



O'ZBEK TILINING TABIIY TILNI QAYTA ISHLASH MODELLARI VA RIVOJLANISH ISTIQBOLLARI

Olimov Muslimbek Ulug'bek o'g'li

Andijon davlat universiteti Axborot texnologiyalari kafedrası o'qituvchisi

ochilov92@list.ru

Annotatsiya

Ushbu maqola O'zbek tilining tabiiy tilni qayta ishlash (NLP) sohasidagi hozirgi holati, mavjud modellar va kelajak istiqbollarini ko'rib chiqadi. Tadqiqotda Hugging Face, Google va OpenAI kabi platformalardagi o'zbekcha modellar, mahalliy ishlanmalar va sohada faoliyat yurituvchi olimlarning hissalarini tahlil qilingan. Maqolada O'zbek tili uchun mukammal NLP ekosistemi yaratishning amaliy yo'llari va bosqichlari batafsil bayon etilgan.

Kalit so'zlar: tabiiy tilni qayta ishlash, o'zbek tili, NLP modellari, sun'iy intellekt, kompyuter lingvistikasi, BERT, transformer modellari

Kirish

Tabiiy tilni qayta ishlash (Natural Language Processing - NLP) zamonaviy sun'iy intellekt texnologiyalarining eng muhim sohalaridan biri hisoblanadi. O'zbek tilining NLP modellari mavjud bo'lib, ularning soni va sifati doimiy ravishda ortib bormoqda. Bugungi kunda o'zbek tili uchun ishlab chiqilayotgan modellar va loyihalar milliy tilimizning raqamli muhitdagi mavqeyini mustahkamlashda muhim rol o'ynamoqda.

O'zbek Tilining Asosiy NLP Modellari va Loyihalari

1. Hugging Face Modellari

O'zbek tiliga mo'ljallangan BERT, GPT va boshqa transformer modellarini Hugging Face platformasida topish mumkin. Eng mashhurlari "uzbert" va "uzbek-roberta" modellari bo'lib, o'zbekcha matnlarni tushunish va tahlil qilish uchun ishlatiladi. Ushbu modellar quyidagi vazifalarni bajarishi mumkin:

- Matn klassifikatsiyasi



- Hissiyot tahlili (sentiment analysis)
- Nom-shakhs tanib olish (Named Entity Recognition - NER)
- Matn xulosalash
- Matn generatsiyasi

2. Ko'p Tilli Katta Modellar

GPT-4, ChatGPT, Gemini kabi umumiy AI modellar o'zbek tilini yaxshi tushunadi va unga javob bera oladi. Google Translate, DeepL va boshqa xizmatlar ham ilg'or NLP texnologiyalari asosida ishlaydi. Ushbu modellar quyidagi imkoniyatlarni taqdim etadi:

- Ko'p tilli tarjima
- Matn generatsiyasi va tahrirlash
- Savol-javob tizimlari
- Matn tahlili va xulosalash
- Kod yozish va tushuntirish

3. O'zbekcha Morfologik Tahlilchilar va Lug'at Tizimlari

Mahalliy tadqiqotchilar tomonidan turli morfologik tahlil va tokenizatsiya modullari ishlab chiqilgan. Bu modullar quyidagi vazifalarni bajaradi:

- So'zlarni asosiy shaklga keltirish (lemmatization)
- Morfologik tahlil va so'z turkumlarini aniqlash
- Sintaktik parsing va gap tuzilishini tahlil qilish
- Semantik tahlil va ma'no aniqlash
- Korpus yaratish va lug'at tuzish

4. O'zbekcha Chatbot va Ovozli Tizimlar

Bir qator mahalliy IT kompaniyalar va startaplar o'zbek tilida chatbot va ovozli yordamchilarni yaratmoqda. Ushbu tizimlar quyidagi sohalarda qo'llanilmoqda:

- Mijozlar bilan muloqot va xizmat ko'rsatish
- Ma'lumot berish va yo'l-yo'riq ko'rsatish



- Buyruqlarni bajarish va vazifalarni avtomatlashtirish
- Dialog tizimlari va virtual assistentlar

