

**TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI
HUZURIDAGI ILMIY DARAJALAR BERUVCHI
DSc.09/2025.27.12. T.01.03 RAQAMLI ILMIY KENGASH**

**MUHAMMAD AL-XORAZMIY NOMIDAGI TOSHKENT AXBOROT
TEXNOLOGIYALARI UNIVERSITETI**

HUSANOV SHERZOD ABDIMANONOVICH

**SUN'IY INTELLEKT TEXNOLOGIYALARI ASOSIDA TARMOQ
MA'LUMOTLARINI TAHLIL QILISH MODEL VA ALGORITMLARI**

05.01.11 – Raqamli texnologiyalar va sun'iy intellekt

**TEXNIKA FANLARI BO'YICHA FALSAFA DOKTORI (PhD)
DISSERTATSIYASI AVTOREFERATI**

Toshkent – 2026

**Texnika fanlari bo'yicha falsafa doktori (PhD) dissertatsiyasi avtoreferati
mundarijasi**

**Оглавления автореферата диссертации доктора философии (PhD) по
техническим наукам**

**Contents of dissertation abstract of doctor of philosophy (PhD)
on technical sciences**

Husanov Sherzod Abdimanonovich

Sun'iy intellekt texnologiyalari asosida tarmoq ma'lumotlarini tahlil qilish
model va algoritmlari..... 3

Хусанов Шерзод Абдуманнонович

Модели и алгоритмы анализа сетевых данных на основе технологий
искусственного интеллекта 24

Husanov Sherzod Abdimanonovich

Models and Algorithms for Analyzing Network Data Based on Artificial
Intelligence Technologies..... 47

E'lon qilingan ishlar ro'yxati

Список опубликованных работ

List of published works..... 50

**TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI
HUZURIDAGI ILMIY DARAJALAR BERUVCHI
DSc.09/2025.27.12. T.01.03 RAQAMLI ILMIY KENGASH**

**MUHAMMAD AL-XORAZMIY NOMIDAGI TOSHKENT AXBOROT
TEXNOLOGIYALARI UNIVERSITETI**

HUSANOV SHERZOD ABDIMANONOVICH

**SUN'IY INTELLEKT TEXNOLOGIYALARI ASOSIDA TARMOQ
MA'LUMOTLARINI TAHLIL QILISH MODEL VA ALGORITMLARI**

05.01.11 – Raqamli texnologiyalar va sun'iy intellekt

**TEXNIKA FANLARI BO'YICHA FALSAFA DOKTORI (PhD)
DISSERTATSIYASI AVTOREFERATI**

Toshkent – 2026

Texnika fanlari bo'yicha falsafa doktori (PhD) dissertatsiyasi mavzusi O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi Oliy attestatsiya komissiyasida B2025.3.PHD/T5888 raqam bilan ro'yxatga olingan.

Dissertatsiya Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universitetida bajarilgan.

Dissertatsiya avtoreferati uch tilda (o'zbek, rus, ingliz (rezyume)) Ilmiy kengash veb-sahifasida (www.tuit.uz) va "Ziyonet" axborot ta'lim portalida (www.ziyonet.uz) joylashtirilgan.

Ilmiy rahbar:

Mo'minov Bahodir Boltayevich
texnika fanlari doktori, professor

Rasmiy opponentlar:

Nosirov Xabibullo Xikmatullo o'g'li
texnika fanlari doktori, dotsent

Hamroyev Alisher Shodmonqulovich
texnika fanlari bo'yicha falsafa doktori, k.i.x.

Yetakchi tashkilot:

Nukus davlat texnika universiteti

Dissertatsiya himoyasi Toshkent axborot texnologiyalari universiteti huzuridagi DSc.09/2025.27.12.T.01.03 raqamli Ilmiy kengashning 2026-yil "9" - sentabr soat 14⁰⁰ dagi majlisida bo'lib o'tadi. (Manzil: 100084, Toshkent shahri, Amir Temur ko'chasi, 108-uy. Tel.: (99871) 207-59-42, e-mail: ilmiy_kengash@tuit.uz).

