

**MANAGEMENT AND ECONOMICS
SCIENTIFIC RESEARCH JOURNAL**

**MENEJMENT VA IQTISODIYOT
ILMIY-TADQIQOT JURNALI**

**НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЖУРНАЛ
МЕНЕДЖМЕНТА И ЭКОНОМИКИ**



MANAGEMENT AND ECONOMICS SCIENTIFIC RESEARCH JOURNAL

☎ +998 95 651 90 00

✉ journals@timeedu.uz

📍 Shota Rustaveli ko'chasi, 114

🌐 tmijournals.ndc-agency.uz

✈ @TMII_ilmiy_bolim

Bosh muharrir:

Narbayev Sharafatdin Kengeshovich

Bosh muharrir o'rinbosari:

Nosirova Nargiza Jamoliddin qizi

Muharrir:

Qarshiyeva Shahnoza Davlatovna

Tahrir hay'ati:

1. Sharipov Kongratbay Avezimbetovich – O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vaziri, t.f.d., prof.;
2. Axmedov Xojiakbarxon Dilshatovich – Toshkent menejment va iqtisodiyot instituti rektori i.f.n., dotsent;
3. Gulyamov Saidasror Saidaxmedovich – O'zbekiston Fanlar akademiyasi akademigi, i.f.d, professor;
4. Gulyamov Saidaxror Saidaxmedovich – O'zbekiston Fanlar akademiyasi akademigi i.f.d, professor;
5. Sultanov Mansur Qilichovich – Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi "Inno" innovatsion o'quv-ishlab chiqarish texnoparki bosh direktori, t.f.f.d., dotsent;
6. Chertovitskiy Aleksandr Stepanovich - „Toshkent irrigatsiya va qishloq xo'jaligini mexanizatsiyalash muxandislari instituti“ Milliy tadqiqot universiteti professori, iqtisodiyot fanlari doktori;
7. Popkova Yelena Gennadyevna - Rossiya Federatsiyasi avtonom notijorat tashkiloti "Ilmiy kommunikatsiyalar instituti" prezidenti, iqtisodiyot fanlari doktori, professor;
8. Dr Goh Khang Wen – Malaziyaning INTI Xalqaro universiteti prorektori;
9. Wang Cheng – XXR "Hebei Institute of Mechanical and Electrical Technology" instituti i.f.f.d., dotsent;
10. Yo'ldoshev Nuriddin Qurbonovich – Toshkent menejment va iqtisodiyot instituti Biznes boshqaruvi va moliya kafedrasini i.f.d., professor;
11. Akramov Tohir Abdirahmonovich Toshkent menejment va iqtisodiyot instituti Biznes boshqaruvi va moliya kafedrasini i.f.d., professor;
12. Temirov Abdulaziz Alimjonovich – Toshkent menejment va iqtisodiyot instituti Biznes boshqaruvi va moliya kafedrasini i.f.n., professor;
13. Rajabov Nazirjon Razzaqovich – Toshkent menejment va iqtisodiyot instituti Iqtisodiyot kafedrasini mudiri, i.f.n., dotsent;
14. Xotamov Ibodullo Sadulloyevich – Toshkent menejment va iqtisodiyot instituti Iqtisodiyot kafedrasini i.f.n., professor;
15. Sobirov Aziz Avazbekovich – Toshkent menejment va iqtisodiyot instituti Biznes boshqaruvi va moliya kafedrasini mudiri, i.f.f.d., dotsent;
16. Raximova Shahlo Baxtiyorovna – Ta'lim sifatini nazorat qilish bo'limi boshlig'i, p.f.f.d., dotsent;
17. Kadirova Istora Bobirovna – Toshkent menejment va iqtisodiyot instituti Karyera markazi direktori, i.f.f.d.;
18. Turakulov Otabek Xolmirzayevich – Toshkent menejment va iqtisodiyot instituti Muhandislik va axborot texnologiyalari kafedrasini mudiri, t.f.f.d.

MUHANDISLIK OYLIGI

“RAQAMLI TEXNOLOGIYALARNI RIVOJLANTIRISH VA INNOVATSION MUHANDISLIK YECHIMLARI: ZAMONAVIY TENDENSIYALAR VA ISTIQBOLLAR” MAVZUSIDA RESPUBLIKA ILMIY-AMALIY ANJUMANI TO‘PLAMI MATERIALLARI