Sohada Faoliyat Yurituvchi Olimlar va Tadqiqot Markazlari

O'zbekistonda tabiiy tilni qayta ishlash (NLP) sohasida bir qator olimlar va tadqiqotchilar faoliyat yuritib, o'zbek tili materiallari asosida tadqiqotlar olib bormoqdalar. Asosiy tadqiqot yo'nalishlari quyidagilarni o'z ichiga oladi:

Akademik tadqiqotchilar:

- Prof. H. Arziqulov - O'zbek tili korpuslari va kompyuter lingvistikasi
- Dots. S. Rizayev - Morfologik tahlil va avtomatik tarjima tizimlari
- Prof. S. Muhamedova - Leksikografiya va raqamli lug'atlar
- Dr. A. Polatov - Semantik ontologiya va ma'lumot olish
- Dr. N. Jo'rayeva - Matn korpuslari va lingvistik annotatsiya

Tadqiqot markazlari:

- Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti - Kompyuter lingvistikasi kafedrası
- Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti
- O'zbekiston Milliy universiteti qoshidagi Raqamli texnologiyalar instituti
- AKFA University - AI va Data Science markazi

O'zbek Tilining Mukammal NLP Modelini Ishlab Chiqish Strategiyasi

O'zbek tilining to'liq lug'atini yaratib, mukammal NLP modelini ishlab chiqish uchun quyidagi bosqichlarni amalga oshirish zarur:

1. Ma'lumot Yig'ish va Korpus Yaratish

Matn manbalarini to'plash:

- O'zbek tilidagi badiiy adabiyot, ilmiy maqolalar va bloglar
- Davlat hujjatlari, qonun-qoidalar va rasmiy matnlar
- OAV materiallari: gazeta, jurnal, televideniye



- Ijtimoiy tarmoqlar va onlayn platformalar
- Audio va video materiallarni matnga o'tkazish

Lug'at va lingvistik ma'lumotlar bazasi:

- So'zlarning grammatik kategoriyalari va ma'nolari
- So'z yasaliş qoidalari va morfologik paradigmalar
- Dialekt xususiyatlari va o'zgarishlar
- Etimologik ma'lumotlar va tarixiy rivojlanish
- Sinonimlar, antonimlar va frazeologik birikmalar

2. Ma'lumotni Tozalash va Standartlashtirish

Sifat nazorati:

- Takrorlanuvchi va noto'g'ri yozilgan so'zlarni filtrlash
- Imlo va grammatik xatolarni avtomatik tuzatish
- Ma'lumot manbalarining ishonchliligini tekshirish
- Copyright va etik masalalarni hisobga olish

Lingvistik qayta ishlash:

- Tokenizatsiya va so'zlarni ajratish
- Lemmatizatsiya (so'zlarni asosiy shaklga keltirish)
- POS (Part-of-Speech) belgilash
- Sintaktik parsing va bog'lanish tahlili
- Semantik rollarni aniqlash

3. Model Arxitekturalari va Treningi

Zamonaviy arxitekturalarni tanlash:

- Transformer modellari (BERT, RoBERTa, T5, GPT)
- Encoder-decoder arxitekturalari
- Multimodal modellar (matn, ovoz, tasvir)
- Retrieval-augmented generation (RAG) tizimlari

Ko'p bosqichli o'qitish strategiyasi:

- Pre-training: katta korpusda til modelini o'qitish



- Fine-tuning: maxsus vazifalar uchun sozlash
- Prompt engineering va in-context learning
- Reinforcement learning from human feedback (RLHF)

4. Baholash va Optimallashtirish

Sifat metrikalari:

- BLEU, ROUGE ko'rsatkichlari tarjima uchun
- F1-score, Precision, Recall klassifikatsiya uchun
- Perplexity til modellari uchun
- Human evaluation va expert judgment

Model optimizatsiyasi:

- Quantization va pruning texnikalari
- Knowledge distillation
- Multi-task learning
- Ensemble methods

5. Amaliy Tatbiq va Joriy Etish

Dasturiy ta'minot va xizmatlar:

- RESTful API va GraphQL interfeyslari
- Web va mobile ilovalar uchun SDK
- Cloud platformalarga deployment (AWS, Azure, GCP)
- Edge computing uchun optimizatsiya

