

TÁBIYIY HÁM TEXNIKALÍQ ILIMLER

Fizika. Matematika. Texnika. Informatika.

QARAQALPAQ TILIN AVTOMATIK TÜRDE ANALIZLEWDE MASHINALÍQ OQÍTÍW ALGORITMLERIN QOLLANÍLÍW

G.R.Abdaliev – docent

N.U.Uteuliev – *fizika-matematika ilimleri doktori, professor*

B.K.Kalmuratov – *áğa oqıtıwshı*

Tashkent informaciyalıq texnologiya universiteti Nókis filiali

Tayanch so'zlar: qoraqalpoq tili, avtomatik analiz, mashinalı o'qıtish, tabiiy tilni qayta ishlash, morfologik tahlil, sintaktik tahlil, semantik tahlil, lemmatizatsiya, tokenizatsiya, texnologik rivojlanish, til korpusi, Python dasturlash, NLTK, scikit-learn, tasniflash algoritmlari, til resurslari.

Ключевые слова: каракалпакский язык, автоматический анализ, машинное обучение, обработка естественного языка, морфологический анализ, синтаксический анализ, семантический анализ, лемматизация, токенизация, технологическое развитие, корпус языка, программирование на Python, NLTK, scikit-learn, алгоритмы классификации, языковые ресурсы.

Key words: karakalpak language, automatic analysis, machine learning, natural language processing, morphological analysis, syntactic analysis, semantic analysis, lemmatization, tokenization, technological development, language corpus, Python programming, NLTK, scikit-learn, classification algorithms, language resources.

Házirgi zaman texnologiyalıq rawajlanıw dáwirinde til texnologiyaları tez pát penen rawajlanbaqta. Mashinalıq oqıtıw algoritmleri til analizi tarawında úlken áhmiyetke iye bolıp, hár túrli tillerdi izertlewde keń qollanımaqta. Usı maqalada mashinalıq oqıtıw algoritmleriniń qaraqalpaq tilin avtomatik türde analizlewde qollanıwı haqqında sóz etiledi.

Mashinalıq oqıtıw - kompyuterlerge anıq program-malastırılmaстан, tájiriybe tiykarında úyreniw imkaniyatın beriwshi jasalma intellektiń bir tarmağı bolıp tabıladı.

Tildi avtomatik analizlew - kompyuter járdeminde tildiń hár túrli aspektlerin (morfologiya, sintaksis, semantika) avtomatik türde izertlew processı.

Algoritm - belgili bir mäseleni sheshiw ushın orınlanıwı kerek bolğan ámellerdiń izbe-iz orınlanıwı.

Bul izertlewdiń tiykarǵı maqseti - mashinalıq oqıtıw algoritmlerin qollanıw, qaraqalpaq tilin avtomatik türde analizlew usılların islep shıǵıw hám onıń nátiyjeligin bahalaw [1].

Qaraqalpaq tiliniń avtomatik analizi házirgi kúnge shekem tolıq izertlenbegen taraw bolıp esaplanadı. Mashinalıq oqıtıw algoritmlerin qollanıw arqalı bul tildiń ózgesheliklerin tereń úyreniw hám onı cıfrlı texnologiyalar menen integraciyalaw múmkinshiligi payda boladı.

Bul izertlewdiń áhmiyeti tómendegilerden ibarat:

1. Qaraqalpaq tiliniń grammatikalıq qurılısın avtomatik türde analizlew;

2. Til korpusın dúziwde járdem beriw;

3. Mashina awdarması sistemaların jetilistiriw;

4. Tildi úyreniw hám oqıtıw processlerin avtomatlastiriw;

5. Qaraqalpaq tilindegi tekstlerdi avtomatik klassifikaciyalaw hám analizlew;

Python programmalastırıw tili mashinalıq oqıtıw tarawında keń qollanıladı. Onıń kóp sanlı kitapxanaları hám qolaylı sintaksisi mashinalıq oqıtıw algoritmlerin ámelge asırıwda úlken jeńillik tuwdıradı. Qaraqalpaq tilin analizlewde tómendegi Python kitapxanaları hám algoritmleri qollanıladı:

1. NLTK (Natural Language Toolkit): Bul kitapxana tildi qayta islew ushın kerekli ásbaplar menen támiyinleydi. Misalı, sózlerdi morfemalargá ajratıw, sóz túrlerin anıqlaw hám sintaksislik analiz júrgiziw ushın qollanıladı [2].

```
import nltk
nltk.download('punkt')
```

```
from nltk.tokenize import word_tokenize
text = "Men qaraqalpaq tilin úyrenemen."
tokens = word_tokenize(text)
print(tokens)
```