17-18 APREL, 2025

TOSHKENT

MUNDARIJA

1.	Primova X.A., Xaniyev N.M., Nodirov M.T. RFID TEXNOLOGIYASI ORQALI OLINGAN MA'LUMOT ASOSIDA QORAMOL KASALLIKLARINI ANIQLASHNING LOYIHALASH	6
2.	Doshanova M.Y., Otaxanova B.I., Aliyev R.R. USING ARTIFICIAL INTELLIGENCE TO AUTOMATE THE PROCESS OF COLLECTING AND ANALYZING DATA FROM ONLINE JOB POSTINGS	10
3.	Eshonqulov E.Sh., Abdiyeva H.S., Abdurahmonov M.S. SUN'IY YO'LDOSH TASVIRLARIDA PANSARPENING YONDASHUVI NATIJALARINI BAHOLASH BAYONNOMALARI	23
4.	Abduraxmanova Z.T. O'ZBEKISTONDA QISHLOQ XO'JALIGINING RAQAMLI TRANSFORMATSIYASI	26
5.	Abdiyeva K.S., Shamsiyeva K., Oblokulov S. GOMPERTZ ALGORITHM FOR THE DETECTION OF BREAST CANCER USING MAMMOGRAPHIC IMAGES	31
6.	Abdiyeva K.S., Shamsiyeva K., Olimjonova S. TOPOGRAPHIC MAP BASED SEGMENTATION OF MEDICAL IMAGES	35
7.	Radjabov B.A. O'ZBEKISTONNING TASHQI SAVDODA LOGISTIKA XARAJATLARI: RAQAMLI YECHIMLAR VA ULARNING EKSPORT RAQOBATBARDOSHLIGIGA TA'SIRI	39
8.	Primova X.A., Vaydullayeva M.F. SUN'IY INTELLEKTNI TALQIN QILISH VA QAROR QABUL QILISH: NORAVSHAN FIKRLASH TIZIMLARINI HISOBLASH MODELLARINI CHUQUR O'RGANISH	49
9.	Ulashov A.E., Obilov H.X. MATEMATIK MASALALARNI ZAMONAVIY DASTURLASH TILLARI YORDAMIDA YECHISH	57
10.	Mamatov N.S., Asqarov A.A. HAVO SIFATI INDEKSINI BASHORATLASHNING REGRESSION MODELI	69
11.	Mirpulatova L.M. AXBOROT TEXNOLOGIYALARI VA RAQAMLI TRANSFORMATSIYANING BANK FAOLIYATIDAGI ROLI	75
12.	Ниезматова Н.А., Жалелов Қ.А., Абдуллаева Б.М. НУТҚ СИГНАЛЛАРИДАН ШОВҚИНЛАРНИ БАРТАРАФ ЭТИШДА ФИЛЬТРЛАР ЖУФТЛИГИНИ ҚЎЛЛАШ ЁНДАШУВИ	84
13.	Mamatov N.S., Jo'rayev I.A., Jalelova M.M., Usarov J.A. TIBBIY TASVIRLAR BAZALARI TAHLILI	88
14.	Mamatov N.S., Jo'rayev I.A., Jalelova M.M., Jumayev B.J. TIBBIY TASVIRLARNI TAHLIL QILISH USUL VA ALGORITMLARI	92
15.	Mamatov N.S., Jalelova M.M., Jo'rayev I.A., Turakulova Sh.A. TIBBIY 2D TASVIRLAR ASOSIDA 3D TASVIRLARNI SHAKLLANTIRISH USULLARI	96
16.	Niyozmatova N.A., Madaminjonov A.D. INSON HIS-TUYG'ULARINI ANIQLASH DASTURIY VOSITALARI VA QURILMALARI	99

17.	Мухамедиева Д.Т., Маматов Н.С., Маматов А.А. МАТНЛИ МАЪЛУМОТЛАРНИ ТАСНИФЛАШДА БЕЛГИЛАРНИ ШАКИЛЛАНТИРИШ УСУЛ ВА АЛГОРИТМЛАРИ	103
18.	Raximov M. KASB KOMPETENSIYASIDA RAQAMLI TA'LIM TRANSFORMATSIYASI	106
19.	Turakulov O.X. MATNLI MA'LUMOTLARNI QAYTA ISHLASH BOSQICHLARI	110
20.	Turakulov O.X., Jalelov R.M. MATNLI MA'LUMOTLARGA DASTLABKI ISHLOV BERISH ALGORITMLARI	116
21.	Мухамедиева Д.К. ПОСТРОЕНИЕ ДЕРЕВА МЕРКЛА	123
22.	Mamatov N.S., Nuritdinov N.D., Almuradova N.A. IJTIMOIY TIZIMLARDA RAQAMLI TRANSFORMATSIYA JARAYONLARINI MODELLASHTIRISH	127
23.	Kabildjanov A.S., Pulatov G'.G. OB-HAVO OMILLARINING YURAK-QON TOMIR KASALLIKLARIGA TA'SIRINING SUN'IY INTELLEKT ASOSIDA TAHLILI	130
24.	Mamatov N.S., Kodirov E.S KAFT TASVIRI ASOSIDA SHAXSNI TANIB OLIH YONDASHUVLARI TAHLILI	133
25.	Raximov I.M. MAHALLIY ISHLAB CHIQUVCHILARNING XALQARO STANDARTLARGA MOSLASHUVI – IMKONIYATLAR, YUTUQLAR VA BARQAROR STRATEGIK YECHIMLAR	136
26.	Raximov F.J. TABIIY VA SUN'IY TOLALAR ASOSIDA QAYTA ISHLANGAN MAHSULOTLARNING TEXNOLOGIK SIFATI VA IQTISODIY SAMARADORLIGI	143
27.	Mamatov N.S., Dusanov X.T. NUTQ SIGNALINI XALAQITLARDAN TOZALASH USULLARI	150
28.	Dauletov A.Y., Sabirova S.T., Oydinov S.SH. ARXIVDA MATNLARNI QIDIRISH USULLARI TAHLILI	155
29.	Toshpulatov M. UNVEILING TOMORROW: EXPLORING THE FRONTIERS OF INNOVATION	162
30.	Маматов Н.С., Ниёзматова Н.А., Тожибоева Ш.Х., Яхяев Б.Ю., Машанпин Т.В. ИНСОН ҲИС-ТУЙҒУЛАРИНИ ЮЗ ТАСВИРИ ОРҚАЛИ АНИҚЛАШ ДАСТУРИЙ ВОСИТАЛАРИ ВА ҚУРИЛМАЛАРИ	165
31.	Ниёзматова Н.А., Маматов Н.С., Турғунова Н.М. МАТНЛИ МАЪЛУМОТЛАРНИ ТАСНИФЛАШ УСУЛ ВА АЛГОРИТМЛАРИ ТАҲЛИЛИ	172
32.	Маматов Н.С., Валижонов И.Х., Шукруллоев Б.Р. ТАСВИРЛАРНИ ТАСНИФЛАШДА ҚўЛЛАНИЛАДИГАН УСУЛЛАР	185
33.	Niyozmatova N.A., Xoitqulov A.A., Mamatov A.A. MATNLI MA'LUMOTLARDAN FAKTLARNI AJRATISH TEXNOLOGIYALARI	178
34.	Buribaeva Z.R. DIGITAL TECHNOLOGIES AND ARTIFICIAL INTELLIGENCE IN EDUCATION: CHALLENGES AND OPPORTUNITIES	181