Dissertatsiya bilan Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universitetining Axborot-resurs markazida tanishish mumkin (445 raqam bilan ro'yxatga olingan). (Manzil: 100084, Toshkent shahri, Amir Temur ko'chasi, 108-uy. Tel.: (99871) 238-64-15).

Dissertatsiya avtoreferati 2026-yil "28" - avgustda tarqatildi.
(2026-yil "28" - avgustdagi 24-raqamli reestr bayonnomasi.)



M.M.Kamilov

Ilmiy darajalar beruvchi ilmiy kengash raisi, texnika fanlari doktori, professor, O'zFA akademigi

N.A.Egamberdiyev

Ilmiy darajalar beruvchi ilmiy kengash ilmiy kotibi, texnika fanlari bo'yicha falsafa doktori

N.O.Raximov

Ilmiy darajalar beruvchi ilmiy kengash qoshidagi ilmiy seminar raisi, texnika fanlari doktori, dotsent

R. Nodir

KIRISH (falsafa doktori (PhD) dissertatsiyasining annotatsiyasi)

Dissertatsiya mavzusining dolzarbligi va zarurati. Jahonda axborot resurslarining asosiy qismini veb muhitda shakllanayotgan hujjatlar tashkil etmoqda. Global internet tarmog'ida har kuni milliardlab yangi veb-sahifalar yaratiladi, mavjudlari yangilanadi yoki modifikatsiyalanadi. Ushbu jarayon natijasida veb muhit ulkan hajmdagi, turli format va tuzilishga ega bo'lgan murakkab axborot tizimiga aylanib bormoqda. Veb hujjatlar o'z tabiatiga ko'ra faqat matndan iborat bo'lib qolmasdan, balki rasm, video, audio, giperhavolalar, skriptlar hamda metama'lumotlar bilan boyitilgan ko'p qatlamli strukturani tashkil etadi. Shu sababli ularni sun'iy intellekt texnologiyalar asosida kompleks tahlil qilish, o'zaro bog'liqligini aniqlash hamda o'xshashlik darajasini baholash dolzarb ilmiy-amaliy muammolardan biri hisoblanadi.

Jahonda turli maqsadlarga mo'ljallangan axborot texnologiyalarning nazariy asoslarini ishlab chiqishda sun'iy intellektni qo'llab tarmoq ma'lumotlarining veb sayt va sahifalarini tahlil qilish model va algoritmlarini ishlab chiqishga yo'naltirilgan ilmiy-tadqiqot ishlari olib borilmoqda. Ma'lumotlarni klasterlash va intellektual tahlil tizimlarida keng qo'llanilmoqda. Xususan, qidiruv tizimlari foydalanuvchiga eng mos natijalarni taqdim etishda sahifalar o'rtasidagi semantik yaqinlikni aniqlashga tayanadi. Shu bilan birga, plagiatni aniqlash tizimlarida matnning yuzaki o'xshashligi emas, balki uning semantik va strukturaviy jihatdan qay darajada mos kelishi muhim ahamiyat kasb etadi. Mavzuning dolzarbligi shundan iboratki, mavjud yondashuvlarda ko'plab veb hujjatlar matnli elementlar asosida baholanadi, biroq veb sahifa tuzilmasining semantik va obyektiv komponentlari (HTML teglari, atributlar, DOM daraxtlari, giperhavolalar tarmoqlari, grafik obyektlar) e'tiborga olinmaydi. Natijada, veb hujjatlar o'xshashligi haqidagi qarorlar noaniq yoki yetarlicha ishonchli bo'lmaydi. Shu nuqtai nazardan, veb hujjatlarning o'xshashligini faqat matn asosida emas, balki ularning strukturaviy, semantik va obyektga yo'naltirilgan xususiyatlarini kompleks hisobga olgan holda baholovchi yangi algoritmlar va matematik model ishlab chiqishga e'tibor berilmoqda.

