ISNN 3030-3907

2026
MAY
ISSUE - 5
VOLUME - 2

DEVELOPMENT OF SCIENCE

scientific journal



Development of science

Ilmiy jurnal

MAY 2026
Issue - 5, VOLUME - 2

ISSN 3030 -3907

Ilmiy jurnal OAK rayosatining 2024 yil 27 sentabrdagi 361-son qaroriga asosan 02.00.00 kimyo fanlari hamda 2025 yil 12 fevral 367- son qaroriga asosan 05.00.00 texnika fanlari bo‘yicha falsafa doktori (PhD) va fan doktori (DSc) ilmiy darajasiga talabgorlarning dissertatsiya ishlari yuzasidan dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan milliy ilmiy nashrlar ro‘yxatiga kiritilgan.

BOSH MUHARRIR:
Temirov Alisher Hoshim o‘g‘li – texnika fanlari falsafa doktori (PhD),Dotsent.
TAHRIRIYAT KENGASH RAISI:
Fozilov Sadriddin Fayzullayevich – texnika fanlari doktori, professor
TAHRIRIYAT KENGASH A’ZOLARI

TEXNIKA FANLARI/ TECHNICAL SCIENCES/ ТЕХНИЧЕСКИЕ НАУКИ

Bozorov G‘ayrat Rashidovich
Buxoro Davlat Texnika Universiteti, professor
Abdurahmonov Olim Rustamovich
Buxoro Davlat Texnika Universiteti, professor
Adizov Bobur Zamirovich
O‘zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti, professor
Hayitov Ruslan Rustamjonovich
Buxoro Davlat Texnika Universiteti, professor
Jumayev Qayum Karimovich
Buxoro Davlat Texnika Universiteti, professor.
Berdiyev Dorob Murodovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor
Maxmudov Rafiq Amonovich
Buxoro Davlat Texnika Universiteti, professor.
Tilloyev Lochin Ismatilloyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Raxmonov Qahramon Sanakulovich
Buxoro Davlat Texnika Universiteti, dotsent.
Barakayev Nusratullo Rajabovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor.
Jo‘rayev Xayrullo Fayziyevich
Buxoro Davlat Texnika Universiteti, professor.
Panoyev Erali Rajabboyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Ataullayev Sherzod Nabiullyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Malik Negmurodovich
Buxoro Davlat Texnika Universiteti, dotsent.
Axmedova Ozoda Bahronovna
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti.
Safarov Bahri Jumayevich
Buxoro Davlat Texnika Universiteti, dotsent.
Xaydarov Axtam Amonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Safarov Jasur Alijon o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.

KIMYO FANLARI/ CHEMICAL SCIENCES/ ХИМИЧЕСКИЕ НАУКИ

Axmedov Voxid Nizomovich
Buxoro Davlat Texnika Universiteti, professor
Umarov Baqo Bafoyevich
Buxoro Davlat Universiteti, professor
Kodirov Abduaxad Abduraximovich
Qarshi davlat Universiteti, professor
Qurbonov Mingniqul Jumagulovich
Qarshi davlat Universiteti, dotsent
Turemuratov Sharibay Naurizbaevich
O‘zbekiston Respublikasi Fanlar akademiyasi Qoraqalpog‘iston bo‘limi Bosh ilmiy kotibi, professor
Tursunov Kaxor Shonazarovich
Qarshi davlat Universiteti, professor
Do‘stov Hamro
Buxoro Davlat Texnika Universiteti, professor
Panoyev Nodir Shavkatovich
Buxoro Davlat Texnika Universiteti, dotsent.
Olimov Bobur Bahodir o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Kasimov Sherzod Abduazimovich
Termiz Davlat Universiteti, professor.
Aliqulov Rustam Valiyevich
Termiz Davlat Universiteti, professor
Tropov Xamza Tursunovich
Samarqand Davlat Universiteti, professor