2. scikit-learn: Bul kitapxana klassifikaciya, regressiya hám klasterizaciya algoritmlerin óz ishine aladı. Qaraqalpaq tilindegi tekstlerdi klassifikaciyalaw ushın qollanıwı múmkin.

```
from sklearn.feature_extraction.text import
CountVectorizer
from sklearn.naive_bayes import MultinomialNB
# Mısal ushın tekstler hám olardıń kategoriyaları
texts = ["Bul jaqsı kitap", "Hawa rayı jaqsı", "Kitap
oqıw paydalı"]
categories = ["kitap", "hawa rayı", "kitap"]
vectorizer = CountVectorizer()
X = vectorizer.fit_transform(texts)
clf = MultinomialNB()
clf.fit(X, categories)
# Jańa tekstti klassifikaciyalaw
new_text = ["Búgin kún ıssı"]
X_new = vectorizer.transform(new_text)
predicted = clf.predict(X_new)
print(predicted)
```

3. Gensim: Bul kitapxana tema modellestiriw hám sóz vektorların dúziw ushın qollanıladı. Qaraqalpaq tilindegi tekstlerdiń semantikalıq analizi ushın paydalanıwǵa boladı.

```
from gensim.models import Word2Vec
# Mısal ushın sóylemler
sentences = [["men", "mektepke", "baraman"], ["ol",
"kitap", "oqıydı"]]
# Word2Vec modeli dúziw
model = Word2Vec(sentences, min_count=1)
# Sózlerdiń uqsaslıǵın tabıw
similarity = model.wv.similarity('mektepke', 'kitap')
print(f'"mektepke" hám "kitap" sózleriniń uqsaslıǵı:
{similarity}')
```

Qaraqalpaq tilin avtomatik türde analizlew processı bir neshe basqıshlardan ibarat boladı. Olardan atap ótsek: Teksti normalizaciyalaw basqıshında tekst birdey qalıpke keltiriledi, yaǵnıy úlken háripler kishi háriplerge ózgeriledi, punktuaciya belgileri alıp taslanadı hám t.b. Al, tokenizaciya basqıshında tekst sózlerge yamasa sóz dizbeklerine bólinedi. Soń "Stop-sózlerdi alıp taslaw" basqıshı ámelge asırılıp, ol tildiń mánisine úlken tásir

etpeytuǵın jiýi ushırasatuǵın sózler (mısalı, "hám", "menen", "biraq") alıp taslanadı. Onnan keyin, Stemming yamasa lemmatizaciya ámelge asırıladı. Onda sózlerdiń túbiri yamasa sózdiktegi tiykarǵı forması anıqlanadı. Keyingi basqısh sóz túrlerin anıqlaw bolıp tabıladı. Hár bir sózdiń grammatikalıq kategoriyası (atlıq, feyil, kelbetlik h.t.b.) belgilenedi. Al sintaksislik analizde gáptiń qurılısı analizlenedi, sózler arasındaǵı baylanıslar anıqlanadı. Semantikalıq analiz bolsa, tekstiń mánisi hám konteksti túsindiriledi.

Bul processlerdi ámelge asırıw ushin mashinalıq oqıtıw algoritmleri qollanıladı. Mısalı:

- Naıv Bayes klassifikatori: Tekstlerdi kategoriyalarǵa bóliw ushin qollanıladı.

- Qollap-quwatlawshı vektor mashinaları (SVM): Tekstlerdi klassifikaciyalaw hám regressiya máselelerin sheshiw ushin paydalanıladı.

- Rekurrent neyron tarmaqları (RNN): Til modellew hám mashina awdarması ushin qollanıladı.

- Transformerler (mısalı, BERT): Kontekстке baylanıslı sóz vektorların dúziw hám hár túrli til máselelerin sheshiw ushin paydalanıladı [3].

Qaraqalpaq tili ushin bul algoritmlerdi qollanıwda geypara qıynshılıqlar bolıwı múmkin: Birinshiden resurslardıń jetispewshiligi yaǵnıy, qaraqalpaq tilinde úlken kólemdegi annotaciyalanǵan korpuslardıń joqlıǵı. Ekinshiden morfologiyalıq quramalılıq. Bunda qaraqalpaq tiliniń agglyutinativ tábiyatı mashinalıq oqıtıw modellerin úyretiwdi qıynlastıradı. Al, standartlastırılǵan resurslardıń jetispewshiligi bolsa, qaraqalpaq tili ushin

standartlastırılǵan sózlikler hám grammatikalıq qaǵıydalar bazasınıń tolıq emesligin kórsetedi. Bul qıynshılıqlardı jeńiw ushin tómendegi usınıslar beriledi:

1. Korpus dúziw: Qaraqalpaq tilinde úlken kólemdegi tekstlerdi jıynaw hám olardı annotaciyalaw. Bul process avtomatlastırılǵan hám qoldan islew usıllarınıń kombinaciyasın talap etedi;

2. Morfologiyalıq analizatorlar dúziw: Qaraqalpaq tiliniń morfologiyalıq ózgesheliklerin esapqa alatuǵın arnawlı algoritmler islep shıǵıw;

3. Til resursların standartlastırıw: Qaraqalpaq tili ushin birdey bolǵan sózlikler, grammatikalıq qaǵıydalar bazasın dúziw;

4. Kóp tilli modellerdi qollanıw: Tuwısqan túrkiy tiller ushin islengen modellerdi qaraqalpaq tiline beyimlestiriw;

5. Transfer oqıtıwdı qollanıw: Basqa tiller ushin úyretilgen modellerdi qaraqalpaq tiline kóshiriw hám jergilikli maǵlıwmatlar menen qayta úyretiw;

Juwmaqlap aytqanda, qaraqalpaq tilin avtomatik analizlewde mashinalıq oqıtıw algoritmlerin qollanıw úlken potencialǵa iye. Bul usıllar tildi izertlew, saqlaw hám rawajlandırıwda áhmiyetli orın tutadı. Biraq, bul baǵdarda aldımızda ele kóp jumıslar bar. Qaraqalpaq tili ushin sapalı resurslar dúziw, algoritmlerdi beyimlestiriw hám jergilikli ózgesheliklerdi esapqa alıw zárúr.

Keleshekte, qaraqalpaq tilin avtomatik analizlew texnologiyaların rawajlandırıw arqalı biz tilimizdi saqlaw, onı úyreniw processin jeńillestiriw hám xalıqaralıq dárejede tanıtıw imkaniyatına iye bolamız. Bul tarawdaǵı izertlewler hám innovaciya qaraqalpaq tiliniń cıfrlı dáwirde de óz ornın saqlap qalıwına járdem beredi.

Ádebiyatlar

1. Jurafsky D., & Martin, J.H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson. 2020.

2. Bird S., Klein E., & Loper, E. Natural Language Processing with Python. O'Reilly Media. 2009.

3. Goldberg Y. Neural Network Methods for Natural Language Processing. Morgan & Claypool Publishers. 2017.

РЕЗЮМЕ. Mazkur maqola Qaraqalpaq tilini avtomatik tahlil qilishda algoritmlarining qo'llanilishi masalasi o'rganildi. Unda morfologiya, sintaksis va semantika kabi til jihatlarini avtomatik tahlil qilish jarayonlari ko'rsatilib, algoritmlarning texnik imkoniyatlari tahlil qilingan. Python dasturlash tili va uning kutubxonalari asosida til korpuslarini yaratish hamda ularni tahlil qilish imkoniyatlari bayon etilgan.

РЕЗЮМЕ. В статье рассматривается использование алгоритмов машинного обучения для автоматического анализа каракалпакского языка. Описаны процессы автоматического анализа языка, включая морфологию, синтаксис и семантику, а также технические возможности алгоритмов. Рассматривается создание и анализ языковых корпусов с использованием программирования на Python и его библиотек.

SUMMARY. This article explores the application of machine learning algorithms for the automatic analysis of the Karakalpak language. It describes the processes of automated analysis, including morphology, syntax, and semantics, and examines the technical capabilities of the algorithms. The creation and analysis of language corpora using Python programming and its libraries are highlighted.

КАРБИДКРЕМНИЕВЫЙ ДЕТЕКТОРНЫЙ ДИОД Au-TiB_x(ZrB_x)-n-SiC 6H

С.К.Абдижалиев – старший преподаватель

Нукусский государственный педагогический институт имени Ажинияза

Таянч сўзлар: Шотки тўсиғи, қувват сенсори, кремний карбиди, детектор диодлари.

Ключевые слова: барьер Шоттки, датчик мощности, карбид кремния, детекторные диоды.

Key words: Schottky barrier, power sensor, silicon carbide, detector diodes.

Введение. Анализ результатов работ последних лет по SiC подтверждает его перспективность для создания электронных устройств, способных функционировать в экстремальных условиях. Представляется возможность определить более конкретные направления перспективных разработок. Наиболее полно его достоинства могут быть реализованы в микроволновых приборах с небольшой рабочей площадью

(эффективные диаметры рабочих областей до 100-300мкм), которые в процессе работы могут иметь выделение чрезвычайно высоких уровней удельной мощности. Этим определяется требование иметь по возможности более высокую рабочую температуру рабочей области чипа и минимальные тепловые сопротивления чипа, а также приемлемую стабильность параметров контактных систем при длительном