35.	Anarova Sh.A., Ibrohimova Z.E., Boliyeva D.N. R-FUNKSIYA USULINI (RFM) QO'LLAGAN HOLDA KO'P O'LCHOVLI FRAKTAL TUZILISHDAGI TASVIRLARNI GEOMETRIK MODELLARINI ISHLAB CHIQUISH	188
36.	Ibrohimova Z.E., Qarshiboyev J.O', Xaniyev N.M. KOMPYUTER GRAFIKASINING GEOMETRIK ALMASHTIRISHLARI ASOSIDA MURAKKAB FRAKTAL TUZILISHLARNI MATEMATIK VA RAQAMLI MODELLASHTIRISHNI AMALGA OSHIRISH	192
37.	Xalimov A.A. SANOAT KORXONALARIDA QAYTA TIKLANUVCHI ENERGIYA MANBALARIDAN FOYDALANISH YO'LLARI	198
38.	Rustamova M.M. INNOVATIVE TECHNOLOGIES AND INVESTMENT EFFICIENCY IN COTTON- TEXTILE CLUSTERS: OPTIMIZING COSTS THROUGH INDUSTRY 4.0 SOLUTIONS.	204
39.	Suyarov A.M. AXBOROTLARNI VIZUAL TAQDIM ETISHDA INFOGRAFIKA VOSITALARINI YARATISH USULLARI	214
40.	Saminjonov N.A. CHAKANA SAVDO FAOLIYATIDA RAQAMLI VA NEYROMARKETING STRATEGIYALARINING INTEGRATSIYASI	219
41.	Alimov A.A. INDUSTRY 4.0 TALABLARI ASOSIDA MUHANDISLIK TA'LIMINI TASHKIL ETISH MASALALARI	224
42.	Sobirov A.A., Zaripbayeva A.A. SANOAT KLASSTERLARINI SHAKLLANISH MEXANIZMLARI VA ZAMONAVIY MODELLARI	228
43.	Sharipov K.A., Karimov V.A. AQLLI ENERGIYA TIZIMLARI VA QAYTA TIKLANADIGAN ENERGIYA MANBALARIDAN FOYDALANISH USULLARI VA SAMARADORLIGI	237
44.	Маматов Н.С., Самижонов А.Н., Самижонов Б.Н. ИНСОН ЭХТИЁЖЛАРИНИ АНГЛОВЧИ ВА УЛАРГА ЖАВОБ БЕРИШ ҚОБИЛИЯТИГА ЭГА ЭМПАТИК РОБОТЛАР	246
45.	Yusupov R.A., G'ulomov A.X. MAYSALARNI O'RISH VA QOR TOZALASH MASHINASINI MASOFADAN BOSHQARISHNI AVTOMATLASHTIRISH	251
46.	Yusupov R.A., Qurbonov Sh.B. HELIX: FOYDALANUVCHIGA MOSLASHA OLADIGAN, KUNDALIK HAYOTDA YORDAM BERUVCHI SUN'IY INTELLEKT	254
47.	Haydarov E.D., Dusmurodov Q.A. OCHIQ WI-FI TARMOQLAR XAVFSIZLIGI	257
48.	Haydarov E.D., Dusmurodov Q.A. SIMSIZ TARMOQDA XAVFSIZLIK	265
49.	Ro'ziyeva F.K. MAISHIY KIMYO TOVARLARI B2B TAQSIMOTIDA KO'P KANALLI MARKETING STRATEGIYASINING AHAMIYATI	269
50.	Qudratov U.I., Abdullayeva N.R. MASHINANI O'QITISH USULLARIDAN FOYDALANGAN HOLDA ZARARLI DASTURLARNI ANIQLASH	277

51.	Muhamediyeva D.T., Raupova M. ELLIPTIK EGRI CHIZIQ ASOSIDA XAVFSIZ KALIT ALMASHINUVI	279
52.	Muhamediyeva D.T., Raupova M. POST-KVANT KRIPTOGRAFIYA ASOSIDA POLINOMLAR BILAN SHIFRLASH	282
53.	Muhamediyeva D.T., Raupova M. KVANT KALIT ALMASHINUVI BILAN INTEGRATSIYALANGAN RSA SHIFRLASH	285
54.	Mamayusupov A.A. ZAMONAVIY RAQAMLI MUHITDA KIBERXAVFSIZLIK TAHDIDLARI VA ULARNING OLDINI OLISH YO'LLARI	288
55.	Тоиров Ш. А. ОПТИМИЗАЦИЯ ФУНКЦИЙ С ПРИМЕНЕНИЕМ ЭВОЛЮЦИОННЫХ АЛГОРИТМОВ	291
56.	Toirov Sh. A. ARCHITECTURE OF OPERATORS IN QUANTUM ALGORITHMS	397
57.	Olimov S.A. TALABALARNING O'ZLASHTIRISH NATIJALARINI BAHOLASHDA RAQAMLI TEXNOLOGIYALARNING O'RNI	302
58.	Olimov S.A. OLIIY TA'LIM MUASSASALARIDA O'QUV JARAYONINI BOSHQARISHDA RAQAMLI TEXNOLOGIYALARNING AHAMIYATI	309
59.	Uzakova G.Z. ENGAGING ESP LEARNERS: STRATEGIES FOR PRODUCTIVE LESSONS	316
60.	Shafkarov F.X. SANATORIY-SOG'LOMLASHTIRISH MUASSASALARINING ICHKI AUDITINI FUNDAMENTAL ASOSDA TASHKIL ETISH	321
61.	Raximov Sh.F. TIJORAT BANKLARIDA MIJOZLARNING SODIQLIK DASTURLARI ASOSIDA DEPOZITLARNI OSHIRISH YO'LLARI	329
62.	Туракулова Ш.А., Суюнова Н. ОПИСАНИЕ КРАЙНИХ ГИББОВСКИХ МЕР ДЛЯ МОДЕЛИ ИЗИНГА С НЕНУЛЕВЫМ ВНЕШНИМ МАГНИТНЫМ ПОЛЕМ НА РЕШЕТКЕ БЕТЕ	337
63.	Ulashov A.E. ZAMONAVIY TA'LIM TIZIMINING RAQAMLI TRANSFORMATSIYASI: TENDENSIYALAR, MUAMMOLAR VA YECHIMLARI	344
64.	Radjabov B.A. O'ZBEKISTONDA TA'LIMNING RAQAMLI TRANSFORMATSIYASI VA SUN'IY INTELLEKT	353
65.	Mamatov N.S., Ibroximov S.R., Almuradova N.A. O'QUV JARAYONIDA SUN'IY INTELLEKT VOSITALARIDAN FOYDALANISH	359
66.	Kuvandikov J.T., Tojiyev A.H. Axborot resurslaridan foydalanish ko'rsatkichlarini normallashtirish asosida bilim oluvchilarning bilim darajasini obyektiv baholash usullari	364
67.	Saidov S.F. TA'LIMDA SUN'IY INTELLEKTDAN SAMARALI FOYDALANISH MASALALARI	374
68.	Suyunova N.A. TA'LIMDA MANTIQ TUSHUNCHASI VA O'QUVCHILARDA MANTIQIY FIKRLASHNI RIVOJLANTIRISHNING INNOVATSION TEXNOLOGIYALARI	382