IQTISODIYOT FANLARI/ ECONOMICS/ ЭКОНОМИЧЕСКИЕ НАУКИ

Nizamov Aslitdin Badreddinovich
Buxoro Davlat Texnika Universiteti, professor.
Boboyev Akmal Choriyevich
Buxoro Davlat Texnika Universiteti, professor.
Ergashxojayeva Shaxnoza Djasurovna
Toshkent davlat iqtisodiyot universiteti, professor
Saidaxmedova Nodira Ilxomovna
Toshkent davlat iqtisodiyot universiteti, professor
Muxsinov Bekzod Toxirovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shakarov Qulmat Ashirovich
Renessans ta'lim universiteti,dotsent.
Jumaeva Zulfiya Qayumovna
Buxoro Davlat Texnika Universiteti, dotsent.
Norova Salomat Yusupovna
Buxoro Davlat Texnika Universiteti, dotsent.
Avezova Shaxnoza Maxmudjonovna
Buxoro Davlat Texnika Universiteti, dotsent.
Jalilov Jamshid G‘anijonovich
Toshkent davlat iqtisodiyot universiteti, dotsent
Xasanova Gulruh Djumanazarovna
Buxoro Davlat Texnika Universiteti, dotsent.

PEDAGOGIKA FANLARI/ PEDAGOGICAL SCIENCES / ПЕДАГОГИЧЕСКИЕ НАУКИ

Siddiqova Sadoqat G‘afforovna
Buxoro Davlat Texnika Universiteti, dotsent.
Zoyirov Erkin Xalilovich
Buxoro Davlat Texnika Universiteti, DSc, professor
Gaffarov Feruz Hasanovich
Puchon Universiteti, *professor*
Ruziyev Davron Yuldashevich
Buxoro Davlat Universiteti, professor
Muradova Firuza Rashidovna
Buxoro Davlat Texnika Universiteti, professor.
Yunusova Gulandon Samiyevna
Buxoro Davlat Texnika Universiteti, professor.
To‘xtayeva Zebo Sharifovna
Buxoro Davlat Texnika Universiteti, professor.
Navro‘zova Gulchixra Nigmatovna
Buxoro Davlat Texnika Universiteti, dotsent.
Uzoqova Laylo Polvonovna
Buxoro muhandislik – texnologiya instituti, dotsent.
To‘raqulova Marjona Qiyom qizi
Buxoro Davlat Texnika Universiteti, dotsent.
Muradova Zarina Rashidovna
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Sanjar Aslonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shodiyev Jahongir Jo‘raqulovich
Buxoro Davlat Texnika Universiteti, dotsent
Gaffarov Laziz Hasanovich
Buxoro muhandislik – texnologiya instituti, dotsent.

Tahririyat a’zolarning to‘liq ro‘yxati jurnalning www.devos.uz saytida keltirilgan.

Tahririyat hay’ati a’zolarining tarkibini o‘zgartirish hamda qo‘shimcha kiritish yoki tahririyat a’zoligidan chiqarish bosh muharrir tomonidan amalga oshiriladi.

ISSN 3030-3907

Development of science ilmiy jurnali 2024 yil 15 aprel kuni №256445 sonli guvohnoma bilan Davlat ro‘yxatidan o‘tgan.

To‘plam elektron ko‘rinishda (pdf formatida) mualliflarga taqdim etiladi. To‘plamga kiritiladigan maqolalarning mazmuni, ulardagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligiga hamda har qanday fikr - mulohazalarga mualliflarning o‘zlari mas’ul hisoblanadilar. Talablarga javob bermaydigan maqolalar to‘plamga kiritilmaydi. Tashkiliy qo‘mita maqola matnini qisqartirish, qisman tuzatish kiritish, seksiyalarga joylashtirish huquqiga ega.

ISSN 3030-3907

KATTA HAJMDAGI MATNLI MA'LUMOTLARNI (BIG DATA) HISSIYOTLAR TAHLILI UCHUN TAYYORLASH METODIKASI

Olimjanova Feruza Sanjar qizi
Kompyuter lingvistikasi mutaxassisligi 1-bosqich magistranti
ToshDO'TAU,
E-mail: feruzaolimjanova78@gmail.com

Annotatsiya. Ushbu maqolada ijtimoiy tarmoqlardan olinadigan ulkan hajmdagi tartibsiz (unstructured) matnlarni hissiyotlar tahlili (Emotion Detection) uchun tayyorlashning nazariy va metodologik asoslari tadqiq etiladi. Tadqiqotda Big Data'ning "5V" tamoyillari lingvistik nuqtai nazardan tahlil qilinib, ma'lumotlarni yig'ish, lingvistik normalizatsiya, annotatsiya metodologiyasi va axborot etikasi masalalari tizimlashtirilgan. Ayniqsa, o'zbek tilining agglyutinativ tabiati va poliorfizm sharoitida matnlarga dastlabki ishlov berishning (preprocessing) o'ziga xos xususiyatlari ilmiy asoslangan. Taklif etilayotgan metodika o'zbek tili uchun avtomatlashtirilgan hissiyotlarni aniqlash tizimlarini yaratishda nazariy poydevor bo'lib xizmat qiladi.

