spalio mėn. surinkto duomenų rinkinio versija.

Teksto anotavimo procesą sudarė propagandinio pobūdžio teksto fragmentų nustatymas bei atitinkamos propagandos technikos priskyrimas konkrečiam fragmentui. Kiekvienam teksto fragmentui, taikant kryžminio anotavimo metodą buvo priskiriama bent viena iš 10-ies skirtingų propagandos technikų.

3 Metodologija

Bendrą tyrimo eigos schemą sudaro: pradinė duomenų analizė, pirminis tekstinių duomenų apdorojimas, lingvistinių požymių analizė ir gautų rezultatų vertinimas.

3.1 Pirminis tekstinių duomenų apdorojimas

Pirmiausia, siekiant tolesnei analizei paruošti tekstinius duomenis atliekamas pirminis teksto apdorojimas. Šis etapas apima tokius veiksmus kaip

didžiųjų raidžių keitimas į mažąsias raides, specialiųjų simbolių (grotelės #, \$, %, @, &), svetainių nuorodų („http“, „https“ ir „www“), skaičių, nelietuviškų raidžių bei skyrybos ženklų šalinimas, ir pan. Taip pat pašalinami dvigubi tarpai, žodžiai, sudaryti iš vienos raidės ir nesuanuototos tuščios eilutės duomenų rinkinyje.

Toliau šiame etape tekstas suskaidomas į atskirus požymius, vadinamus teksto vienetais (angl. *tokens*), t. y. žodžiai atskiriami vienas nuo kito, atsižvelgiant į tarp žodžių esančius tarpus.

Iš teksto pašalinami nereikšmingi žodžiai. Nereikšmingais laikomi dažnai pasikartojantys žodžiai, tai gali būti prielinksniai, jungtukai ar kiti panašūs žodžiai, kurie nesuteikia jokios naudingos informacijos, kuri būtų reikšminga konstruojant duomenų modelį [9].

3.2 Lingvistinių požymių analizė

Atlikus pradinę analizuojamų duomenų analizę ir pirminį teksto apdorojimą, kitas žingsnis – propagandos technikų lingvistinių požymių analizė. Natūraliosios kalbos apdorojimo metodai, tokie kaip stilistinių požymių išskyrimas, *n*-gramų junginių sudarymas ir kalbos dalių žymėjimas naudojami siekiant išskirti pagrindinius kalbinius požymius, kurie leistų lengviau nustatyti konkrečius žodžius, dažniausiai siejamus su kiekviena propagandos technika [10].

Tekstinių duomenų požymiams išskirti duomenys gali būti atvaizduojami realiųjų reikšmių vektoriais arba tam gali būti naudojami skiriamieji stilistiniai teksto požymiai, tokie kaip teksto ilgis (teksto simbolių ar žodžių skaičius), skyrybos ženklų skaičius, vidutinis žodžių ilgis ar nereikšmingų žodžių kiekis tekste [1, 11]. Žodžių ilgio analizė leidžia nustatyti žodžių ilgį tiek paties žodžio, tiek pagrindinės žodžių formos lygmenimis, tuo tarpu sakinių ilgio analizė vykdoma aukštesniu lygmeniu pagal žodžių skaičių [1].

Duomenų požymiams parinkti taip pat gali būti taikomas *n*-gramų junginių metodas. *N*-gramos – tekstinių dokumentų sekos, žodžių arba simbolių junginiai, sudaryti iš *n* gretimų žodžių arba simbolių, dar dažnai vadinami pastoviųjų žodžių junginiais (angl. *multiword expressions*) arba leksinėmis samplaikomis (angl. *lexical bundles*) [12, 13]. *N*-gramos gali būti sudarytos pagal įvairias dedamąsias (raides, skiemenis), tačiau dažniausiai sudaromos žodžių ir pagrindinių žodžių formų *n*-gramos. Sakiniai yra padalijami į žodžius, o žodžiai sugrupuojami pagal iš anksto apibrėžtus *n*-gramų junginius. Specialus *n*-gramų atvejis, kai sekos sudarytos iš vieno žodžio, vadinamas

unigramomis ($n = 1$). Dvižodžiai junginiai vadinami bigramomis ($n = 2$), o junginiai, sudaryti iš trijų žodžių – trigramomis ($n = 3$) [13].

