

O'ZBEK TILI MATNLARIDA UNDOV GAPLARNI TAHLIL QILISHNING QOIDAGA ASOSLANGAN ALGORITMI

Sharipov Maksud Siddiqovich,

Texnika fanlari nomzodi, dotsent

maqsbek72@gmail.com

Urganch davlat universiteti “Kompyuter ilmlari” kafedrasida dotsenti

Adinayev Xushnudbek Saylboyevich,

hushnudbek.adinaev@gmail.com

TATU Urganch filiali o'quv-uslubiy boshqarma boshlig'i

Yusupova Mukaddas Muratovna,

mukaddas0783@gmail.com

Urganch Ranch texnologiya universiteti 1-kurs magistranti

ANNOTATSIYA. Bu ishda biz o'zbek tili matnlari undov gaplarini tahlil qilishning qoidaga asoslangan algoritmini ishlab chiqishni ko'rib chiqamiz. Undov gaplar — gaplovchi shaxsning his-tuyg'ularini (hayrat, quvonch, g'azab, achinish, chaqiriq va h.k.) ifodalovchi gaplar bo'lib, ko'pincha undov so'zlar, emotsional ohang yoki maxsus so'z tartibi bilan ajralib turadi. O'zbek tili kam resursli til hisoblangaligi sababali hozirgacha bunday algoritmlar ishlab chiqilmagan. O'zbek tili matnlaridagi gaplarning undov belgilari to'g'ri yoki noto'g'ri qo'yilganligini aniqlash masalasini yechishning qoidaga asoslangan algoritmi ishlab chiqilgan va qoidalar bazasi yaratilgan.

ABSTRACT. In this work, we consider the development of a rule-based algorithm for analyzing exclamatory sentences in Uzbek texts. Exclamatory sentences are sentences that express the speaker's emotions (surprise, joy, anger, pity, appeal, etc.) and are often distinguished by exclamatory words, emotional tone, or special word order. Due to the fact that the Uzbek language is considered a low-resource language, such algorithms have not been developed so far. A rule-based algorithm for solving the problem of determining whether exclamation marks are placed correctly or incorrectly in sentences in Uzbek texts has been developed and a rule base has been created.

АННОТАЦИЯ. В данной работе рассматривается разработка алгоритма анализа восклицательных предложений в узбекских текстах на основе правил. Восклицательные предложения — это предложения, которые выражают чувства говорящего (удивление, радость, гнев, жалость, восклицание и т. д.) и часто характеризуются восклицательными словами, эмоциональным тоном или особым порядком слов. В связи с тем, что узбекский язык считается языком с ограниченными ресурсами, такие алгоритмы пока не разработаны. Разработан алгоритм на основе правил и создана база правил для решения

задачи определения правильности или неправильности расстановки восклицательных знаков в предложениях узбекских текстов.

Kalit soʻzlar. *Tinish belgilari, NLP, undov belgisi, undov gaplar.*

KIRISH

Oʻzbek tili matnlarida undov belgisining xato yoki toʻgʻri qoʻyilganligini aniqlash (punktatsion tahlil) masalasi tabiiy tilni qayta ishlash (NLP) masalalaridan biri hisoblanadi. Punktatsion tahlil tabiiy tilni qayta ishlashda mashinali tarjima, matn tahlili, semantik tahlil, sintaktik tahlil kabi masalalarni hal qilishda ishlatiladi.

Punktatsiya qoidolari va meʼyorlarini bilish esa yozma fikrni aniq va toʻgʻri, mantiqli va ifodali, ravon va tushunarli hamda maqsadga muvofiq bayon qilishda muhimdir. Tinish belgilarining asosiy vazifasi yozuvda ifodalanadigan nutqning fikriy boʻlinishini, gapning qanday qismlarga ajralishini va u boʻlaklarning oʻzaro munosabatlarini koʻrsatishdan iboratdir. Tinish belgilari gap mazmuni, tuzilishi va ohangiga koʻra qoʻllaniladi. Globallashuv davrida axborot texnologiyalari kun sayin hayotimizdan chuqur oʻrin egallamoqda. Bugun yozma nutq deyarli axborot texnologiyalari orqali amalga oshmoqda. Biron matn yoki xabarni elektron holatda yozilar ekan, bevosita tinish belgilariga ham murojaat qilinadi. Undov belgisi oʻzbek yozuvlarida XX asrdan koʻpchilik tinish belgilari kabioha davrning yagona adabiy til meʼyorlariga amal qilingan “Turkiston viloyatining gazetasi”da muntazam qoʻllangan.[1].