Kalit so'zlar: Big Data, hissiyotlar tahlili, NLP, o'zbek tili, lingvistik normalizatsiya, lemmatizatsiya, annotatsiya, axborot etikasi, "5V" tamoyili, korpus lingvistikasi.

Abstract. This article explores the theoretical and methodological foundations of preparing massive unstructured text data from social networks for emotion detection. The study analyzes the "5V" principles of Big Data from a linguistic perspective and systematizes the issues of data collection, linguistic normalization, annotation methodology, and information ethics. Special scientific emphasis is placed on the features of text preprocessing in the context of the agglutinative nature of the Uzbek language and poly-orthography (Cyrillic and Latin scripts). The proposed methodology serves as a theoretical framework for developing automated emotion detection systems for the Uzbek language.

Keywords: Big Data, emotion detection, NLP, Uzbek language, linguistic normalization, lemmatization, annotation, information ethics, "5V" principles, corpus linguistics.

Аннотация. В данной статье исследуются теоретические и методологические основы подготовки огромных массивов неструктурированных текстовых данных из социальных сетей для анализа эмоций (Emotion Detection). В исследовании анализируются принципы «5V» Big Data с лингвистической точки зрения, а также систематизируются вопросы сбора данных, лингвистической нормализации, методологии аннотирования и информационной этики. Научно обоснованы особенности предварительной обработки текстов (preprocessing) в условиях

ISSN 3030-3907

агглютинативной природы узбекского языка и полиорфографии. Предложенная методика служит теоретическим фундаментом для создания автоматизированных систем распознавания эмоций для узбекского языка.

Ключевые слова: Big Data, анализ эмоций, NLP, узбекский язык, лингвистическая нормализация, лемматизация, аннотирование, информационная этика, принцип «5V», корпусная лингвистика.

KIRISH.

Bugungi kunda axborot texnologiyalarining jadal rivojlanishi natijasida insoniyat tomonidan ishlab chiqilayotgan ma’lumotlarning salmoqli qismi matnli kontent ko‘rinishida namoyon bo‘lmoqda. Ayniqsa, raqamli muloqot platformalari – ijtimoiy tarmoqlar, blogpostlar va onlayn xizmatlarga qoldirilayotgan izohlar ulkan hajmdagi "Katta ma’lumotlar" (Big Data) oqimini yuzaga keltirdi [1]. Ushbu ma’lumotlarning lingvistik tabiati an’anaviy yozma nutqdan tubdan farq qilib, u o‘zida ham og‘zaki, ham yozma nutq belgilarini mujassamlashtiradi [2].

Ma’lumotlarning "tartibsiz" (unstructured) ekanligi muammosi

Ijtimoiy tarmoqlardagi matnli ma’lumotlarning asosiy xususiyati va ayni paytda tadqiqotchilar uchun bosh muammosi ularning strukturalashmagan (tartibsiz) ko‘rinishga egaligidir [3]. Akademik matnlardan farqli o‘laroq, ijtimoiy media kontenti qat’iy grammatik qoidalarga bo‘ysunmaydi. Unda imloviy xatolar, qisqartmalar, jargonlar, emojilar va ko‘p tilli (code-switching) elementlarning aralashib ketishi kuzatiladi. Bunday "lingvistik shovqin" matnning mazmuni va undagi hissiy yuklamani (emotion) avtomatik aniqlash jarayonini murakkablashtiradi [4]. Binobarin, tartibsiz matnlarni mashinali o‘rganish algoritmlari uchun tushunarli formatga keltirish metodikasini ishlab chiqish dolzarb ilmiy vazifa hisoblanadi.

