



PAREMIOLOGIK BIRLIKLARNI AVTOMATIK ANIQLASH TEXNOLOGIYASI

Muxlisa Pardaboyeva Zokir qizi

Jizzax viloyati Paxtakor tumani 11-maktab ona tili va adabiyot o'qituvchisi.

Annotatsiya: Ushbu maqolada paremiologik birliklarni (maqol, matal va hikmatli so'zlar) avtomatik aniqlash masalasi, uning lingvistik va kompyuter lingvistikasi nuqtai nazaridan yondashuvlari hamda zamonaviy sun'iy intellekt texnologiyalari asosida ishlab chiqilayotgan algoritmlar tahlil qilinadi. Shuningdek, tabiiy tilni qayta ishlash (NLP) usullari, mashinali o'rganish va chuqur o'rganish modellari yordamida paremiologik birliklarni aniqlashning samarali metodlari yoritiladi.

Kirish

Paremiologik birliklar xalq og'zaki ijodining eng muhim qatlamlaridan biri bo'lib, ular milliy tafakkur, madaniyat va qadriyatlarni ifodalaydi. Maqol va matallar qisqa, ammo mazmunan chuqur bo'lib, ko'pincha metaforik va ko'chma ma'noda ishlatiladi.

Bugungi raqamli davrda katta hajmdagi matnlarni avtomatik qayta ishlash zarurati ortib bormoqda. Shu sababli paremiologik birliklarni avtomatik aniqlash texnologiyasi tilshunoslik va sun'iy intellektning muhim yo'nalishlaridan biriga aylandi.

Paremiologik birliklarning lingvistik xususiyatlari

Paremiologik birliklar quyidagi asosiy belgilar bilan ajralib turadi:

Barqaror tuzilma – o'zgarmas yoki kam o'zgaruvchan sintaktik shakl

Ko'chma ma'no – literal emas, metaforik ifoda

Qisqalik va ixchamlik

Didaktik funksiyaga ega bo'lishi

Masalan:

“Yetti o'lchab, bir kes”



“Mehnat qilgan — to‘q yashar”

Ushbu xususiyatlar ularni avtomatik aniqlashni murakkablashtiradi, chunki oddiy grammatik yoki kalit so‘zlarga asoslangan yondashuv yetarli bo‘lmaydi.

Avtomatik aniqlash muammosi

Paremiologik birliklarni avtomatik aniqlash quyidagi muammolar bilan bog‘liq:

Polisemiya (ko‘p ma’nodlilik)

Bir maqol turli kontekstlarda turlicha ma’noda ishlatilishi mumkin.

Erkin iboralar bilan o‘xshashlik

Ko‘plab maqollar oddiy gap tuzilishiga o‘xshab ketadi.

Til resurslarining cheklanganligi

O‘zbek tilida to‘liq raqamlashtirilgan paremiologik korpuslar kam.

Texnologik yondashuvlar

1. Qoidaga asoslangan yondashuv

Bu usulda lingvistik qoidalar va oldindan tuzilgan lug‘atlar ishlatiladi.

Afzalliklari:

Aniqlik yuqori bo‘lishi mumkin

Tushuntirilishi oson

Kamchiliklari:

Yangi maqollarni aniqlay olmaydi

Kengaytirish qiyin

2. Statistik yondashuv

Bu usulda matndagi so‘z birikmalarining chastotasi va birga uchrash ehtimoli hisoblanadi.

Eng ko‘p ishlatiladigan usullar:

n-gram modellari

TF-IDF tahlili

kollokatsiya aniqlash

3. Mashinali o‘rganish (Machine Learning)



Bu yondashuvda modelga katta hajmdagi belgilangan (annotatsiyalangan) ma'lumotlar beriladi.

Qo'llaniladigan algoritmlar:

Support Vector Machine (SVM)

Random Forest

Naive Bayes

Xususiyatlar:

sintaktik belgilari

semantik embeddinglar

so'zlarning kontekst vektorlari

4. Chuqur o'rganish (Deep Learning)

Eng zamonaviy yondashuv bo'lib, u kontekstni chuqur tahlil qilish imkonini beradi.

Asosiy modellar:

LSTM (Long Short-Term Memory)

GRU

Transformer modellar (BERT, RoBERTa)

Bu modellar maqollarning kontekstga bog'liq ma'nosini aniqlashda juda samarali hisoblanadi.

Transformer asosidagi yondashuv

So'nggi yillarda Transformer arxitekturasi paremiologik birliklarni aniqlashda eng samarali vositalardan biri bo'lib kelmoqda.

Uning afzalliklari:

Kontekstni to'liq hisobga oladi

Uzoq bog'lanishlarni aniqlaydi

Ko'p tilli modellarga moslashuvchan

Masalan, BERT modeli matndagi gapning maqol ekanligini yuqori aniqlik bilan bashorat qila oladi.

Tizim arxitekturasi (umumiy model)



Paremiologik birliklarni aniqlash tizimi odatda quyidagi bosqichlardan iborat:

Matnni yig'ish
(ijtimoiy tarmoqlar, kitoblar, korpuslar)
Oldindan ishlov berish
tokenizatsiya
stop-so'zlarni olib tashlash
lemmatizatsiya
Xususiyat ajratish
n-gramlar
embeddinglar
Modelga kiritish
ML yoki DL algoritmlar
Natijani aniqlash
maqol / maqol emas klassifikatsiyasi
O'zbek tili uchun muammolar

O'zbek tilida paremiologik birliklarni avtomatik aniqlashda quyidagi muammolar mavjud:

Korpuslarning yetishmasligi
Dialektal farqlar
Imlo variantlari (masalan, "bo'ladi / boladi")
Raqamli resurslarning kamligi
Shu sababli lokal til resurslarini yaratish juda muhim.
Amaliy qo'llanilishi
Paremiologik birliklarni avtomatik aniqlash quyidagi sohalarda qo'llaniladi:
Ta'lim tizimi (til o'rgatish platformalari)
Sun'iy intellekt chat-botlar
Matn tahlili va sentiment analiz
Tarjima tizimlari
Raqamli lug'atlar yaratish



Xulosa

Paremiologik birliklarni avtomatik aniqlash texnologiyasi zamonaviy kompyuter lingvistikasi va sun'iy intellektning muhim yo'nalishlaridan biridir. Qoidaga asoslangan, statistik va chuqur o'rganish yondashuvlari birgalikda qo'llanganda yuqori aniqlikka erishish mumkin.

Kelajakda o'zbek tili uchun maxsus yirik paremiologik korpuslar va neyron tarmoqlarga asoslangan modellar ishlab chiqilishi ushbu sohaning yanada rivojlanishiga xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR:

Mamatov, A. E. O'zbek tili frazeologiyasi va paremiologiyasi. Toshkent: Fan nashriyoti, 2011.

Sayfullayeva Roza va boshqalar. Hozirgi o'zbek adabiy tili. Toshkent: Universitet, 2010.