69.	Kuvandikov J.T. Tojiyev A.H. TA'LIM OLUVCHILAR BILIMINI BAHOLASHDA NORAVSHAN MANTIQDAN FOYDALANISH	389
70.	Djabbarova M.D. KASB-HUNAR TA'LIMIDA ELEKTRON TA'LIMNING PEDAGOGIK VA PSIXOLOGIK ASOSLARI	393
71.	Anarkulova G.M., Abduraxmanova F.N. OLIIY TA'LIMDA MAXSUS FANLAR BO'YICHA MASHG'ULOTLARDA ISHCHAN VA ROLLI O'YINLAR METODINI QO'LLASH INNOVATSION YONDASHUV	396
72.	Анаркулова Г.М., Абдиганиева З.С. ПРОБЛЕМЫ ИСПОЛЬЗОВАНИЕ СОВРЕМЕННЫХ МЕТОДОВ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ СТУДЕНТОВ КАК ФАКТОР УСПЕВАЕМОСТИ	406
73.	Obidov R.A. TA'LIMDA KEYS TEXNOLOGIYALARINING AHAMIYATI	417
74.	Narbayev Sh.K., Maxkamova N.X. MAXSUS FANLARNI O'QITISHDA RAQAMLI VA INNOVATSION TEXNOLOGIYALARIDAN FOYDALANISHNING METODIK SHART-SHAROITLARI	422
75.	Erhanova N.A. MAKTABGACHA TA'LIM TASHKILOTLARIDA MULTIMEDIA RESURLARIDAN FOYDALANISH SAMARADORLIGI	430
76.	Xidoyatov M.B. OLIIY TA'LIM MUASSASALARIDA RAQAMLI MARKETING FAOLIYATINI BAHOLASH: AXBOROT TEXNOLOGIYALARIGA ASOSLANGAN YONDASHUVLAR	435
77.	Темиров А.А. АКЦИЯДОРЛИК ЖАМИЯТЛАРИ ФАОЛИЯТИГА РАҚАМЛИ БОШҚАРУВНИ ЖОРИЙ ЭТИШ ИСТИҚБОЛЛАРИ	444
78.	В.А.Ахмадалиев ЧОРВАЧИЛИК ХЎЖАЛИКЛАРИНИ ТАШКИЛ ЭТИШ ВА УЛАРГА ЕР АЖРАТИШ ТАРТИБИ	454
79.	Носирова Н.Дж. ЭКСПОРТНЫЙ ПОТЕНЦИАЛ СУБЪЕКТОВ МАЛОГО БИЗНЕСА УЗБЕКИСТАНА: СОВРЕМЕННОЕ СОСТОЯНИЕ И ПЕРСПЕКТИВЫ (2018–2025 ГГ)	459
80.	Матлюбов Х.Ж. ТРАНСФОРМАЦИЯ ПРОМЫШЛЕННОЙ ПОЛИТИКИ: ОСНОВНЫЕ ТЕНДЕНЦИИ И ВЫЗОВЫ	463
81.	Sa'dulloyev H. I. SANOAT KORXONALARINING EKO-MARKETING TIZIMINI BAHOLAH AMALIYOTI	469

MATNLI MA'LUMOTLARGA DASTLABKI ISHLOV BERISH ALGORITMLARI

Turakulov O.X., Jalelov R.M.

“Toshkent menejment va iqtisodiyot instituti”

Annotatsiya. Maqolada matnli ma'lumotlarga dastlabki ishlov berish jarayonining algoritmik asoslari tahlil qilinadi. Dastlab HTML teglar, to'xtash so'zlar, tinish belgilar va o'zaklash kabi usullar orqali matnni tozalash bosqichlari ko'rib chiqiladi. Matnni raqamli ifodalash uchun TF-IDF, so'zlar sumkasi modeli, vektor fazosi modeli kabi klassik usullar yoritilgan. Shuningdek, Word2Vec, CBOW va Skip-Gram kabi neyron tarmoqlarga asoslangan zamonaviy modellar bilan tanishtiriladi. Bu modellar yordamida matndagi semantik bog'liqlik va kontekst saqlanadi, bu esa matnni chuqur tahlil qilish imkonini beradi. Maqolada algoritmlarning nazariy asoslari hamda ularning real sohalaridagi qo'llanish imkoniyatlari ham yoritilgan.

Kalit so'zlar: Matnli ma'lumot, dastlabki ishlov, TF-IDF, so'zlar sumkasi modeli, vektor fazosi, Word2Vec, CBOW, Skip-Gram, semantika, normallashtirish, to'xtash so'zlar.

Abstract. The article analyzes the algorithmic foundations of the pre-processing of text data. First, the stages of text cleaning using methods such as HTML tags, stop words, punctuation marks, and kernelization are considered. Classical methods for numerical representation of text, such as TF-IDF, bag-of-words model, and vector space model, are covered. Modern models based on neural networks, such as Word2Vec, CBOW, and Skip-Gram, are also introduced. These models preserve the semantic connections and context in the text, which allows for in-depth analysis of the text. The article also covers the theoretical foundations of algorithms and their application in real areas.