Big Data konsepsiyasining fundamental "5V" tamoyillari hissiyotlar tahlili (sentiment analysis) kontekstida quyidagicha namoyon bo‘ladi [5]:

Volume (Hajm): Terabaytlab matnli ma’lumotlarni an’anaviy usulda (qo‘lda) qayta ishlashning imkonsizligi.

Velocity (Tezlik): Ijtimoiy tarmoqlardagi fikrlarning real vaqt rejimida o‘zgarishi va tahlil tizimlaridan ham shunday tezlik talab qilinishi [6].

Variety (Xilma-xillik): Ma’lumotlarning turli tillarda, har xil alifbolarda va turli formatlarda (post, izoh, rasm ostidagi qisqa matn) ekanligi.

(Ishonchlilik): Matnlardagi sarkazm, istehzo yoki feyk xabarlar tufayli hissiyotning haqiqiyligini aniqlashdagi qiyinchiliklar.

Value (Qiymat): Ushbu ulkan va tartibsiz matnlar ichidan jamiyat kayfiyatini aks ettiruvchi foydali xulosalarni ajratib olish [7].

MAVZUNING DOLZARBLIGI.

ISSN 3030-3907

An’anaviy matn tahlili usullari, asosan, lug‘atga asoslangan yoki statik qoidalar tizimiga tayanadi. Biroq, Big Data sharoitida tilning dinamik o‘zgarishi, yangi neologizmlar va hissiyot ifodalashning norasmiy usullari (masalan, "yaxshshshiii" kabi so‘z cho‘zilishi) an’anaviy tizimlarning aniqligini pasaytirib yubormoqda [8]. Shuningdek, ma’lumotlar hajmining eksponentsial o‘sishi hisoblash quvvatlaridan samarali foydalanishni talab etuvchi yangi metodologik yondashuvlarga ehtiyoj tug‘dirmoqda. Shu bois, matnlarni dastlabki qayta ishlashning nazariy asoslarini takomillashtirish, katta hajmdagi ma’lumotlarni hissiyotlar tahlili uchun sifatli tayyorlashning zamonaviy metodikasini yaratish bugungi kun tilshunosligi va IT sohasining kesishmasidagi eng dolzarb masalalardan biridir.

ADABIYOTLAR TAHLILI.

Katta hajmdagi matnlarni tahlil qilish va ulardagi hissiy indikatorlarni aniqlash masalasi jahon va mahalliy ilm-fanda turlicha yondashuvlar asosida o‘rganib kelinmoqda. Hissiyotlar tahlili (Emotion Detection) va munosabat tahlili (Sentiment Analysis) sohalari dastlab ingliz tili materiallari asosida shakllangan bo‘lsa-da, bugungi kunda o‘zbek tili segmentida ham salmoqli tadqiqotlar amalga oshirilmoqda.

Jahon miqyosida B. Liu [4] va Pang hamda Lee [7] kabi tadqiqotchilar matnlarni tayyorlashda mashinali o‘rganish algoritmlarining o‘rni va kontekstual tahlilning nazariy asoslarini yaratganlar. Ular matndagi subyektivlikni aniqlashda "shovqin"larni tozalashning ahamiyatini fundamental darajada isbotlashgan.

O‘zbek filologiyasi va kompyuter lingvistikasi kesishmasida professor B. Elov va uning maktabi tomonidan amalga oshirilgan ishlar alohida ahamiyatga ega. Xususan, B. Elov o‘zbek tili korpusini yaratish va matnlarni avtomatik tahlil qilish metodologiyasi bo‘yicha fundamental tadqiqotlar olib borgan. Uning ishlarida o‘zbek tilining agglyutinativ tabiati, so‘z yasovchi va shakl yasovchi affikslarning semantik tahlili masalalari keng yoritilgan [6]. Bu esa Big Data sharoitida lemmatizatsiya jarayonini to‘g‘ri tashkil etish uchun nazariy manba bo‘lib xizmat qiladi.