4 Eksperimento rezultatai

Iš viso duomenų rinkinyje yra 10265 propagandinio pobūdžio teksto fragmentų, kuriems priskirta atitinkama propagandos technika. Lyginant skirtingas propagandos technikas stebimi ryškūs technikų fragmentų kiekio skirtumai. Gauti rezultatai rodo, kad dažniausia propagandos technika „Emocinė raiška“ pasikartoja 3978 kartus ir sudaro 38,8 % visų fragmentų, o rečiausia „Sekimas iš paskos“, pasikartoja 79 kartus ir sudaro 0,8 % visų fragmentų.

Taip pat atsiskleidžia propagandos technikų skaitinių teksto požymių skirtumai. Propagandos technikų skaitinių teksto požymių statistika, vaizduojanti vidutinį simbolių, žodžių ir skyrybos ženklų skaičių, vidutinį žodžių ilgį ir nereikšmingų žodžių kiekį bei standartinį nuokrypį, pateikta 1 lentelėje.

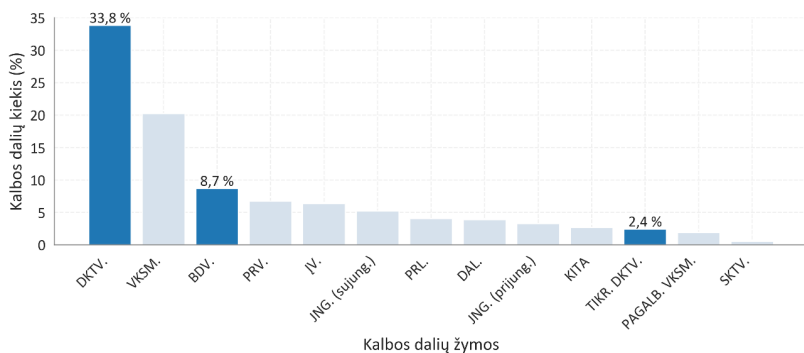
1 lentelė. Propagandos technikų skaitinių teksto požymių statistika.

Technika	Simbolių skaičius	Žodžių skaičius	Skyrybos ženklų skaičius	Vidutinis žodžių ilgis	Nereikšmin- gų žodžių kiekis
Emocinė raiška	227±247	30±34	7±9	6±1	9±11
Whataboutism / Red Herring / Straw Man	267±277	36±38	9±11	6±1	11±12
Supaprastinimas	211±327	28±43	6±10	6±1	8±13
Neapibrėžtumas	175±209	24±28	5±6	6±1	7±9
Apeliavimas į autoritetą	131±135	17±18	5±7	7±1	4±5
Mojavimas vėliava	106±96	14±13	3±3	6±1	4±4
Sekimas iš paskos	151±231	20±28	4±5	6±1	6±7
Abejojimas	195±201	26±26	6±7	6±1	7±8
Reductio ad Hitlerum	129±152	17±20	3±5	7±2	4±7
Pakartojimas	34±37	4±5	1±1	7±2	1±2

* Žydra spalva pažymėtas didžiausias, šviesiai pilka – mažiausias skaitinių požymių kiekis.

Tarp propagandos technikų ypač išsiskiria „Whataboutism / Red Herring / Straw Man“ ir „Pakartojimas“ technikos. Pavyzdžiui, „Whataboutism / Red Herring / Straw Man“ vidutinis simbolių skaičius yra 267 simboliai su 277 simbolių standartiniu nuokrypiu, tuo tarpu „Pakartojimas“ technikos vidutinis simbolių skaičius yra 34 simboliai su 37 simbolių standartiniu nuokrypiu. Propagandos technikos „Whataboutism / Red Herring / Straw Man“ teksto fragmentuose taip pat stebimas ir didžiausias žodžių skaičius, skyrybos ženklus skaičius bei nereikšmingų žodžių kiekis. „Pakartojimas“ technika nukreipta į tekste kartojamą tą pačią frazę, todėl žodžių skaičius, skyrybos ženklų ir nereikšmingų žodžių kiekis yra mažesnis, bet vidutiniškai tekste žodžiai ilgesni, fragmentus vidutiniškai sudaro 7 žodžiai su 2 žodžių standartiniu nuokrypiu.

Taikant kalbos dalių žymėjimo metodą, nustatyta, kad dažniausiai tekste pasitaikanti kalbos dalis yra daiktavardis (DKTV.), tai sudaro didžiąją dalį, 33,8 % visų kalbos dalių, todėl daiktavardžiai yra išrenkami kaip raktiniai žodžiai. Būdvardžiai ir tikriniai daiktavardžiai, kurie taip pat dažnai laikomi potencialiais raktiniais žodžiais, atitinkamai sudaro 8,7 % ir 2,4 % visų kalbos dalių. Likusių kalbos dalių pasiskirstymas tekste pavaizduotas 1 pav. Žodžiai, kurių apdorojimo tikimybės yra per žemos, nėra apdorojamos ir nurodoma kaip KITA.