Keywords: Text data, pre-processing, TF-IDF, bag-of-words model, vector space, Word2Vec, CBOW, Skip-Gram, semantics, normalization, stop words.

Аннотация. В статье анализируются алгоритмические основы процесса предварительной обработки текстовых данных. Сначала проверяются этапы очистки текста с использованием таких методов, как HTML-теги, стоп-слова, пунктуация и корневой анализ. Рассматриваются классические методы численного представления текста, такие как TF-IDF, модель «мешка слов» и модель векторного пространства. Также будут представлены современные модели на основе нейронных сетей, такие как Word2Vec, CBOW и Skip-Gram. Эти модели сохраняют смысловые связи и контекст в тексте, что позволяет проводить глубокий анализ текста. В статье также рассматриваются теоретические основы алгоритмов и их потенциальное применение в реальных приложениях.

Ключевые слова: текстовые данные, предварительная обработка, TF-IDF, модель «мешка слов», векторное пространство, Word2Vec, CBOW, Skip-Gram, семантика, нормализация, стоп-слова.

Hozirgi kunda zamonaviy ma'lumotlarni tahlil qilishdan oldin, axborot muhitida matnli ma'lumotlar ularni tozalash, normallashtirish va katta hajmda va turli ko'rinishlarda raqamli ifodaga keltirish zarur. mavjud bo'lib, ular ko'pincha Dastlabki ishlov berish bosqichi – bu strukturalanmagan bo'ladi. Bunday matnni keyingi intellektual tahlil

jarayonlariga tayyorlash bosqichi bo'lib, u orqali ma'lumotlarning aniqligi va sifati ta'minlanadi.

Matnli ma'lumotlar ko'pincha yuqori darajada strukturalanmagan muhitda uchraydi va ko'pincha odamlar ishtirokida yaratiladi. Ko'p hollarda, veb-hujjatlar ichida matn *Hyper Text Markup Language* (HTML) teglar, xato yozishlar, noaniq so'zlar ishtirok etgan bo'ladi. Bundan tashqari, bitta veb-sahifada bir nechta bo'limlar bo'lishi mumkin, ularning aksariyati reklama yoki boshqa aloqador bo'lmagan axborotlarni taqdim qilishi mumkin. Ushbu effektlarni tegishli ishlov berish bilan yaxshilash mumkin. Ishlov berishda keng tarqalgan usullar quyidagilar:

1. Platformaga asoslangan ajratish va o'girish. Matn platformaga asoslangan kontentlarni o'z ichiga olishi mumkin, masalan, HTML teglari. Bunday hujjatlarni platformaga asoslangan holda tozalash va o'girish kerak. Matnni o'girishda hujjatlardan individual bo'laklar ajratib olinadi. Bo'lak - bu qayta ishlash uchun bo'linmas birlik sifatida ko'rib chiqilgan matn tarkibidagi belgilar ketma-ketligi. Hujjatda bitta so'zning har bir qayd etilishi alohida bo'lak sifatida ko'rib chiqiladi.

2. Bo'laklarni qayta ishlash. Bo'laklarga ajratilgan matnni to'plamda ishlatiladigan atamalarga aylantirish uchun qayta ishlanadi. To'plamda juda tez-tez uchraydigan "ammo", "lekin" va "biroq" kabi so'zlarni olib tashlash mumkin. Ushbu so'zlar, odatda, ko'plab intellektual tahlil dasturlari uchun ahamiyatsiz va

ular faqat katta miqdordagi shovqinni hosil qiladi. Bunday so'zlarga to'xtash so'zlari ham deyiladi. Umumiy predloglar, birikmalar va olmoshlar to'xtash so'zlari hisoblanadi. Umuman olganda, to'xtatish so'zlarining tilga oid lug'atlari ko'pincha mavjud. Bir xil o'zakka ega bo'lgan so'zlar (masalan, turli xil zamondagi so'z) birlashtirilishi uchun so'zlar o'zakka keltiriladi. Tinish belgilar va bosh harflar bilan bog'liq muammolar hal qilindi. Shu nuqtada, individual so'zlarning chastotalarini o'z ichiga olgan ko'p o'lchovli vektor fazosini yaratish mumkin.

3. Normallashtirish. Yuqorida muhokama qilinganidek, analitik vazifalarda hamma so'zlar bir xil ahamiyatga ega emas. To'xtatish so'zlari spektrning bir uchida juda tez-tez uchraydigan so'zlarni anglatadi, ularni ko'rib chiqish kerak emas. Qolgan so'zlarning o'zgaruvchan chastotalari haqida nima qilinadi? Aniqlanishicha, ularni korpusga xos bo'lgan chastotalarga asoslangan holda hujjatdagi atama chastotalarini o'zgartirish orqali ularni biroz boshqacha qilish mumkin. Ko'proq korpusga xos bo'lgan chastotalarga ega bo'lgan atamalar kam og'irliklarga ega. Ushbu uslub teskari hujjat chastotasini normallashtirish deb nomlanadi [78; 50-57-b., 79; 27-34-b.].

Dastlabki ishlov berish ko'p o'lchovli matritsa yaratadi. D bu $n \times d$ hujjat-atama matrisasi bo'lsin. Hujjatlar soni n bilan belgilanadi va atamalar soni d bilan belgilanadi.

Birinchi qadam formatlanmagan matnni belgilar ketma-ketligiga aylantirishdan iborat. Matn ba'zida

binar formatlarda bo'ladi, masalan, Microsoft Word yoki Adobe portable document (PDF) formatida uchraydi. Boshqacha qilib aytganda, baytlarni quyidagi omillar asosida belgilar ketma-ketligiga aylantirishimiz lozim:

1. Muayyan matnli hujjat Microsoft Word fayli, PDF formati yoki ZIP fayl kabi format turiga qarab ma'lum bir kodlash turida taqdim etilishi mumkin.