Ayniqsa, professor Shahlo Hamroyeva tomonidan o‘zbek tili milliy korpusining lingvistik asoslari va so‘z shakllarini modellashtirish bo‘yicha olib borilgan tadqiqotlar ushbu mavzu uchun metodologik poydevor bo‘lib xizmat qiladi [8]. Sh. Hamroyevaning matn birliklarini leksik-semantik jihatdan tizimlashtirish va korpus lingvistikasining nazariy tamoyillari bo‘yicha ishlari, ijtimoiy tarmoqlardagi "tartibsiz" matnlarni ilmiy nuqtai nazardan normalizatsiya qilish va ularni tahlilga tayyorlash jarayonida muhim ilmiy asos bo‘ladi. Uning agglyutinativ tillarda morfologik tahlil modellarini yaratish borasidagi

ISSN 3030-3907

yondashuvlari hissiyotlar tahlilida so‘z o‘zagini aniqlash aniqligini ta’minlashda yuqori qimmatga ega.

So‘nggi yillarda o‘zbek tili segmentida sentiment tahlil va hissiyotlarni aniqlash bo‘yicha A. Abdullayev va G. Raximova kabi tadqiqotchilar tomonidan olib borilayotgan izlanishlar ham diqqatga sazovordir. Ular tomonidan ishlab chiqilgan hissiy lug‘atlar (lexicons) va datasetlarni shakllantirish tamoyillari, ijtimoiy media kontentini klassifikatsiyalashda yangi metodik yondashuvlarni taklif etadi [9].

Biroq, tahlillar shuni ko‘rsatadiki, katta hajmdagi (Big Data) ma’lumotlarni yig‘ishdan to ularni annotatsiya qilishgacha bo‘lgan barcha bosqichlarni qamrab oluvchi yaxlit va tizimli metodika hali ham takomillashtirishga muhtojdir. Ayniqsa, o‘zbek tili ijtimoiy segmentidagi poliorfizm (lotin va kirill aralashuvi) va dialektal xususiyatlarning hissiyot tahliliga ta’siri masalasi ushbu maqolaning tadqiqot obyektini belgilab beradi.

Ma’lumotlarni tanlash va yig‘ishning konseptual tamoyillari

Katta hajmdagi matnli ma’lumotlar bilan ishlashda sifat ko‘rsatkichi miqdor ko‘rsatkichidan ustun turadi. Hissiyotlar tahlili uchun bazani shakllantirish shunchaki ma’lumot yig‘ish emas, balki muayyan ilmiy tamoyillarga asoslangan strategik jarayondir [11].

Reprezentativlik tamoyili

Reprezentativlik – yig‘ilgan ma’lumotlar to‘plamining (dataset) o‘rganilayotgan obekt yoki ijtimoiy guruhning umumiy kayfiyatini qanchalik xolis aks ettira olish darajasidir [12]. Nazariy jihatdan, agar tanlab olingan ma’lumotlar faqat bir turdagi platformadan (masalan, faqat Telegram) yig‘ilsa, tadqiqot xulosalari bir yoqlama bo‘lib qolishi mumkin.

Maqsadli guruh mutanosibligi: Tadqiqot ijtimoiy qatlamlarning yoshi, jinsi va hududiy mansubligini qamrab olishi lozim.

Statistik ishonchlilik: Tanlanma hajmi shunday bo‘lishi kerakki, undan olingan natijalar katta ehtimollik bilan umumiy ijtimoiy fikrga mos kelsin. Bu tamoyil tahlil natijalarining umumlashtirish (generalization) imkoniyatini ta’minlaydi [13].

Ekstraktsiya metodologiyasi: Web-mining va Text-mining

Ma’lumotlarni avtomatlashtirilgan holda ajratib olish (extraction) jarayoni ikki xil nazariy yondashuvning kesishmasida amalga oshiriladi:

Web-mining (Veb-qazib olish): Bu jarayonda asosiy e’tibor internetdagi strukturalashmagan ma’lumotlarni qidirish va yig‘ishga qaratiladi [10]. Bu yerda Web Content Mining (kontentni ajratish) va Web Structure Mining (sahifalararo bog‘liqlikni tahlil qilish) muhim rol o‘ynaydi.

ISSN 3030-3907

Text-mining (Matn-qazib olish): Web-mining orqali yig‘ilgan xom ma’lumotlar ichidan aynan matnli qismlarni, ya’ni til birliklarini, sintaktik qurilmalarni va semantik ma’nolarni ajratib olish jarayonidir [4].