Respublikamizda ham mazkur yo'nalishda suniy intellekt texnologiyalar asosida tarmoq va veb hujjat ma'lumotlariga ishlov berish, ijtimoiy-iqtisodiy sohalarida amaliyotga joriy etish yuzasidan keng qamrovli chora-tadbirlar amalga oshirilmoqda. Mamlakatimizda milliy iqtisodiyotning turli jabhalarida samaradorlikni oshirish va innovatsiyalarni qo'llab-quvvatlashga alohida e'tibor qaratilmoqda. Xususan "Raqamli O'zbekiston-2030" strategiyasi doirasida sun'iy intellekt texnologiyalarini rivojlantirish milliy raqamlashtirish dasturlariga kiritilgan bo'lib, unda "iqtisodiyot tarmoqlarida virtual va to'ldirilgan reallik, sun'iy intellekt, kriptografiya, mashinaviy o'rganish, katta ma'lumotlarni tahlil qilish va "bulutli" hisoblash texnologiyalaridan foydalanish imkoniyatlarini o'rganish va ularni amaliyotga tatbiq etish;" vazifalari belgilab berilgan. Shuningdek "Sun'iy intellekt texnologiyalarini 2030-yilgacha rivojlantirish strategiyasini tasdiqlash" qarorida "katta ma'lumotlar"ni sun'iy intellekt asosida qayta ishlash va tahlil qilish" vazifasi belgilangan. Mazkur vazifalarni amalga oshirish, xususan tashkilotlar va ta'lim

muassasalari veb saytlarini tahlil qilishda, sun'iy intellekt texnologiyalarini joriy etish muhim ahamiyat kasb etadi.

O'zbekiston Respublikasi Prezidentining 2025-yil 22-oktyabrdagi PF-189-son "Sun'iy intellekt texnologiyalarini yanada rivojlantirishga oid qo'shimcha chora tadbirlari to'g'risida"gi Farmoni, 2024-yil 14-oktyabrdagi PQ-358-son "Sun'iy intellekt texnologiyalarini 2030-yilga qadar rivojlantirish strategiyasini tasdiqlash to'g'risida"gi, 2021-yil 17-fevraldagi PQ-4996-son "Sun'iy intellekt texnologiyalarini jadal joriy etish uchun shart-sharoitlar yaratish chora-tadbirlari to'g'risida"gi, hamda Vazirlar Mahkamasining 2025-yil 10-iyuldagi 425-son "2025-2026-yillarda sun'iy intellekt texnologiyalari sohasida ustuvor loyihalarni amalga oshirish chora-tadbirlari to'g'risida" qarorlari va boshqa mazkur faoliyatga tegishli meyoriy-huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishda ushbu dissertatsiya tadqiqoti muayyan darajada xizmat qiladi.

Tadqiqotning Respublika fan va texnologiyalari rivojlanishining ustuvor yo'nalishlariga mosligi. Mazkur tadqiqot Respublika fan va texnologiyalari rivojlanishining IV. "Axborotlashtirish va axborot-kommunikatsiya texnologiyalarini rivojlantirish" ustuvor yo'nalishi doirasida bajarilgan.

Muammoning o'rganilganlik darajasi. Xorijda chuqur va mashinali o'qitish texnologiyalarga asoslangan katta hajimli ma'lumotlarga ishlov berish hamda veb hujjatlarni tahlil qilish va baholash bo'yicha ilmiy izlanishlar jadal sur'atlar bilan rivojlanmoqda. Xorijiy olimlar Larry Page va Sergey Brin tomonidan PageRank algoritmi ishlab chiqilgan. Ular bu algoritmnini Stanford Universityda doktorantura davrida yaratgan. PageRank algoritmi keyinchalik Google qidiruv tizimining asosiy ishlash mexanizmlaridan biriga aylangan. Veb hujjatlar mazmunini semantik tahlil qilish sohasida Christopher D. Manning, Hinrich Schütze hamda Daniel Jurafsky tomonidan olib borilgan ilmiy ishlar muhim o'rin tutadi. Ularning ishlari natijasida matnli ma'lumotlarni vektor fazoda ifodalash, semantik o'xshashlikni aniqlash va avtomatik tahlil qilish usullari ishlab chiqilgan.