2. Hujjat tili uning belgilar to'plamini va kodlashni belgilaydi.

Hujjat xitoy tili kabi ma'lum bir tilda yozilganda, u ingliz tilida yozilgan holatga qaraganda boshqa belgilar to'plamidan foydalanadi. Ingliz va boshqa ko'plab Yevropa tillari lotin belgilar to'plamiga asoslangan. Ushbu belgilar to'plamini ASCII (American Standard Code for Information Interchange) da oson ko'rsatish mumkin. Hamma kodlash tizimlari hamma belgilar to'plamlarini qamrab ololmaydi.

Unicode bu Unicode Consortium tomonidan yaratilgan standart kod. Unda har bir belgi unikal identifikator bilan ifodalanadi. Unicode-da turli tillardan bizga ma'lum bo'lgan deyarli barcha belgilar taqdim etilishi mumkin. Shu sababli Unicode barcha tillarni namoyish qilish uchun standart to'plam hisoblanadi. Unicode ning turli o'lchamli ko'rinishlari mavjud bo'lib, UTF-8 bir baytdan foydalanadi va ASCII uchun mos keladi, UTF-16 ikki baytda foydalanadi va ko'p hollarda Osiyo tillari uchun mos keladi [85; 241-256-b., 86; 61-75-b.].

Matn ma'lumotlari uchun vektor fazosini shakllantirishda so'zlar

sumkasi modeli oilasiga tegishli modellar an'anaviy (sanash asosidagi) modellar deyiladi. Bunga termin chastotalari, TF-IDF (termin chastotasiga teskari hujjat chastotasi), N -gramm, mavzu modellari va boshqalar kiradi. Ular matndan belgilarni ajratib olishning samarali usullari bo'lishiga qaramay, modelning o'ziga xos xususiyati shunchaki tuzilmagan so'zlar to'plami bo'lgani uchun, har bir matnli hujjatdagi yaqin so'zlar atrofidagi semantika, tuzilish, ketma-ketlik va kontekst kabi qo'shimcha ma'lumotlar yo'qotiladi. Ushbu jihatlarga e'tibor qaratadigan yanada zamonaviy modellar mavjud va ularni ushbu bobning keyingi qismida ko'rib chiqiladi. An'anaviy modellar matematik va statistik metodologiyalar yordamida qurilgan. Bu strukturalanmagan matn uchun eng oddiy vektor fazosini ifodalash modelidir.

Vektor fazosi modeli - bu strukturalanmagan matnni (yoki boshqa har qanday ma'lumotlarni) raqamli vektorlar sifatida ifodalash uchun matematik model bo'lib, vektorning har bir o'lchami o'ziga xos belgini ifodalaydi.

Masalan, ijtimoiy tarmoqlardagi foydalanuvchilarning xabarlari (yozishmalari) ko'rinishida matnli ma'lumotlar tanlanmasi olingan bo'lsin. Har i -xabar $X_i = (x_1, x_2, \dots, x_{n_i})$ kabi so'zlar ketma-ketligida ifodalangan, bu yerda x_{n_i} - i -xabardagi so'zlar soni. Avvalo, ijtimoiy tarmoqdagi matnlar tanlanmasida unikal so'zlar lug'ati L tuzib olingan bo'lishi kerak. Shunda, $|L|$ - lug'atdagi

unikal soʻzlar soni boʻladi. Shunday holatda, har bir i -xabarlarni sonli vektor koʻrinishida ifodalash mumkin. Bunda A_i vektor $|L|$ oʻlchamda boʻlib, undagi har bir A_{ij} komponent X_i xabardagi L lugʻatga mansub x_j soʻzlar soniga mos keladi. Shunday qilib, formal shaklda vektor $A_i = (A_{i1}, A_{i2}, \dots, A_{i|L|})$ koʻrinishida ifodalanadi.

Soʻzlar sumkasi modeli har bir matnli hujjatni raqamli vektor sifatida ifodalaydi, bu yerda har bir oʻlchov korpusning oʻziga xos soʻzidir va qiymat hujjatdagi chastota, vujudga kelishi (1 yoki 0 bilan belgilanadi) yoki hatto oʻlchangan qiymatlar boʻlishi mumkin. Modelning nomi shundan iboratki, har bir hujjat soʻzma-soʻz, tartib va grammatikani inobatga olmay, oʻz soʻzlari sumkasi sifatida taqdim etiladi [2; 97-108-b.].

Katta hajmli korpus bilan ishlaganda, “soʻzlar sumkasi” modeli bilan bogʻliq baʼzi muammolar boʻlishi mumkin. Belgi vektorlari absolyut termin chastotalariga asoslanganligi sababli, barcha hujjatlar boʻylab tez-tez uchraydigan baʼzi atamalar boʻlishi mumkin va ular xususiyatlar toʻplamidagi boshqa atamalarga soya qilishga moyil boʻlishi mumkin. Ayniqsa tez-tez uchraydigan soʻzlar, ammo maʼlum toifalarni aniqlash uchun belgilar sifatida yanada qiziqarli va samarali boʻlishi mumkin. Bunday holatda TF-IDF samarali boʻladi. TF-IDF bu termin chastotasiga teskari hujjat chastotasini anglatadi. Bu ikkita oʻlchov, termin chastota (tf) va teskari hujjat chastotasi (idf) kombinatsiyasi. Ushbu uslub dastlab foydalanuvchi

soʻrovlari asosida qidiruv tizimlarini natijalarini reytinglashtirish sifatida ishlab chiqilgan va maʼlumot qidirish va matnni ajratib olishning bir qismi boʻlgan.