Nazariy modelda Web-mining "qobiq" bilan ishlasa, Text-mining bevosita "mag‘iz" (matn mazmuni) bilan ishlaydi. Hissiyotlar tahlili uchun ushbu ikki metodologiyaning sinxron ishlashi ma’lumotlar sifatini oshiradi.

Manba ishonchliligi va "shovqin" tasnifi

Big Data sharoitida ma’lumotlar manbasining ishonchliligi (Veracity) tahlil aniqligini belgilovchi eng asosiy omildir. Nazariy jihatdan manbalarni "shovqin" darajasiga ko‘ra quyidagicha tasniflash mumkin:

Texnik shovqinlar: Matndagi HTML teglar, reklamalar va takrorlanuvchi tizimli xabarlar [3].

Ijtimoiy shovqinlar (Spam va Botlar): Sun’iy ravishda yaratilgan, hissiyot tahlilini manipulyatsiya qilishga qaratilgan avtomatik xabarlar. Ular haqiqiy inson hissiyotini ifodalamaydi.

Feyk va noto‘g‘ri ma’lumotlar: Mazmunan yolg‘on bo‘lgan xabarlar, garchi ular hissiy bo‘yoqqa ega bo‘lsa-da, ijtimoiy kayfiyatni noto‘g‘ri aks ettirishi mumkin.

Manbalarni saralashda ishonchlilik filtrlari (reliability filters) metodikasini qo‘llash, tadqiqotning ilmiy qiymatini saqlab qolishga xizmat qiladi. Faqat verifikatsiyadan o‘tgan yoki foydalanuvchi faolligi tabiiy bo‘lgan manbalarni tanlash hissiyotlar tahlilidagi xatoliklar ehtimolini kamaytiradi.

Matnga ishlov berishning iyerarxik bosqichlari:

Katta hajmdagi tartibsiz matnlarni hissiyotlar tahlili (Emotion Detection) uchun tayyorlash jarayoni murakkab lingvistik bosqichlarni talab etadi. Bu jarayon xom ma’lumotni (raw data) semantik birlikka aylantirishning iyerarxik tizimidir.

Lingvistik normalizatsiya va standartlashtirish muammolari

Lingvistik normalizatsiya – bu matndagi har xil ko‘rinishdagi leksik birliklarni yagona kanonik (standart) shaklga keltirish jarayonidir [14]. Ijtimoiy tarmoqlar kontekstida bu bosqich bir qator nazariy muammolarni yuzaga keltiradi:

Dialektizmlar va adabiy til munosabati: Ijtimoiy media foydalanuvchilari ko‘pincha hududiy dialektlar yoki submadaniy jargonlardan foydalanishadi. Nazariy jihatdan, ushbu birliklarni adabiy tilga o‘tkazishda so‘zning hissiy bo‘yog‘i yo‘qolishi xavfi mavjud.

Grafik normalizatsiya: O‘zbek tili segmenti uchun xos bo‘lgan lotin va kirill alifbolarining aralash qo‘llanilishi matnni kompyuter tomonidan tahlil qilishda "so‘zlar xilma-xilligi"ni sun’iy oshirib yuboradi.

ISSN 3030-3907

Morfologik va sintaktik tahlil: Lemmatizatsiyaning roli

Morfologik tahlilning eng muhim bo‘g‘ini – lemmatizatsiya (so‘zni uning boshlang‘ich lug‘aviy shakliga keltirish) hisoblanadi. Hissiyotlarni aniqlashda ushbu jarayonning nazariy ahamiyati quyidagilardan iborat:

Leksik zichlikni kamaytirish: Agglyutinativ tillarda birgina so‘z o‘zagi yuzlab affikslar bilan kelishi mumkin. Lemmatizatsiya algoritmi so‘z o‘zagini aniqlash orqali modelga hissiy lug‘at (emotion lexicon) bilan samarali bog‘lanish imkonini beradi [3].

Hissiy yuklamaning saqlanishi: Nazariy tahlilda "lemmatizatsiya" va "stemming" farqlanishi shart. Lemmatizatsiya so‘zning morfologik tabiatini saqlagan holda tahlil aniqligini ta’minlaydi.

