

“Turk tilidagi so‘zlarning turkumlari bo‘yicha modellar yaratish va ularga
ishlov beruvchi protsessorni yaratish”

Atoyeva Nafosat Husniddin qizi
Özbekiston Milliy Universiteti

Annotatsiya:

Ushbu tezis turk tilidagi so‘zlarni avtomatik ravishda turkumlashtirish uchun mo‘ljallangan modellar va ularni qayta ishlovchi maxsus protsessorni yaratishga bag‘ishlangan. Tezisdagi turk tilining o‘ziga xos grammatik xususiyatlari hisobga olingan holda so‘zlarni morfologik, sintaktik va semantik jihatdan tahlil qilish uchun mashinaviy o‘rganish usullari, neyron tarmoqlari va lingvistik qoidalar asosida modellar ishlab chiqiladi. Shuningdek, yaratilgan modellar asosida real vaqt rejimida yuqori unumdorlikni ta‘minlaydigan apparat ta‘minoti – maxsus protsessor loyihalashtiriladi. Tezisning asosiy maqsadi tabiiy tilni qayta ishlash (NLP) sohasida turk tili uchun samarali va tezkor yechimlarni taklif qilishdir.

Kalit so‘zlar: Turk tili, so‘z turkumlari, avtomatik turkumlashtirish, tabiiy tilni qayta ishlash (NLP), mashinaviy o‘rganish, neyron tarmoqlari, protsessor arxitekturasi, apparat ta‘minoti, unumdorlik.

Kirish:

Tabiiy tilni qayta ishlash (NLP) sohasining rivojlanishi kompyuterlarga inson tilini tushunish va qayta ishlash imkoniyatini beradi. So‘z turkumlarini aniqlash (Part-of-Speech Tagging - POS Tagging) NLPning fundamental vazifalaridan biri bo‘lib, matni tahlil qilish, mashinaviy tarjima, axborotni qidirish va boshqa ko‘plab amaliy ilovalarda muhim rol o‘ynaydi.

Turk tili o‘ziga xos aglutinativ (qo‘shimchalar orqali so‘z yasash) tuzilishga ega bo‘lib, bu so‘zlarning morfologik jihatdan boy va murakkab bo‘lishiga olib keladi. Bu xususiyat turk tilidagi so‘z turkumlarini avtomatik aniqlashni boshqa tillarga nisbatan qiyinlashtiradi. Mavjud NLP tizimlari ko‘pincha turk tilining grammatik xususiyatlarini to‘liq hisobga olmasligi sababli yetarli darajada aniqlik ko‘rsatmaydi.

Shu sababli, turk tilidagi so‘zlarning turkumlarini aniqlash uchun maxsus modellar va ularni tezkor qayta ishlashga mo‘ljallangan protsessor yaratish dolzarb masala hisoblanadi. Ushbu tadqiqot turk tilining lingvistik xususiyatlarini chuqur

o‘rganish asosida samaraliroq modellar ishlab chiqish va ularni apparat darajasida optimallashtirish orqali NLP sohasida sezilarli yutuqlarga erishishga qaratilgan.

Asosiy qism: Tabiiy tilni qayta ishlash (NLP) sohasining jadal rivojlanishi: XXI asrda axborot texnologiyalarining rivojlanishi bilan inson va kompyuter o‘rtasidagi muloqotning ahamiyati ortib bormoqda. NLP texnologiyalari, jumladan so‘z turkumlarini aniqlash (POS tagging), mashinaviy tarjima, sentiment tahlili, savol-javob tizimlari kabi ko‘plab amaliy sohalarda keng qo‘llanilmoqda. Bu esa turli tillar, shu jumladan turk tili uchun ham samarali NLP yechimlarini yaratish zaruratini keltirib chiqaradi.

Turk tilining o‘ziga xos lingvistik xususiyatlari: Turk tili aglutinativ til bo‘lib, so‘zlarga ko‘plab qo‘shimchalar birikishi orqali grammatik ma‘nolar ifodalanadi. Bu xususiyat so‘zlarning morfologik tuzilishini murakkablashtiradi va so‘z turkumlarini avtomatik aniqlashni boshqa tillarga nisbatan qiyinlashtiradi. Mavjud umumiy maqsadli NLP vositalari ko‘pincha turk tilining bu o‘ziga xosliklarini to‘liq hisobga olmasligi sababli yetarli darajada yuqori aniqlikka erisha olmaydi. Shuning uchun turk tilining lingvistik xususiyatlariga moslashtirilgan maxsus modellar va algoritmlarni ishlab chiqish muhim ahamiyatga ega.