METODOLOGIYASI

Tarkibida emotsionallikni kuchaytiradigan qanday, qanchalik, qanchalar, qanaqa, naqadar, muncha, bunchalik kabi soʻzlar mavjud boʻlgan gaplar oxirida undov belgisi qoʻyiladi.

Yashash qanday yaxshi, tagʻin shunday katta, goʻzal shaharda!

Emotsionallik darajasi kuchli boʻlgan ritorik soʻroq gaplar oxirida baʼzan undov belgisi qoʻyiladi.

Ona oʻz bolasiga yomonlikni ravo koʻradimi!

1. Undov gapning mazmuni (semantik belgi)

- Gapda **gaplovchining his-tuygʻusi, emotsiyasi yoki kayfiyati** ifodalanadi: hayrat, quvonch, gʻazab, chaqiriq, ogʻriq, achinish va hokazo.

Misollar:

- Voy, qanday goʻzal manzara!
- Oh, belim!

2. Undov soʻzlarning ishtiroki (leksik belgi)

Undov gaplarda koʻpincha maxsus **undov soʻzlar** ishlatiladi:

Kategoriya	Namuna undov soʻzlar
Hayrat	Voy, opa-ey, eh-ha, hay-hay
Ogʻriq	Oh, oʻh
Achinish	Bechora, afsus, voy-boʻldim
Chaqiriq	Qani, yur, ana, mana

Qoidasi: Gapda ushbu undov soʻzlardan biri qatnashsa, bu gap undov gap boʻlishi ehtimoli yuqori.

3. Gap oxiridagi punktuatsiya belgisi (grafik belgi)

Undov gaplar odatda quyidagi belgilar bilan tugaydi:

- ! — Undov belgisi
- ?! — Hayrat, soʻroq va undovni birlashtirgan belgilar

Qoidasi: Gap ! yoki ?! bilan tugasa, bu undov gap boʻlish ehtimoli yuqori.

4. Gap tuzilmasi (sintaktik belgi)

- Undov gaplar koʻpincha **egasiz yoki iboraviy** boʻladi:
 - "Voy!" – egasiz
 - "Qani yur!" – buyruq + chaqiriq
 - "Bechora bola!" – iboraviy, emotsional

Qoidasi: Gapda ot+ot, undov+ot, yoki buyruq shaklida tuzilgan qisqa gaplar boʻlsa — undov gap boʻlishi mumkin.

5. Gapning ohangidan foydalanish (nutq belgi)

Yozma matnda toʻgʻridan-toʻgʻri ohangni aniqlash qiyin, lekin:

- ! belgisi
- undov soʻzlar
- qisqa va emotsional tuzilma

— bu belgilar **emotsional ohang** borligini koʻrsatadi.

6. Chaqiriq va buyruqni ifodalovchi gaplar

- Koʻpincha harakatga daʼvat yoki tezkor amal ifodalangan:
 - "Qani boshladik!"
 - "Yur, ketdik!"

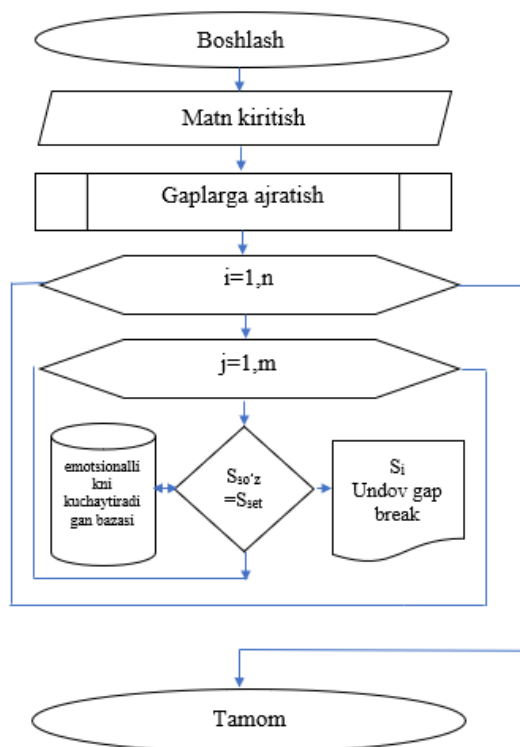
Qoidasi: Gap chaqiriq yoki topshiriq boʻlsa, lekin emotsional tusda boʻlsa — undov gap hisoblanadi.