To‘xtovchi so‘zlar (Stop-words) nazariyasi

Matn tahlilida informatsion yuklamaga ega bo‘lmagan so‘zlarni olib tashlash an’anaviy usul hisoblanadi. Biroq, hissiyotlar tahlilida bu jarayonga ehtiyotkorona nazariy yondashuv talab etiladi:

Funksional so‘zlar va hissiyot: Odatda "va", "bilan", "uchun" kabi yordamchi so‘zlar olib tashlanadi. Ammo ba’zi yordamchi so‘zlar (masalan, "-mi", "nahotki") gapning so‘roq yoki hayrat ohangini belgilaydi.

Inkor muammosi (Negation handling): Nazariy jihatdan "emas", "yo‘q" kabi inkor so‘zlari stop-words ro‘yxatiga kiritilishi katta xatodir [13]. Hissiyotlar tahlili metodikasida inkor birliklarini matn tarkibida saqlab qolish tamoyili ilgari suriladi.

2-jadval.

Hissiyotlar tahlili (Emotion Detection) jarayonida matnga dastlabki ishlov berish (preprocessing) bosqichlarining funksional tavsifi

Ishlov berish bosqichi	Maqsadi	Hissiyot tahlilidagi ahamiyati
Normalizatsiya	Matnni yagona standartga keltirish	Kirill/Lotin variantlarini birlashtirish, xatolarni tuzatish.
Tokenizatsiya	Matnni so‘zlarga ajratish	Hissiy lug‘atlar bilan solishtirish uchun asosiy birlikni yaratish.
Lemmatizatsiya	So‘zni o‘zak shakliga keltirish	Agglyutinativ qo‘shimchalarni ajratib, o‘zakdagi hissiyotni topish.

ISSN 3030-3907

Stop-words filtri	Ma’nosiz so‘zlarni olib tashlash	Modelni keraksiz so‘zlardan tozalash (Inkor yuklamalaridan tashqari).
Negation handling	Inkor birliklarini saqlash	"Yaxshi emas" birikmasidagi negativlikni yo‘qotmaslik.

Yuqorida keltirilgan 2-jadvalda matnli ma’lumotlarni hissiyotlar tahlili uchun tayyorlashning iyerarxik bosqichlari tizimlashtirilgan. Jadval tahlili shuni ko‘rsatadiki, **normalizatsiya** va **lemmatizatsiya** bosqichlari agglyutinativ tillar, xususan, o‘zbek tili uchun bazaviy filtr hisoblanadi. Zero, o‘zbek tilidagi lotin va kirill alifbolari poliorfizmi sharoitida matnlarni yagona standartga keltirmasdan turib, keyingi tahlil bosqichlarida yuqori aniqlikka erishib bo‘lmaydi.

Jadvalning e’tiborli jihati shundaki, unda **Stop-words filtri** va **Negation handling (inkorni hisobga olish)** o‘rtasidagi farq aniq ko‘rsatilgan. An’anaviy tahlildan farqli o‘laroq, hissiyotlarni aniqlash metodikasida inkor birliklarini saqlab qolish gapning semantik yo‘nalishini (ijobiy/salbiy) saqlab qolishda fundamental ahamiyat kasb etadi. Ushbu bosqichlar ketma-ketligi "shovqinli" Big Data ma’lumotlarini ilmiy jihatdan asoslangan "Oltin standart" datasetiga aylantirishning asosiy mexanizmi bo‘lib xizmat qiladi.

Ekspertli baholash va annotatsiya metodologiyasi

Katta hajmdagi matnli ma’lumotlarni tayyorlashning yakuniy va eng mas’uliyatli bosqichi bu – matnlarga tegishli hissiyot toifalarini biriktirish, ya’ni annotatsiya jarayonidir. Bu jarayon mashinali o‘rganish modellari uchun "o‘qituvchi" vazifasini o‘taydi [10].

Subyektivlik muammosi va annotatorlararo kelishuv

Hissiyotlarni aniqlash (Emotion Detection) obektiv o‘lchov birliklariga ega bo‘lmagan soha bo‘lgani uchun, bu yerda subyektivlik muammosi birinchi planga chiqadi.

Idrok tafovuti: Bir xil ijtimoiy tarmoq izohi turli insonlar tomonidan turlicha qabul qilinishi mumkin.

Nazariy yechim: Ushbu muammoni bartaraf etish uchun nazariyada Inter-Annotator Agreement (Annotatorlararo kelishuv) koeffitsiyentlaridan (masalan, Koen kappasi) foydalaniladi [11].