Mashinali o'rganish va chuqur o'rganish texnologiyalarini rivojlantirishda Geoffrey Hinton, Yoshua Bengio va Yann LeCun kabi olimlarning hissasi katta. Ular tomonidan ishlab chiqilgan neyron tarmoqlar modellari (CNN, RNN, Transformer arxitekturasini) veb hujjatlardagi matn, rasm va boshqa multimodal ma'lumotlarni tahlil qilish imkonini kengaytirdi. Tim Berners-Lee, Ricardo Baeza-Yates, Christopher D. Manning, Andrew Ng, hamda mahalliy tadqiqotchilar tomonidan bir qator ilmiy ishlar amalga oshirilgan.

O'zbekistonda Respublikasida ham mashinali va chuqur o'qitish texnologiyalari yo'nalishida ilmiy izlanishlar olib borayotgan olimlar O'zFA akademiklari T.F.Bekmurodov, M.M.Kamilov, professorlar Sh.X.Fozilov, M.M.Musayev, A.Nishanov, N.S.Mamatov, D.Muhamediyeva, M.Rahmatullayev, B.B.Mo'minov, N.Raximov va boshqa olimlar sun'iy intellekt sohalariga o'z hissalarini qo'shib kelmoqdalar. Shu kabi olimlarning sun'iy intellekt va intellectual hisoblash sohasini rivojlantirishda muhim ilmiy asos vazifasini bajarmoqda.

O'zbekistonda ushbu soha kesimida muayyan ilmiy ishlar olib borilayotgan bo'lsada, bu yo'nalish haliyetarli darajada rivojlangan deb bo'lmaydi. Birinchidan,

ilmiy laboratoriyalar va yuqori hisovlash imkoniyatlariga ega markazlar kamligi katta hajimli ma'lumotlar bilan ishlashni cheklab qo'yadi. Ikkinchidan, lokal ma'lumotlar bazasi va milliy datasetlar yetishmasligi sababli ko'pgina tadqiqotlar xorijiy ma'lumotlar asosida olib borilmoqda, bu esa natijalarni mahalliy sharoitga mosligini pasaytiradi. Uchunchidan, soha bo'yicha mutaxassis kadrlar soni kam, natijada sun'iy intellekt texnologiyalar asosida tarmoq ma'lumotlarining veb sayt va sahifalarini tahlili qilish model va algoritmlari ishlab chiqish bo'yicha ilmiy izlanishlar O'zbekistonda hali to'liq shakllanmagan, yetarli darajada o'rganilmagan.

Tadqiqotning maqsadi. Sun'iy intellekt texnologiyalari asosida tarmoq ma'lumotlarining veb sayt sahifalarini tahlil qilish model va algoritmlarini ishlab chiqishdan iborat.

Tadqiqotning vazifalari:

sun'iy intellekt texnologiyalari asosida veb hujjat tuzilishi tahlilini amalga oshirish va takomillashtirish model va algoritmlarini tahlil qilish;

sun'iy intellekt texnologiyalari asosida veb-sayt hodisalarni tahlil qilish va loyihalashtirishga yo'naltirilgan gibridd algoritmini ishlab chiqish;

sun'iy intellekt texnologiyalari asosida veb hujjatning tuzilishini klasterlash algoritmlarini takomillashtirish;

veb hujjat obyektlari o'xshashligi asosida baholashning matematik modeli va algoritmini ishlab chiqish;

produksion bilimlar asosida obyekt-aloqa modelini va boshqarish algoritmini takomillashtirish;

veb hujjatga intellektual tahlil qilish va baholash dasturiy vositasini loyihalashtirish.