(1.2.2)da TF-IDF mexanizmining matematik ifodasida d matnli hujjat va t soʻzlar (termlar) qaraladi, TF(term frequency) – t soʻzlarning d hujjatga mansubligining sonli munosabatini bildiradi, yaʼni:

$$TF(t, d) = \frac{d \text{ matnli hujjatda } t \text{ soʻzlar soni}}{d \text{ matnli hujjatdagi soʻzlar soni}}$$

IDF(invers document frequency) – hujjatlar sonining t soʻzlar uchragan hujjatlar soniga munosabatining inversiyasidir:

$$IDF(t, D) = \log \frac{D \text{ korpusdagi hujjatlar soni}}{D \text{ korpusda } t \text{ soʻzlar uchragan hujjatlar soni}}$$

Shunday qilib, (1.2.2) formulani parametrik koʻrinishida quyidagicha

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (2.2.1)$$

matematik tavsiflanadi. Agar soʻzning TF-IDF qiymati qancha katta boʻlsa, shu soʻz matnning kalit soʻzi hisoblanadi. Masalan, X hujjat (masalan, “Jizzax ovozi” gazetasida) matni uchun

$$TF-IDF (“Jizzax”, x) = 0,09,$$

$$TF-IDF (“oila”, x) = 0,16,$$

$$TF-IDF (“ayollar”, x) = 0,21$$

baholardan “oila” va “ayollar” soʻzlari koʻproq uchrashi kalit soʻzlar sifatida aniqlanadi.

Matnli maʼlumotlarda belgilar fazosini shakllantirishning anʼanaviy modellari matndan belgilarni ajratib olishning samarali usullari boʻlishiga qaramay, modelning oʻziga xos xususiyati shunchaki

strukturalanmagan soʻzlar toʻplami boʻlgani uchun, har bir matnli hujjatdagi yaqin soʻzlar atrofidagi semantika, tuzilish, ketma-ketlik va kontekst kabi qoʻshimcha maʼlumotlar yoʻqotiladi. Bu ushbu maʼlumotlarni toʻplash mumkin boʻlgan va yanada keng tarqalgan joylashtirish deb nomlanuvchi soʻzlarni vektorli ifodalash xususiyatlarini beradigan yanada murakkab modellarni oʻrganish uchun yetarli asos boʻladi. Nutqni yoki tasvirni aniqlash tizimlarida barcha maʼlumotlar allaqachon audio spektr va tasvir piksellerining intensivligi kabi yuqori oʻlchovli maʼlumotlar toʻplamiga kiritilgan boy zich belgili vektorlar shaklida mavjud boʻladi. Shu bilan birga, boshlangʻich matn maʼlumotlari, xususan, “Soʻzlar sumkasi” kabi hisoblash modellari haqida gap ketganda, oʻzlarining identifikatorlariga ega boʻlishi mumkin boʻlgan va soʻzlar oʻrtasidagi semantik aloqani ushlab turmaydigan individual soʻzlar bilan ishlanadi. Bu matnli maʼlumotlar uchun siyrak soʻzli vektorlarni keltirib chiqaradi.

Soʻzlar sumkasi modelining kamchiliklarini bartaraf etish uchun vektor fazoli modellardan foydalaniladi. Shunda soʻz vektorlarini shu uzluksiz vektor makoniga semantik va kontekst boʻyicha oʻxshashlik asosida joylashtiriladi. Aslida, taqsimot semantikasi sohasidagi taqsimot gipotezasi bir xil kontekstda yuzaga keladigan va ishlatiladigan soʻzlar semantik jihatdan bir-biriga oʻxshash va oʻxshash maʼnolarga ega ekanligini aytadi. Oddiy soʻzlar bilan aytganda, “soʻzni oʻzi saqlab turgan kompaniya

bilan tavsiflanadi”. Kontekstli soʻz vektorlari uchun ikkita asosiy usul mavjud. *Latent Semantical Analysis* (LSA) kabi hisoblash usullaridan korpusda qoʻshni soʻzlar bilan qoʻshilib kelgan soʻzlarning statistik oʻlchovlarini hisoblash va ushbu oʻlchovlardan har bir soʻz uchun zich soʻzli vektorlar yaratish uchun foydalanish mumkin. Neyron tarmoqqa asoslangan til modellari kabi bashoratli usullar qoʻshni soʻzlardan korpusdagi soʻzlar ketma-ketligiga qarab soʻzlarni bashorat qilishga harakat qilishdan iborat.

Word2Vec modeli. Ushbu model Google tomonidan 2013 yilda yaratilgan va kontekstual va semantik oʻxshashliklarga ega boʻlgan soʻzlarning yuqori sifatli, taqsimlangan va uzluksiz zich vektorli ifodalarini hisoblash va yaratish uchun chuqur oʻrganishga asoslangan bashoratli model hisoblanadi. Aslida, bu katta matnli korpusni egallashi mumkin boʻlgan soʻzlarning lugʻatini yaratadigan va ushbu lugʻatni ifodalovchi vektor fazosidagi har bir soʻz uchun zich soʻz birikmalarini yaratadigan modellar hisoblanadi. Odatda, soʻzni kiritish vektorlarining hajmini belgilash mumkin va vektorlarning umumiy soni asosan lugʻat hajmiga toʻgʻri keladi. Bu zich vektor boʻshligʻining oʻlchamliligini anʼanaviy soʻzlar sumkasi modellariidan foydalangan holda qurilgan yuqori oʻlchamli siyrak vektor fazosidan ancha past qiladi. Ushbu soʻzlarni oʻz ichiga olgan joylashtirishlarni yaratish uchun Word2Vec yordamida ikkita turli xil

arxitektura mavjud. Bularga quyidagilar kiradi:

Continuous Bag of Words (CBoW) modeli. Continuous Bag of Words (CBoW) modeli tabiiy tillarni qayta ishlashda qollaniladigan neyron tarmoqqa asoslangan modeldir. So'zlarni mustaqil ravishda ko'rib chiqadigan "So'zlar sumkasi" (BoW) modelidan farqli o'laroq, CBOW modeli atrofdagi kontekst so'zlari asosida maqsadli so'zni bashorat qilish orqali so'zlar kontekstini qamrab oladi. CBOW modeli manba kontekst so'zlari asosida joriy maqsadli so'zni bashorat qilishga qaratilgan. U kirish sifatida atrofdagi so'zlarning oynasini oladi va oynaning markazida maqsadli so'zni taxmin qilishni o'rganadi. CBOW modelini katta matn korpusida o'rgatish orqali u so'zlar va ularning konteksti o'rtasidagi munosabatlarni o'rganadi, natijada semantik va sintaktik ma'lumotlarni kodlaydigan so'z vektorlari paydo bo'ladi. O'rganilgan so'z vektorlari matnni tasniflash, hissiyotlarni tahlil qilish va mashina tarjimai kabi turli xil NLP vazifalari uchun xususiyat sifatida ishlatilishi mumkin.