Hissiyotlarni toifalashning nazariy modellari

Nazariy tahlilda asosan ikki xil yondashuv ajralib turadi:

ISSN 3030-3907

Ekman modeli (Diskret model): Pol Ekman tomonidan ilgari surilgan ushbu model hissiyotlarni 6 ta bazaviy toifaga ajratadi: shodlik, qayg‘u, qo‘rquv, g‘azab, hayrat va jirkanish [13].

Plutchik modeli (Hissiyotlar g‘ildiragi): Robert Plutchik modeli hissiyotlarning yanada murakkab va iyerarxik tuzilishini taklif etadi. Bu yondashuv matndagi hissiyotning nafaqat turini, balki uning kuchini aniqlashda ham nazariy asos bo‘lib xizmat qiladi [13].

Oltin standart (Gold Standard) yaratish talablari

Metodologiyaning cho‘qqisi bu – "Oltin standart" (Gold Standard) datasetini yaratishdir. Bu ideal darajada aniq belgilangan va tekshirilgan ma’lumotlar to‘plami bo‘lib, u sifat nazorati, muvozanatlilik (Class Balance) va kontekstual boylik kabi nazariy talablarga javob berishi kerak [14].

Axborot etikasi va huquqiy aspektlar:

Big Data tahlili nafaqat texnik, balki jiddiy etik va huquqiy masalalarni ham yuzaga keltiradi. Tadqiqot obyekti sifatida insonning raqamli nutqi olinganda, axborot etikasi (information ethics) tamoyillariga rioya qilish shart [15].

Maxfiylik nazariyasi va etik chegaralar.

Anonimlashtirish tamoyili: Katta hajmdagi matnlarni tayyorlash jarayonida foydalanuvchining shaxsini identifikatsiya qiluvchi barcha ma’lumotlar matn tarkibidan olib tashlanishi shart [16].

Ma’lumotlardan foydalanishning maqsadliligi: Yig‘ilgan ma’lumotlar faqat e’lon qilingan ilmiy maqsadda qo‘llanilishi, foydalanuvchining obro‘siga putur yetkazadigan maqsadlarda foydalanmaslik kafolatlanishi kerak.

Ma’lumotlar egaligi va manfaatlar muvozanati

Ijtimoiy tarmoq foydalanuvchisi kontenti (User-Generated Content - UGC) va tadqiqotchining huquqlari o‘rtasidagi muvozanatni saqlash huquqiy tahlilning markazida turadi. Nazariy yondashuv shuni ko‘rsatadiki, tadqiqotchi foydalanuvchining "shaxsiy hududi" va "ommaviy hududi" o‘rtasidagi farqni anglagan holda ish tutishi kerak [27].

Shu nuqtai nazardan, Big Data bilan ishlovchi har bir metodologiya xalqaro ma’lumotlarni himoya qilish reglamentlari (masalan, GDPR) va milliy qonunchilik talablariga mos kelishi, tadqiqotning huquqiy legitimligini ta’minlashi zarur [28].

FOYDALANILGAN ADABIYOTLAR RO‘YXATI

1. Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think.
2. Inmon, W. H., & Linstedt, D. (2014). Data Architecture: A Primer for the Data Scientist.

ISSN 3030-3907

3. Liu, B. (2020). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions.
4. Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety.
5. Zikopoulos, P., & Eaton, C. (2011). Understanding Big Data.
6. Elov, B. B. (2020). O‘zbek tili kompyuter lingvistikasining nazariy va amaliy masalalari.
7. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis.
8. Hamroyeva, Sh. M. (2018). O‘zbek tili milliy korpusining lingvistik asoslari.
9. Abdullayev, A. (2023). Hissiyotlar tahlilida o‘zbek tili korpusidan foydalanish va annotatsiya tamoyillari.
10. Liu, B. (2011). Web Data Mining.
11. Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing.
12. Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis.
13. Ekman, P. (1992). An argument for basic emotions.
14. Plutchik, R. (1980). Emotion: A Psychoevolutionary Synthesis.
15. Boyd, D., & Crawford, K. (2012). Critical questions for Big Data.
16. Voigt, P., & Von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR).

ISSN 3030-3907