O‘ZBEK TILIDAGI MATNLARDA KALIT SO‘ZLARNI AVTOMATIK ANIQLASHda AKT TEXNOLOGIYALARINING O‘RNI VA USULLARI

Vaisova Muxayyo Akbarovna
Urganch davlat universiteti
Kompyuter lingvistikasi yo’nalishi
2-kurs magistranti

Annotatsiya

Mazkur ilmiy maqolada o‘zbek tilidagi matnlarda kalit so‘zlarni avtomatik aniqlash jarayonida axborot-kommunikatsiya texnologiyalarining (AKT) roli va zamonaviy dasturiy yondashuvlar tahlil qilinadi. Tadqiqot doirasida TF-IDF, TextRank algoritmlari, mashinaviy o‘rganish modellaridan foydalanish hamda Python dasturlash muhiti asosida kalit so‘zlarni ajratib olish imkoniyatlari yoritiladi. Shuningdek, o‘zbek tili uchun NLP resurslarini rivojlantirish istiqbollari ko‘rsatib beriladi.

Kalit so‘zlar: AKT, kalit so‘z aniqlash, NLP, TF-IDF, TextRank, Python, UzbekBERT.

Kirish

Raqamli transformatsiya sharoitida matnli ma’lumotlar hajmining ortishi AKT asosida avtomatik matn tahlili texnologiyalarini jadal rivojlantirmoqda. Xususan, elektron hujjatlar, ilmiy maqolalar, internet resurslari va ta’lim platformalarida matnlarni qayta ishlashda kalit so‘zlarni aniqlash muhim vazifalardan biri hisoblanadi. AKT vositalari orqali kalit so‘zlarni avtomatik ajratish quyidagi jarayonlarda qo‘llaniladi:

• elektron kutubxonalar va kataloglash tizimlarida,
• onlayn ta’lim platformalarida mavzularni indekslashda,
• sun’iy intellekt asosidagi qidiruv tizimlarida,
• matnlarni avtomatik qisqartirish va annotatsiya yaratishda.

O‘zbek tili uchun bunday texnologiyalarni joriy etish dolzarb ilmiy-amaliy ahamiyatga ega.

1. Kalit so‘zlarni aniqlashda AKT asoslari

Kalit so‘zlarni avtomatik aniqlash — bu matndan mazmuniy jihatdan eng muhim so‘z va iboralarni ajratib olish jarayonidir. Ushbu jarayon AKT vositalarisiz, ya’ni dasturiy ta’minot va algoritmlarsiz amalga oshirilishi qiyin.

Bugungi kunda kalit so‘z ajratishda quyidagi AKT texnologiyalar qo‘llaniladi:

• NLP (Natural Language Processing) platformalari
• Python dasturlash muhiti
• Mashinaviy o‘rganish algoritmlari
• Sun’iy neyron tarmoqlar (Transformer modellar)
• Matn korpuslari va lingvistik ma’lumotlar bazasi

2. Statistik algoritmlar va AKT yechimlari

2.1. TF-IDF algoritmi va dasturiy qo‘llanishi

TF-IDF algoritmi AKT tizimlarida eng ko‘p qo‘llaniladigan statistik usul bo‘lib, so‘zning matndagi muhimlik darajasini aniqlaydi.

[TF-IDF = TF(t) × IDF(t)]

AKT misolida bu algoritm Python dasturida quyidagi kutubxonalar yordamida bajariladi:

• scikit-learn
• numpy
• pandas

Misol: o‘zbek matnidan avtomatik kalit so‘z ajratish jarayoni elektron hujjatlarni indekslash tizimlarida qo‘llanadi.

TF-IDF algoritmining afzalligi shundaki, u avtomatik ravishda matndagi terminlarni ajratib beradi. Ammo o‘zbek tilidagi qo‘shimchalar sababli lemmatizatsiya talab etiladi.

2.2. N-gram texnologiyasi

AKT amaliyotida kalit so‘zlar ko‘pincha birikma ko‘rinishida uchraydi:

• “raqamli ta’lim texnologiyalari”

• “sun’iy intellekt tizimlari”

• “matnlarni avtomatik tahlil qilish”

Shu sababli bigram va trigram modellar AKT tizimlarida keng qo‘llaniladi.

3. TextRank algoritmi va raqamli matn tahlili

TextRank algoritmi graf asosidagi avtomatik ajratish metodidir. Ushbu algoritm AKT tizimlarida quyidagi platformalarda ishlatiladi:

• Google Scholar metadata tahlilida

• onlayn referatlash dasturlarida

• sun’iy intellekt asosidagi qidiruv tizimlarida

TextRank jarayoni:

1. matndagi so‘zlar tugun sifatida olinadi,

2. ular o‘zaro bog‘lanish asosida graf hosil qiladi,

3. eng markaziy tugunlar kalit so‘z bo‘ladi.

AKT nuqtayi nazaridan TextRank algoritmi avtomatik va korpussiz ishlashi bilan qulay.

4. Mashinaviy o‘rganish va AKT asosidagi intellektual yondashuvlar

4.1. Klassik ML modellari

AKT asosida kalit so‘z aniqlashda quyidagi modellar qo‘llanadi:

• Support Vector Machine (SVM)

• Logistic Regression

• Random Forest

Bu modellar elektron hujjatlarni avtomatik tasniflashda muhim rol o‘ynaydi.

4.2. Transformer va UzbekBERT texnologiyasi

So‘nggi yillarda NLP sohasida transformer modellar katta burilish yasadi. O‘zbek tili uchun ishlab chiqilgan UzbekBERT modeli matn kontekstini tushunib, kalit so‘zlarni aniq ajratib beradi.

AKT misolida:

• chatbot tizimlari,

• intellektual qidiruv platformalari,

• avtomatik tarjima dasturlari

kalit so‘zlarni semantik aniqlash orqali ishlaydi.

5. O‘zbek tilida AKT asosida kalit so‘z ajratish muammolari

O‘zbek tili uchun AKT asosida kalit so‘z aniqlashda bir qancha muammolar mavjud:

• morfologik murakkablik (qo‘shimchalar tizimi),

• lemmatizatorlarning yetishmasligi,

• stop-so‘zlar bazasining cheklanganligi,

• katta annotatsiyalangan datasetlarning yo‘qligi.

Bu muammolarni hal qilish uchun milliy NLP platformalarini rivojlantirish talab etiladi.

Xulosa

O‘zbek tilidagi matnlarda kalit so‘zlarni avtomatik aniqlashda AKT texnologiyalarining o‘rni juda katta. Statistik algoritmlar (TF-IDF), graf asosidagi metodlar (TextRank) hamda mashinaviy o‘rganish va transformer modellar yordamida matn mazmunini avtomatik ajratib olish imkoniyatlari kengaymoqda. Kelajakda o‘zbek tili uchun maxsus korpuslar yaratish, UzbekBERT modellarini rivojlantirish va raqamli lingvistik resurslarni boyitish ushbu sohaning istiqbolli yo‘nalishidir.

Foydalanilgan adabiyotlar

1. Manning C., Raghavan P., Schütze H. Introduction to Information Retrieval. 2008.
2. Mihalcea R., Tarau P. TextRank Algorithm. EMNLP, 2004.
3. Jurafsky D., Martin J. Speech and Language Processing. 2021.
4. Devlin J. et al. BERT Model. 2019.

5. Karimov U. O‘zbek tilida NLP muammolari. Toshkent, 2022.
6. Abdurahmonov A. Kalit so‘z ajratish algoritmlari. Filologiya, 2023.
7. UzbekBERT loyihasi bo‘yicha ilmiy manbalar, 2024.