Tezkor va samarali NLP tizimlariga talabning ortishi: Zamonaviy ilovalarda, masalan real vaqt rejimida ishlaydigan tarjima tizimlari, ovozli yordamchilar va katta hajmdagi matnlarni tahlil qilishda yuqori unumdorlik va tezlik talab etiladi. Mavjud dasturiy ta‘minotga asoslangan NLP yechimlari bu talablarga har doim ham javob bera olmaydi. Maxsus apparat ta‘minoti – so‘z turkumlarini aniqlashga yo‘naltirilgan protsessor yaratish orqali hisoblash jarayonini tezlashtirish va energiya samaradorligini oshirish mumkin.

Turk tili uchun resurslarning cheklanganligi: Boshqa keng tarqalgan tillarga nisbatan turk tili uchun annotatsiyalangan korpuslar va lingvistik resurslar nisbatan kamroq. Bu esa yuqori aniqlikdagi NLP modellarini yaratishni qiyinlashtiradi. Shu sababli, mavjud resurslardan maksimal darajada foydalanish va yangi metodlarni ishlab chiqish dolzarb vazifa hisoblanadi.

Akademik va amaliy ahamiyat: Ushbu mavzu bo‘yicha olib borilgan tadqiqotlar nafaqat turk tilshunosligi va kompyuter lingvistikasi sohasiga nazariy hissa qo‘shadi, balki turk tilidagi NLP ilovalarini rivojlantirish uchun amaliy asos yaratadi. Yaratilgan modellar va protsessor turk tilida so‘zlashuvchi foydalanuvchilar uchun qulay va samarali texnologiyalarni yaratishga imkon beradi.

Turk tilining so‘z turkumlarining lingvistik tahlili:

Turk tilidagi asosiy va yordamchi so‘z turkumlarining grammatik xususiyatlari (ot, sifat, fe‘l, ravish, olmosh, ko‘makchi, bog‘lovchi, undov).

So‘z o‘zgartiruvchi va so‘z yasovchi qo‘shimchalarning so‘z turkumlariga ta’siri.

So‘z tartibi va sintaktik aloqalarning turkum aniqlashdagi roli.

Turk tilidagi so‘z turkumlarining universal tagsetlar (masalan, Universal Dependencies) bilan qiyosiy tahlili.

Turk tilidagi so‘z turkumlarini aniqlash uchun modellar yaratish:

Morfologik tahlilga asoslangan modellar:

Qoidalar asosidagi tizimlar (rule-based systems).

Finite-state automata (chekli avtomatlar) va ularning turk tili morfologiyasini modellashtirishdagi imkoniyatlari.

Leksikografik resurslar va ularning turkum aniqlashda qo‘llanilishi.

Statistik usullarga asoslangan modellar:

N-gram modellar.

Hidden Markov Models (yashirin Markov modellari).

Conditional Random Fields (shartli tasodifiy maydonlar).

Turk tilining morfologik boyligi statistik modellarga qanday ta’sir qilishi.

Mashinaviy o‘rganishga asoslangan modellar:

Support Vector Machines (tayanch vektor mashinalari).

Decision Trees (qaror daraxtlari) va Random Forests (tasodifiy o‘rmonlar).

Turk tili uchun moslashtirilgan feature injiniring (belgi yaratish) usullari.

Neyron tarmoqlariga asoslangan chuqur o‘rganish modellari:

Recurrent Neural Networks (qaytariluvchi neyron tarmoqlar) - LSTM va GRU.

Transformer arxitekturasi va uning turk tilidagi ketma-ketliklarni modellashtirishdagi afzalliklari.

So‘z embeddinglari (Word Embeddings) - Word2Vec, GloVe, FastText va ularning turk tili uchun o‘ziga xos xususiyatlari.

Til modellari (Language Models) - BERT, mBERT va ularning turk tilidagi so‘z turkumlarini aniqlashdagi roli.

Ko‘p tilli (multilingual) modellar va ularning turk tili uchun transfer o‘rganish imkoniyatlari.

Yaratilgan modellarni baholash va solishtirish:

Turk tilidagi maxsus korpuslar va ularning turkumlashtirish uchun ahamiyati.

Model sifatini baholash metrikalari (aniqlik, aniqlik, to‘liqlik, F1-o‘lchov).

Turli modellarning turk tilidagi so‘z turkumlarini aniqlashdagi natijalarini qiyosiy tahlil qilish.