Integratsiya imkoniyatlari:

- CRM va ERP tizimlar bilan integratsiya
- E-commerce platformalar uchun maxsus echimlar
- Ta'lim texnologiyalari va LMS tizimlari
- Tibbiy ma'lumotlar bilan ishlash tizimlari

Kelajak Istiqbollari va Rivojlanish Yo'nalishlari

O'zbek tilining NLP sohasidagi rivojlanishi quyidagi yo'nalishlarda davom etishi kutilmoqda:



Texnologik innovatsiyalar:

- **Multimodal AI:** Matn, ovoz, tasvir va videolarni birlashtiruvchi modellar
- **Generative AI:** O'zbek tilida kreativ kontent yaratuvchi modellar
- **Conversational AI:** Tabiiy dialog olib boruvchi chatbotlar
- **Specialized LLMs:** Soha-maxsus katta til modellari

Amaliy tatbiqotlar:

- **Ta'lim:** Adaptiv o'qitish tizimlari va virtual o'qituvchilar
- **Tibbiyot:** Tibbiy hujjatlarni tahlil qilish va diagnostika yordamchisi
- **Huquq:** Qonuniy hujjatlarni tahlil qilish va maslahat tizimlari
- **Biznes:** Mijozlar bilan ishlash va marketing avtomatizatsiyasi

Ilmiy tadqiqotlar:

- Cognitive linguistics va psycholinguistics tadqiqotlari
- Code-switching va ko'p tillilik masalalari
- Ethical AI va bias detection
- Explainable AI va interpretability

Xulosa

O'zbek tilining NLP modellari rivojlanishi milliy tilimizning raqamli kelajagini ta'minlash uchun strategik ahamiyatga ega. Ilmiy tadqiqotchilar, IT mutaxassislari, davlat tashkilotlari va xalqaro hamkorlarning birgalikdagi sa'y-harakatlari orqali o'zbek tili uchun zamonaviy va samarali NLP ekosistemi yaratish mumkin.

Bu jarayon nafaqat texnologik taraqqiyotni ta'minlaydi, balki milliy madaniyat va tilimizning saqlanib qolishi, kelajak avlodlarga yetkazilishi uchun ham muhim zamin yaratadi. O'zbek tilining raqamli mavjudligi uning global miqyosda tanilishi va rivojlanishiga katta hissa qo'shadi.



Foydalanilgan Adabiyotlar

1. Hugging Face Hub - O'zbek tiliga mo'ljallangan modellar kolleksiyasi. URL: <https://huggingface.co/models?language=uz>
2. OpenAI Platform Documentation - GPT modellari va API hujjatlari. URL: <https://platform.openai.com/docs>
3. Google AI Research - Neural Machine Translation va Multilingual Models. URL: <https://ai.google/research/>
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
5. Radford, A., Wu, J., Child, R., et al. (2019). Language models are unsupervised multitask learners. *OpenAI Technical Report*.
6. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
7. Rogers, A., Kovaleva, O., Rumshisky, A. (2020). A primer in neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57, 615-686.
8. Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.
9. Eisenstein, J. (2019). *Introduction to Natural Language Processing*. MIT Press.
10. Manning, C.D., Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.
11. Arziqulov, H. (2023). O'zbek tili korpuslari va kompyuter lingvistikasi metodlari. *O'zbekiston Milliy universiteti ilmiy axborotnomasi*, 2(1), 45-52.
12. Rizayev, S. (2022). Morfologik tahlil va o'zbek tilida NLP algoritmlari. *Kompyuter lingvistikasi jurnali*, 15(3), 78-89.



13. Muhamedova, S. (2023). Raqamli lug'atlar va leksikografik ma'lumotlar bazasi. *Tilshunoslik tadqiqotlari*, 4(2), 123-135.
14. O'zbekiston Respublikasi Prezidenti Farmoni - "Raqamli O'zbekiston-2030" strategiyasi to'g'risida. (2020). Toshkent.
15. UNESCO Atlas of the World's Languages in Danger - Language Vitality and Endangerment Framework. (2022).