1 pav. Kalbos dalių pasiskirstymas tekstinuose fragmentuose. DKTV. – daiktavardis, VKSM. – veiksmažodis, BDV. – būdvardis, PRV. – prieveiksmis, IV. – įvardis, JNG. (sujung. / prijung.) – jungtukas (sujungiamasis / prijungiamasis), PRL. – prielinksnis, DAL. – dalelytė, TIKR. DKTV. – tikrinis daiktavardis, PAGALB. VKSM. – pagalbinis veiksmažodis, SKTV. – skaitvardis.

Toliau 2 lentelėje pateikiama, kokios kalbos dalys yra dažniausiai pasitaikančios atitinkamai technikai. Sužymėjus kiekvienos technikos žodžius pagal dažniausiai pasitaikančias kalbos dalis pastebima, kad didžiausiu daiktavardžių procentiniu kiekiu tekstiniuose fragmentuose išsiskiria „Pakartojimas“ technika, daiktavardžiai šios technikos fragmentuose sudaro 38,2 % visų kalbos dalių žymų. Technikos „Apeliavimas į autoritetą“ tekstuose cituojami garsūs žmonės, autoritetai, todėl juose yra aptinkama daugiau tikrinių daiktavardžių (10,2 %) lyginant su kitomis technikomis. Šiai technikai taip pat būdingas panašus daiktavardžių ir veiksmažodžių procentinis kiekis kaip ir „Reductio ad Hitlerum“. Visgi, „Reductio ad Hitlerum“ technika skiriasi nuo kitų didesniu būdvardžių procentu, aptinkamu fragmentuose (12,4 %). Mažesniu daiktavardžių procentiniu kiekiu, lyginant su kitomis propagandos technikomis, pasižymi „Whataboutism / Red Herring / Straw Man“, „Sekimas iš paskos“, ir „Neapibrėžtumas“. Kadangi technika „Sekimas iš paskos“ nukreipta į apeliavimą į bandos jausmą, fragmentuose labiau išryškėja didesnis procentas veiksmažodžių ir įvardžių.

2 lentelė. Kiekvienos propagandos technikos pasirinktų kalbos dalių procentinis pasiskirstymas.

	DKTV.	VKSM.	BŪDV.	PRV.	ĮV.	TIKR. DKTV.
Emocinė raiška	32,5	20,1	8,4	6,8	6,8	3,2
Whataboutism / Red Herring / Straw Man	31,7	19,6	7,6	6,2	6,9	4,9
Supaprastinimas	32,3	19,9	8,7	7,1	6,5	3,1
Neapibrėžtumas	30,5	21,0	8,6	7,4	6,8	3,0
Apeliavimas į autoritetą	34,2	17,2	7,2	4,6	4,8	10,2
Mojavimas vėliava	36,0	18,2	8,7	4,8	7,2	5,5
Sekimas iš paskos	30,5	22,6	8,3	5,9	8,9	3,1
Abejojimas	34,0	19,2	7,8	6,4	5,7	4,7
Reductio ad Hitlerum	33,5	17,1	12,4	5,8	5,4	5,0
Pakartojimas	38,2	16,9	9,6	4,8	6,4	7,0

* Žydra spalva pažymėtas didžiausias, šviesiai pilka – mažiausias kalbos dalių pasiskirstymas (%).

2 pav. pateikti žodžių debesys, vaizduojantys tarpusavyje išsiskiriančių propagandos technikų 30 dažniausiai pasitaikančių žodžių. Žodžių debesys-