Tadqiqotning obyekti. Veb hujjatlar va ularning strukturali komponentlari (HTML elementlari, meta ma'lumotlar, giperhavolalar, kontent bloklari).

Tadqiqotning predmeti. Veb hujjat obyektlari o'xshashligini matematik modellashtirish, ularning semantik va strukturaviy aloqadorligini aniqlash algoritmlari.

Tadqiqotning usullari. Tadqiqotda matematik modellashtirish, klasterlash, Jaccard o'xshashlik koeffitsienti, kosinus o'xshashligi, hamda mashinali o'rganish va tabiiy tilni qayta ishlash (NLP) usullari qo'llanilgan. Shuningdek, Python dasturlash tili, BeautifulSoup, Scikit-learn, TensorFlow kutubxonalarini hamda GPT texnologiyalaridan foydalanilgan.

Tadqiqotning ilmiy yangiligi quyidagilardan iborat:

sun'iy intellekt texnologiyalari asosida veb hujjat hodisalarni loyihalashtirishga yo'naltirilgan gibridd algoritmi ishlab chiqilgan;

veb hujjatlarning tuzilishi asosida obyektlarini orasidagi o'zaro o'xshashligini aniqlashga yo'naltirilgan metrikalarni tanlash algoritmlari ishlab chiqilgan;

veb hujjat obyektlarining hodisalari asosida o'zaro o'xshashligini baholashning matematik modeli va algoritmi ishlab chiqilgan;

produksion qoidalarini asosida bilimlarni loyihalash obyekt-aloqa modeli va boshqarish algoritmi ishlab chiqilgan.

Tadqiqotning amaliy natijalari quyidagilardan iborat:

tarmoq ma'lumotlarini veb hujjatning tuzilishi asosida sinflashtirish modeli va algoritmlarini ishlab chiqildi obyektlarni o'zaro sinflashtirish amalga oshirilgan;

veb hujjatni havollari asosida baholash algoritmlarini ishlab chiqilib amaliy natijalari tadbiq etilgan;

hodisalarni loyihalashtirish nazoratiga mo'ljallangan gibrid algoritmi ishlab chiqildi veb hujjatdagi barcha hodislar o'rganilib ularga qayta ishlov berilgan;

veb hujjatning tuzilishi asosida sinflashtirish va baholash dasturiy vositasini loyihalashtirilgan.

Tadqiqot natijalarining ishonchliligi. Izlanishlarning sun'iy intellektga asoslangan usul va vositalaridan foydalangan holda o'tkazilganligi, tadqiqot apparat-dasturiy vositasi yordamida amalga oshirilgan nazariy va eksperimental tadqiqotlar natijalarining o'zaro mutanosibligi sun'iy intellekt texnologiyalar asosida veb hujjatlarni tahlil qilish uchun ishlab chiqilgan model va algoritmlari tashkilot veb saytlarida o'tkazilgan tajriba-sinovlarining ijobiy natijalari, tegishli guvohnomalar va tadqiqot natijalarini amaliyotga joriy etilganligi bilan asoslanadi.

Tadqiqot natijalarining ilmiy va amaliy ahamiyati. Tadqiqotning ilmiy ahamiyati, unda veb-hujjatlar va tarmoq ma'lumotlarini tahlil qilish jarayonida sun'iy intellekt texnologiyalaridan foydalanishning nazariy va amaliy asoslari doirasida veb hujjatlarning tuzilishi asosida obyektlarini orasidagi o'zaro o'xshashligini aniqlashga yo'naltirilgan metrikalarni tanlash algoritmlari ishlab chiqildi. Ushbu yondashuvlar obyektlar o'xshashligini inobatga olgan holda klasterlash jarayonini amalga oshirib yangi ilmiy yechimlarni taqdim etadi. Shu tariqa tadqiqot klasterlash nazariyasi va mashenali o'qitish usullarini hamda o'xshashliklarni aniqlashda muhim ilmiy ahamiyat kasb etganligi bilan izohlanadi.