UNDOV GAPLAR TAHLILINING ALGORITMI

Ushbu maqolada matndagi undov gaplarni tahlil qilish algoritmi ishlab chiqildi. Aloritm matndagi undov belgisini toʻgʻri qoʻyilgan yoki notoʻgʻri qoʻyilganligini tahlil qiladi. Quyida algortmning ishlash jarayonini koʻib chiqamiz.

1. Dasturni boshlash.
2. Dasturga matnni kiritish.

3. Tokenizatsiya jarayoni.
4. Ikkita sikl ochiladi. Dastlab sikl gaplarning sonigacha keying sikl gapdagi so'zlar sonigacha, Xar bir so'zni emotsionallikni kuchaytiradigan so'zlar bor yo'qligi bilan tekshiradi. Agar shunday so'z bo'lsa ekranga undov gap deb chiqaradi.
5. Dasturni yakunlash.



1-rasm: O'zbek tili matnlarida undov gaplarni aniqlash algoritmi.

NATIJA

Dastur 30 turdagi matn, 137 848 gap, 1 642 860 so'zdan iborat Uzbek korpusidagi matnlar bilan sinovdan o'tkazildi. Korpusda 3492 ta undov gap mavjud. Dastur 3492 ta undov gapni topdi va 111 tasini undov gap emas deb topdi. Jami 133 945 undov emas gapning 300 tasini undov gap deb topdi. Bu 74.3% aniqlikni beradi.

№	Matn turlari	Gaplar soni	Undov gap		Undov gap emas		F1 Score
			undov gap	undov gap emas	undov gap	undov gap emas	
	Adabiyot	11659	1143	11	14	10491	98,9
	Anatomiya	10565	1	1	1	10562	50,0
	Biologiya	4757	0	0	3	4754	0,0



	Botonika	2717	2	1	1	2713	66,7
	Diniy	6079	20	2	10	6047	76,9
	Fizika	12078	6	1	6	12065	63,2
	Geografiya	6237	0	0	19	6218	0,0
	Huquq	1964	5	1	2	1956	76,9
	Ijtimoiy soha	421	0	0	4	417	0,0
	Informatika	2512	22	3	12	2475	74,6
	Iqtisodiyot	520	1	0	1	518	66,7
	Kimyo	3566	18	2	14	3532	69,2
	Kino	668	3	1	1	663	75,0
	Matematika	7921	22	3	12	7884	74,6
	O'qish kitobi	3382	193	12	34	3143	89,4
	O'zbek tili kitob	11020	56	9	26	10929	76,2
	Qishloq xo'jaligi	1209	0	0	12	1197	0,0
	Qissa	10529	1729	32	26	8742	98,4
	Qonun va qarorlar	4360	1	0	1	4358	66,7
	Rassomchilik	3673	12	1	12	3648	64,9
	Roman	1896	80	8	19	1789	85,6
	San'at	9570	129	12	23	9406	88,1
	Siyosat	1222	7	1	2	1212	82,1
	Sport	2565	18	5	5	2537	78,3
	Tabobat	421	1	1	1	418	51,3
	Tarix	4619	5	1	2	4611	76,9
	Texnologiya	796	7	2	2	785	77,8
	Tibbiyot	5490	0	0	23	5467	0,0
	Urush	1084	10	1	11	1062	62,8
	Zoologiya	4348	1	0	1	4346	66,7
	jami	137848	3492	111	300	133945	74,3

XULOSA

Hozirgi kungacha o'zbek tili matnlari uchun undov belgilarini tahlil qilish agoritmlari deyarli ishlab chiqilmaganligi sababli ushbu ishda taklif qilingan algoritmlar dolzarb hisoblanadi. O'zbek tilida tabiiy tilni qayta ishlash masalasi bo'lgan punktuatsion tahlil (undov belgilari) qilishishni qoidaga asoslangan algoritmlari ishlab chiqildi, punktuatsion tahlil (undov belgilari) uchun qoidalar bazasi yaratildi. Ishlab chiqilgan algoritmlar yordamida o'zbek tilining 30 ta kategoriyadan iborat 1 642 860 ta so'z va 137 848 ta gapdan iborat korpus tahlil qilinib natijalar olindi. Ushbu ishning davomi sifatida keljakda o'zbek tili matnlarini punktuatsion (undov gaplarni) tahlil qilish masalasini sun'iy intellekt modellaridan foydalanib yechish va yaxshiroq aniqlikdagi natijalarni olish maqsad qilingan.