se vaizduojamų žodžių dydis rodo jų dažnumą duomenų rinkinyje. Remiantis žodžių dydžiu, propagandos technikos „Supaprastinimas“ fragmentuose trys svarbiausi žodžiai yra „Lietuvos“, „žmonių“ ir „gali“, kurie duomenų rinkinyje pasitaiko atitinkamai 139, 136 ir 129 kartus (2 pav. A). Žodžiai „Lietuvos“, „žmonių“ rodo, kad kaltė perkeliama Lietuvai, ar žmonėms. Žodis „Lietuvos“ taip pat yra dažniausiai pasikartojęs žodis teknikose: „Emocinė raiška“, „Abejojimas“, „Mojavimas vėliava“, „Apeliavimas į autoritetą“, „Whataboutism / Red Herring / Straw Man“ ir „Sekimas iš paskos“. Likusios trys propagandos technikos išsiskiria nuo anksčiau paminėtų. Propagandos technikoje „Neapibrėžtumas“ dažniausi žodžiai yra „gali“, „nėra“, pasikartojantys 19 ir 16 kartų atitinkamai (2 pav. B). Tuo tarpu, „Reductio ad Hitlerum“ (2 pav. C), technikos dažniausi žodžiai – „revoliucijos“, „komunistinė“. „Pakartojimas“ tekste kartojama ta pati žinutė, dažniausiai vartojant žodžius „komunistas“, „Stambulo“ (2 pav. D).



2 pav. Propagandos technikų dažniausiai pasitaikančius žodžius vaizduojantys žodžių debesy. A – „Supaprastinimas“, B – „Neapibrėžtumas“, C – „Reductio ad Hitlerum“, D – „Pakartojimas“.

Nustatytos dažniausiai naudojamos žodžių junginių struktūros (bigramos). Pavyzdžiui, „Apeliavimas į autoritetą“ technikoje dažniausiai pasitaikančios junginių struktūros yra „užsienio reikalų“, „Anna Louri“ ir „John Hylan“. Šie junginiai vaizduoja autoritetus, viena iš jų, „Anna Louri“ prancūzų žurnalistė, pasakojusi apie karą Donbase (3 pav. D).



3 pav. Propagandos technikų dažniausiai pasitaikančius žodžių junginius (bigramas) vaizduojantys žodžių debesy. A – „Supaprastinimas“, B – „Emocinė raiška“, C – „Mojavimas vėliava“, D – „Apeliavimas į autoritetą“.

Taip pat ieškotos kolokacijos arba žodžių, kurie dažnai pasitaiko kartu, kombinacijos (angl. *Collocations*). 3 lentelėje pateiktos dažniausios kolokacijos (bigramos). Pavyzdžiui, tekstiniuose fragmentuose žodis „konservatorių“ dažnai yra susijęs su žodžiu „valstybė“, taip pat dažnai kartu pasitaiko „Emocinės raiškos“ technikos žodžiai „konstitucinio“ ir „teismo“.

3 lentelė. Dažniausios kolokacijos (bigramos).

Bigramos	Kiekis	Technika	Kiekis
konservatorių valstybė	47	Abejojimas	27
konstitucinio teismo	25	Emocinė raiška	9
sveiko proto	23	Emocinė raiška	12
mokesčių mokėtojų	32	Emocinė raiška	15
užsienio reikalų	29	Emocinė raiška	12

Išvados

Šiame darbe atlikta propagandos technikų lingvistinių požymių lyginamoji analizė parodė bendrąsias ir unikaliąsias kiekvienos technikos savybes. Rezultatai rodo, kad tarp skirtingų propagandos technikų būdingi ryškūs techninių fragmentų kiekio, skaitinių požymių skirtumai. „Whataboutism / Red Herring / Straw Man“ technikos tekstiniai fragmentai ilgesni bei turintys daugiau skyrybos ženklų ir nereikšmingų žodžių, lyginant su kitomis technikomis.

Taikant kalbos dalių žymėjimo metodą pastebima, kad daugiausia daiktavardžių yra „Pakartojimas“ fragmentuose. Mažiau daiktavardžių, lyginant su kitomis propagandos technikomis, aptinkama „Whataboutism / Red Herring / Straw Man“, „Sekimas iš paskos“, ir „Neapibrėžtumas“ technikoje. Technikoje „Apeliavimas į autoritetą“ priešingai negu kitose aptinkama daugiau tikrinių daiktavardžių, tačiau jai būdingas panašus daiktavardžių ir veiksmažodžių kiekis kaip ir „Reductio ad Hitlerum“.

Nustatyta, kad technikoje „Supaprastinimas“ „Emocinė raiška“, „Abejojimas“ „Mojavimas vėliava“, „Apeliavimas į autoritetą“, „Whataboutism / Red Herring / Straw Man“, „Sekimas iš paskos“ stebimas tas pats dažniausiai pasikartojantis žodis – „Lietuvos“. Tuo tarpu, dažniausiai pasitaikantys žodžiai „gali“, „revoliucijos“, „komunistas“ atitinkamai technikoje „Neapibrėžtumas“, „Reductio ad Hitlerum“, „Pakartojimas“ rodo, kad šios technikos išsiskiria nuo anksčiau paminėtų technikų.