Tadqiqot natijalarining amaliy ahamiyati esa taklif etilgan algoritmlar majmuasi veb hujjat ma'lumotlarini tahlil qilish imkoniyatini berishi bilan ifodalanadi. Ishlab chiqilgan algoritmlar asosida dasturiy majmua yaratilgan bo'lib, u amaliy ma'lumotlar tahliliy tizimlariga integratsiya qilish, hisoblash resurslarini tejash va qaror qabul qilish jarayonini tezlashtirish uchun metodik asos bo'lib xizmat qiladi. Bu esa tadqiqot natijalarining amaliyotga yo'naltirilganligini va keng qo'llanish imkoniyatiga ega ekanligi bilan izohlanadi.

Tadqiqot natijalarining joriy qilinishi. Tadqiqotda ishlab chiqilgan veb hujjatga intellektual ishlov berish va baholash dasturiy vositasi asosida:

sun'iy intellekt texnologiyalari asosida veb hujjat hodisalarni loyihalashtirishga yo'naltirilgan gibrid algoritmi asosida ishlab chiqilgan dasturiy vosita O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi "Toshkent davlat iqtisodiyot universiteti"da rasmiy veb saytini tahlili qilish va rivojlantirish maqsadida amaliyotga joriy qilingan (Raqamli texnologiyalar vazirligining 2026-yil 2-iyundagi 33-8/4770-sonli ma'lumotnomasi). Natijada universitet rasmiy veb saytining (texnik, funksional, iqtisodiy, foydalanish) samaradorlik ko'rsatkichlari 1,26 barobarga oshirilgan;

veb hujjatlarning tuzilishi asosida obyektlarini orasidagi o'zaro o'xshashligini aniqlashga yo'naltirilgan metrikalarni tanlash hamda veb hujjat tuzilishini sinflashtirish algoritmlari asosida ishlab chiqilgan dasturiy vosita O'zbekiston

Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi qoshidagi "Raqamli ta'lim texnologiyalarini rivojlantirish markazi"da ishlab chiqilgan axborot tizimlarini tahlili qilish va rivojlantirish maqsadida amaliyotga joriy qilingan (Raqamli texnologiyalar vazirligining 2026-yil 2-iyundagi 33-8/4770-sonli ma'lumotnomasi). Natijada markazda ishlab chiqilgan axborot tizimlarining (texnik, funksional, iqtisodiy, foydalanish) samaradorlik ko'rsatkichlari 1,8 barobarga oshirilgan;

vab hujjat obyektlarining hodisalari asosida o'zaro o'xshashligini baholashning matematik modeli va algoritmi asosida ishlab chiqilgan dasturiy vosita Raqamli texnologiyalar vazirligi qoshidagi "Raqamli hukumat loyihalarni boshqarish markazi"da axborot tizimlarini tahlili qilish va rivojlantirish maqsadida amaliyotga joriy qilingan (Raqamli texnologiyalar vazirligining 2026-yil 2-iyundagi 33-8/4770-sonli ma'lumotnomasi). Natijada markazda ishlab chiqilgan axborot tizimlarining (texnik, funksional, iqtisodiy, foydalanish) samaradorlik ko'rsatkichlari 1,3 barobarga oshirilgan;

produksion qoidalari asosida bilimlarni loyihalash obyekt-aloqa madeli va boshqarish algoritmi hamda vab hujjat tuzilishini sinflashtirish algoritmlari asosida ishlab chiqilgan dasturiy vosita O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi "Toshkent davlat tibbiyot universiteti"da rasmiy vab saytini tahlili qilish va rivojlantirish maqsadida amaliyotga joriy qilingan (Raqamli texnologiyalar vazirligining 2026-yil 2-iyundagi 33-8/4770-sonli ma'lumotnomasi). Natijada universitet rasmiy vab saytining (texnik, funksional, iqtisodiy, foydalanish) samaradorlik ko'rsatkichlari 1,62 barobarga oshirish imkoniyati yaratilgan.