Skip-Gram modeli. Skip-Gram - bu bir neyron tarmoq tushunchasi, ko'p yoki ko'p qatlamli so'zlar orasidagi o'zaro bog'liqlikni aniqlash uchun ishlatiladigan algoritm. Bu model,

so'zlar orasidagi bog'liqlikni aniqlashda yaxshi natijalarni ko'rsatadi va ko'p maqsadli so'zlar orasidagi mantiqiy bog'liqliklarni o'rganishda yordam beradi. Skip-Gram modeli, so'zlar orasidagi o'zaro bog'liqliklarni aniqlash va bu bog'liqliklarni o'rganishda yaxshi natijalarni ko'rsatadi. Bu model, muhim matnlar tahlilini osonlashtirish uchun keng qo'llaniladi va ko'p maqsadli so'zlar orasidagi mantiqiy bog'liqliklarni o'rganishda yordam beradi.

Xulosa. Matnli ma'lumotlarga dastlabki ishlov berish - bu matnni tahlilga tayyorlashdagi muhim bosqich bo'lib, u ma'lumot sifati va keyingi tahlil natijalariga bevosita ta'sir ko'rsatadi. Klassik yondashuvlar, jumladan TF-IDF va so'zlar sumkasi modellari matnni matematik jihatdan ifodalash imkonini bersa, zamonaviy neyron tarmoqlarga asoslangan modellar esa kontekst va semantikani saqlab qolish orqali yanada chuqur tahlilni amalga oshiradi. Tadqiqot natijalari shuni ko'rsatadiki, matnli ma'lumotlarga dastlabki ishlov berishda algoritmning to'g'ri tanlanishi tahlilning aniqligi va samaradorligini oshiradi. Kelajakda bu sohada yanada optimallashtirilgan va kontekstga mos modellarni ishlab chiqish dolzarb vazifalardan biri bo'lib qolmoqda.

Foydalanilgan adabiyotlar ro'yxati:

7. A.I.Saleh, A.E.Abulwafa, and M.F.Al Rahmawy, "A web page distillation strategy for efficient focused crawling based on optimized Na{\'i}ve bayes (ONB) classifier," *Applied Soft Computing*, vol. 53, pp. 181-204, 2017.

8. Бабомурадов О.Ж., Қўйлиева Ф.А., Туракулов О.Х. "Clustering approaches to text documents" *The scientific journal vehicles and roads*, 2022 №1

Научный журнал транспортных средств и дорог, 2022 №1 2 НАУЧНЫЙ ЖУРНАЛ
ТРАНСПОРТНЫХ СРЕДСТВ И ДОРОГ Издаётся с 2021 год 109-114б.

9. G.Karystianis, K.Thayer, M.Wolfe, and G.Tsafnat, "Evaluation of a rule-based method for epidemiological document classification towards the automation of systematic reviews," *Journal of biomedical informatics*, vol. 70, pp. 27–34, 2017.

10. O.J.Babomuradov, O.X.Turakulov "Matnni intellektual tahlillash yondashuvlari", *O'zbekiston Respublikasi fanlar akademiyasi O'zbekiston respublikasi aloqa, axborotlashtirish va telekommunikatsiya texnologiyalari davlat qo'mitasi Informatika va energetika muammolari O'zbekiston jurnali* 4 son 2016 y., 68-72 b.

11. O.J.Babomuradov, O.X.Turakulov "Matn belgilari o'lchovini qisqartirish hamda tabiiy til matnini tahlillash yondashuvi" *Innovative Developments And Research In Education, International scientific-online conference, Part 24 December 23 rd colletions of scientific works Canada, 2023.*,320-p.

12. O.J.Babomuradov, N.S.Mamatov, L.B.Boboyev, B.I.Otaxonova, "Text documents classification in Uzbek language," *International journal of recent technology and engineering*, vol. 8, no. 2, pp. 39-43, 2019.

Ingliz tili muharriri: Nosirova Nargiza Jamoliddin qizi
Musahhih: Vahobova Marfua Mirkamalovna
Sahifalovchi va dizayner: Zoirov Sardor Ziyodin o'g'li

1-maxsus son

© Materiallar ko'chirib bosilganda **"Management and Economics Scientific Research Journal"** jurnali manba sifatida ko'rsatilishi shart. Jurnalda bosilgan material va reklamalardagi dalillarning aniqligiga mualliflar ma'sul. Tahririyat fikri har vaqt ham mualliflar fikriga mos kelamasligi mumkin. Tahririyatga yuborilgan materiallar qaytarilmaydi.

Mazkur jurnalda maqolalar chop etish uchun quyidagi havolalarga maqola, reklama, hikoya va boshqa ijodiy materiallar yuborishingiz mumkin.
Materiallar va reklamalar pullik asosda chop etiladi.

El. pochta: journals@timeedu.uz

Tel.: [+998 95 651 90 00](tel:+998956519000)

"Management and Economics Scientific Research Journal" jurnali 19.09.2024-yildan O'zbekiston Respublikasi Prezidenti Adminstratsiyasi huzuridagi Axborot va ommaviy kommunikatsiyalar agentligi tomonidan 404368-son reyestr raqami tartibi bo'yicha ro'yxatdan o'tkazilgan. Litsenziya raqami: C-5669639
PNFL: 308906510

Manzilimiz: Toshkent shahar, Yakkasaroy tumani
Shota Rustaveli ko'chasi, 114