Xatoliklar tahlili va ularning sabablarini aniqlash.

Turk tilidagi so‘z turkumlarini qayta ishlovchi protsessor arxitekturasini loyihalash:

Yaratilgan modellarning hisoblash talablarini aniqlash.

Parallel hisoblash imkoniyatlarini tahlil qilish.

Maxsus apparat arxitekturasining afzalliklari (tezlik, energiya samaradorligi).

Protsessorning asosiy bloklari: xotira boshqaruv bloki, hisoblash bloki, boshqaruv bloki.

Neyron tarmoqlarini tezlatish uchun optimallashtirilgan arxitekturalar (masalan, systolic array).

FPGA (Field-Programmable Gate Array) yoki ASIC (Application-Specific Integrated Circuit) yechimlarini ko‘rib chiqish.

Protsessorning quvvat sarfi va o‘lchamlarini hisoblash.

Protsessorning ishlashini simulyatsiya qilish va baholash:

Loyihalashtirilgan protsessor arxitekturasining ishlash modelini yaratish.

Turk tilidagi matnlar asosida simulyatsiya o‘tkazish.

Protsessorning tezligi, unumdorligi va resurslardan foydalanish samaradorligini baholash.

Mavjud umumiy maqsadli protsessorlar bilan qiyosiy tahlil qilish.

Muhokama:

Ushbu bo‘limda tezisning asosiy natijalari muhokama qilinadi. Yaratilgan modellar va protsessorning afzalliklari va kamchiliklari tahlil qilinadi. Turk tilining o‘ziga xos xususiyatlarini hisobga olgan holda erishilgan yutuqlar va mavjud muammolar ko‘rib chiqiladi. Shuningdek, tadqiqotning NLP sohasidagi boshqa ishlanmalar bilan aloqasi va kelajakdagi tadqiqotlar uchun yo‘nalishlar muhokama etiladi. Xususan:

- Turk tilidagi morfologik murakkablikni yengishda qanday yondashuvlar eng samarali bo‘ldi?
- Chuqur o‘rganish modellari statistik va qoidalar asosidagi modellarga nisbatan qanday ustunliklarga ega?

- Maxsus protsessor arxitekturasini yaratishning zarurati va uning potentsial afzalliklari nimada ko‘rinadi?
- Yaratilgan protsessorning real vaqt rejimida ishlash imkoniyatlari qanday baholanadi?
- Tadqiqot natijalarining turk tilidagi NLP ilovalari (masalan, mashinaviy tarjima, sentiment tahlili) uchun ahamiyati nimada?

Xulosa va takliflar:

Tezisning xulosa qismida tadqiqotning asosiy natijalari umumlashtiriladi va qo‘yilgan maqsadlarga erishilganligi ta’kidlanadi. Yaratilgan modellar va protsessorning turk tilidagi so‘z turkumlarini avtomatik aniqlash sohasiga qo‘shgan hissasi baholanadi. Kelgusida ushbu yo‘nalishda olib borilishi mumkin bo‘lgan tadqiqotlar va takliflar keltiriladi. Masalan:

Yaratilgan modellarning aniqligini oshirish uchun qo‘shimcha ma’lumotlardan foydalanish.

Protsessor arxitekturasini yanada optimallashtirish va energiya samaradorligini oshirish.

Yaratilgan tizimni boshqa NLP vazifalarida (masalan, nutqni tanish, matnni tahlil qilish) qo‘llash imkoniyatlarini o‘rganish.

Turk tili uchun katta hajmli annotatsiyalangan korpuslar yaratishning ahamiyati.

Kam resursli tillar uchun transfer o‘rganish usullarini turk tiliga tadbiq etish.

Foydalanilgan adabiyotlar

1. Çöltekin, Ç. (2010). A Freely Available Morphological Analyzer for Turkish. Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC).
2. Oflazer, K. (1994). Two-level description of Turkish morphology. Literary and Linguistic Computing, 9(2), 137–148.
3. Yıldız, O., Çetinoğlu, Ö., & Schulz, S. (2016). Morphological Disambiguation for Turkish: A Deep Learning Approach. Proceedings of COLING.
4. Jumaniyozov, I. (2022). Kompyuter lingvistikasi asoslari. Toshkent: Fan va Texnologiya nashriyoti.
5. Yusupov, B. (2018). Sun’iy intellekt va tabiiy tilni qayta ishlash texnologiyalari. Toshkent: Innovatsiya.