Šio darbo rezultatai leidžia išplėsti tyrimą analizuojamoje srityje. Analizės automatizavimo procesas bei natūralios kalbos apdorojimo sprendimų taikymas padėtų tyrėjams palengvinti lingvistikos tyrimus. Tolesnio tyrimo metu propagandos technikų lingvistinei analizei atlikti, planuojama taikyti mašininio mokymo metodus, įtraukiant stilistinius požymius, *n*-gramų junginius ir žodžių dažnius bei kalbos dalių žymėjimo požymius, taip pat transformieriamais pagrįstus, iš anksto apmokytus daugiakalbius modelius.

Padėka. Autoriai dėkoja Lietuvos mokslo tarybos (LMTLT) finansuojamam projektui „Propagandos ir dezinformacijos tyrimai: automatinis atpažinimas mašininio mokymo metodais, poveikis ir visuomenės atsparumas“ (sutarties Nr. S-VIS-23-8) už tyrimui suteiktus duomenis.

Literatūra

- [1] Horák, A., Sabol, R., Herman, O., & Baisa, V. (2024). Recognition of propaganda techniques in newspaper texts: Fusion of content and style analysis. *Expert Systems with Applications*, 251, 124085.
- [2] Ramberdiyeva, G., Dildabekova, A., Abikenova, Z., Karabayeva, L., & Zhuasbaeva, A. (2024). The Functional and Semantic Category of Appeal as a Linguistic Tool in Political Propaganda Texts (in the Example of the English Language). *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique*, 1-14.
- [3] Mishra, A., Shaikh, S. H., & Sanyal, R. (2022). Context based NLP framework of textual tagging for low resource language. *Multimedia Tools and Applications*, 81(25), 35655-35670.
- [4] Mengliev, D., Barakhnin, V., Abdurakhmonova, N., & Eshkulov, M. (2024). Developing named entity recognition algorithms for Uzbek: Dataset insights and implementation. *Data in Brief*, 54, 110413.
- [5] Da San Martino, G., Seunghak, Y., Barrón-Cedeno, A., Petrov, R., & Nakov, P. (2019). Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 5636-5646). Association for Computational Linguistics.
- [6] Barrón-Cedeno, A., Jaradat, I., Da San Martino, G., & Nakov, P. (2019). Propopy: Organizing the news based on their propagandistic content. *Information Processing & Management*, 56(5), 1849-1864.
- [7] Projektas ATSPARA. „Propagandos ir dezinformacijos tyrimai: automatinis atpažinimas mašininio mokymo metodais, poveikis ir visuomenės atsparumas“. Prieiga per internetą: <https://www.atspara.mif.vu.lt> (žiūrėta: 2025-03-06).
- [8] LRT tyrimas. Lietuvos „penktoji kolona“: Rusijos propagandą platina šeimos gynėjai, sekto ir knygų apie Staliną leidėjai. Prieiga per internetą: <https://www.lrt.lt/naujienos/lrt-tyrimai/5/1700792/lrt-tyrimas-lietuvas-penktoji-%20kolona-rusijos-propaganda-platina-seimos-gynejai-sektos-ir-knygu-apie-stalina-leidejai> (žiūrėta: 2025-03-06).

- [9] Ahmad, P. N., Guo, J., AboElenein, N. M., Haq, Q. M. U., Ahmad, S., Algarni, A. D., & A. Ateya, A. (2025). Hierarchical graph-based integration network for propaganda detection in textual news articles on social media. *Scientific Reports*, 15(1), 1827.
- [10] Pradhan, A., & Yajnik, A. (2024). Parts-of-speech tagging of Nepali texts with Bidirectional LSTM, Conditional Random Fields and HMM. *Multimedia Tools and Applications*, 83(4), 9893-9909.
- [11] Leonavičienė, A. (2010). Sakinių ilgis–publicistinio ir šnekamojo stiliaus sandūros tekstuose požymis. *Kalbotyra*, (62 (3)), 95-107.
- [12] Esmaeilzadeh, A., Cacho, J. R. F., Taghva, K., Kambar, M. E. Z. N., & Hajiali, M. (2022, July). Building wikipedia n-grams with apache spark. In *Science and information conference* (pp. 672-684). Cham: Springer International Publishing.
- [13] Kapusta, J., Drlik, M., & Munk, M. (2021). Using of n-grams from morphological tags for fake news classification. *PeerJ Computer Science*, 7, e624.