Foyadalanilgan adabiyotlar:

1. M. A., M. N., A.D. Mahmudov N., O'zbek tili me'yorlari (Punktuatsiya), vol. 1. Tashkent: Zamin nashr, 2021.
2. J. Pokrywka, “Punctuation Prediction for Polish Texts using Transformers,” Feb. 2024. doi: 10.48550/arXiv.2410.04621.
3. P. Żelasko, P. Szymański, J. Mizgajski, A. Szymczak, Y. Carmiel, and N. Dehak, “Punctuation Prediction Model for Conversational Speech,” Feb. 2018, pp. 2633–2637. doi: 10.21437/Interspeech.2018-1096.
4. M.S. Sharipov, H. S. Adinaev, and E. R. Kuriyozov, “Rule-Based Punctuation Algorithm for the Uzbek Language,” in International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices, EDM, 2024, pp. 2410 – 2414. doi: 10.1109/EDM61683.2024.10615061.
5. M. Bayraktar, B. Say, and V. Akman, “An analysis of English punctuation: the special case of comma,” International Journal of Corpus Linguistics, vol. 3, Jul. 1998, doi: 10.1075/ijcl.3.1.03bay.
6. D.Hardt, “Comma checking in Danish,” 2001.
7. U.Salaev, E.Kuriyozov, and C. Gomez-Rodríguez, “SimRelUz: Similarity and Relatedness scores as a Semantic Evaluation Dataset for Uzbek Language,” in 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages, SIGUL 2022 - held in conjunction with the International Conference on Language Resources and Evaluation, LREC 2022 - Proceedings, 2022, pp. 199 – 206. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85138700420&partnerID=40&md5=bf476cd74317f06577dd0548c5c600d6>
8. A. M.Abdurashetona and I. O. Ismailovich, “Methods of Tagging Part of Speech of Uzbek Language,” in Proceedings - 6th International Conference on Computer Science and Engineering, UBMK 2021, 2021, pp. 82 – 85. doi: 10.1109/UBMK52708.2021.9558900.
9. A. M.Abdurashetona and U. Mokhiyakon, “Software Features and Linguistic Features of Uzbek Synonymizer,” in Proceedings - 7th International Conference on Computer Science and Engineering, UBMK 2022, 2022, pp. 171 – 175. doi: 10.1109/UBMK55850.2022.9919447.
10. B. Mengliyev, Sh.Shahabitdinova, Sh.Khamroeva, Sh.Gulyamova, and A.Botirova, “The morphological analysis and synthesis of word forms in the linguistic analyzer,” Journal of Language and Linguistic Studies, vol. 17, no. 1, pp. 558 – 564, 2021, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103797274&partnerID=40&md5=8a4052419f721c3c734f8fe1c48984ec>
11. K. Madatov, S. Bekchanov, and J. Vicic, “Dataset of Karakalpak language stop words,” Data Brief, vol. 48, 2023, doi: 10.1016/j.dib.2023.109111.



12. M. Sharipov, J. Mattiev, J. Sobirov, and R. Baltayev, “Creating a Morphological and Syntactic Tagged Corpus for the Uzbek Language”, CEUR Workshop Proceedings, vol. 3315, pp. 93–98, 2022.
13. D. Mengliev, E. Akhmedov, V. Barakhnin, Z. Hakimov, and O. Alloyorov, “Utilizing Lexicographic Resources for Sentiment Classification in Uzbek Language,” Jan. 2023, pp. 1720–1724. doi: 10.1109/APEIE59731.2023.10347765.
14. K. Madatov, S. Bekchanov, and J. Vicic, “Dataset of stopwords extracted from Uzbek texts”, Data in Brief, vol. 43, 2022.