Tadqiqot natijalarining aprobatsiyasi. Mazkur tadqiqot natijalari 2 ta xalqaro va 2 ta respublika ilmiy-amaliy anjumanlarida ma'ruza qilingan va muhokamadan o'tkazilgan.

Tadqiqot natijalarining e'lon qilinganligi. Tadqiqotning asosiy natijalari 14 ta ilmiy ishlarda e'lon qilingan, ulardan 5 tasi O'zbekiston Respublikasi Oliy attestatsiya komissiyasi tomonidan doktorlik dissertatsiyalarining asosiy ilmiy natijalarini e'lon qilish uchun tavsiya qilingan jurnallarda, jumladan 2 ta xorijiy jurnallarda va 2 ta Respublika jurnallarida, 1 ta xalqaro konferensiyada hamda EHM uchun dasturlarni rasmiy ro'yxatdan o'tganligi to'g'risida 2 ta guvohnoma olingan.

Dissertatsiyaning tuzilishi va hajmi. Dissertatsiya ishi kirish, to'rtta bob, xulosa, foydalanilgan adabiyotlar ro'yxatidan iborat. Dissertatsiyaning umumiy hajmi 124 betni tashkil etgan.

DISSERTATSIYANING ASOSIY MAZMUNI

Dissertatsiya mavzusining dolzarbligi va zaruriyati asoslangan, tadqiqotning respublika fan va texnologiyalari taraqqiyotining ustuvor yo'nalishlariga mosligi ko'rsatilgan. Tadqiqotning maqsad va vazifalari belgilab olingan hamda tadqiqot obyekti va predmeti aniqlangan, olingan natijalarning ishonchligi asoslab berilgan, ularning nazariy va amaliy ahamiyati ko'rsatilgan, tadqiqot natijalarini amaliyotga joriy qilish holati, nashr etilgan ishlar va dissertatsiya tuzilishi bo'yicha ma'lumotlar keltirilgan.

Dissertatsiyaning "Tarmoq ma'lumotlar tahlilining nazariy asoslari va vositalari" deb nomlangan birinchi bobda vab-hujjatlarni tahlil qilish - elektron

hujjatlarni texnik usulda ko‘rib chiqish yoki baholash jarayoni hisoblanadi. Veb-hujjatlarni tahlil qilish usuli, boshqa ko‘plab tadqiqot usullari singari, ma’noni ochish, tushunish va xulosaga kelish uchun ma’lumotlarga intellektual yoki qisman intellektual ishlov berish orqali tekshirish va sharhlash muammolari o‘rganilib, ularni bartaraf etish yo‘llari va tadqiqot ishi masalasining qo‘yilishi keltirilgan.

1.1-paragrafda qaror qabul qilishga ko‘maklashuvchi tizimlarning predmet soxalardagi ahamiyati yoritilgan bo‘lib, ularning qo‘llanilishi va ishlab chiqishga oid har bir bosqichga mos usul va algoritmlar tadqiq qilingan. Bundan tashqari, jahonda veb hujjatlar tahlilida qo‘llanilib kelayotgan intellektual tizimlar haqida ma’lumotlar keltirib o‘tilgan.

1.2-paragrafda veb hujjat kontentlarga yo‘naltirilgan veb analitika modellari va algoritmlarining asosiy tushuncha va ta’riflari keltirilgan bo‘lib, unda qoida, bilimlar bazasi xususan veb-saytlarda foydalaniladigan fayl tuzilmalari haqida batafsil ma’lumotlar berilgan. Shuningdek, intellektual tizimlar yaratishning zamonaviy yondashuvlari haqida ma’lumot keltirilgan.

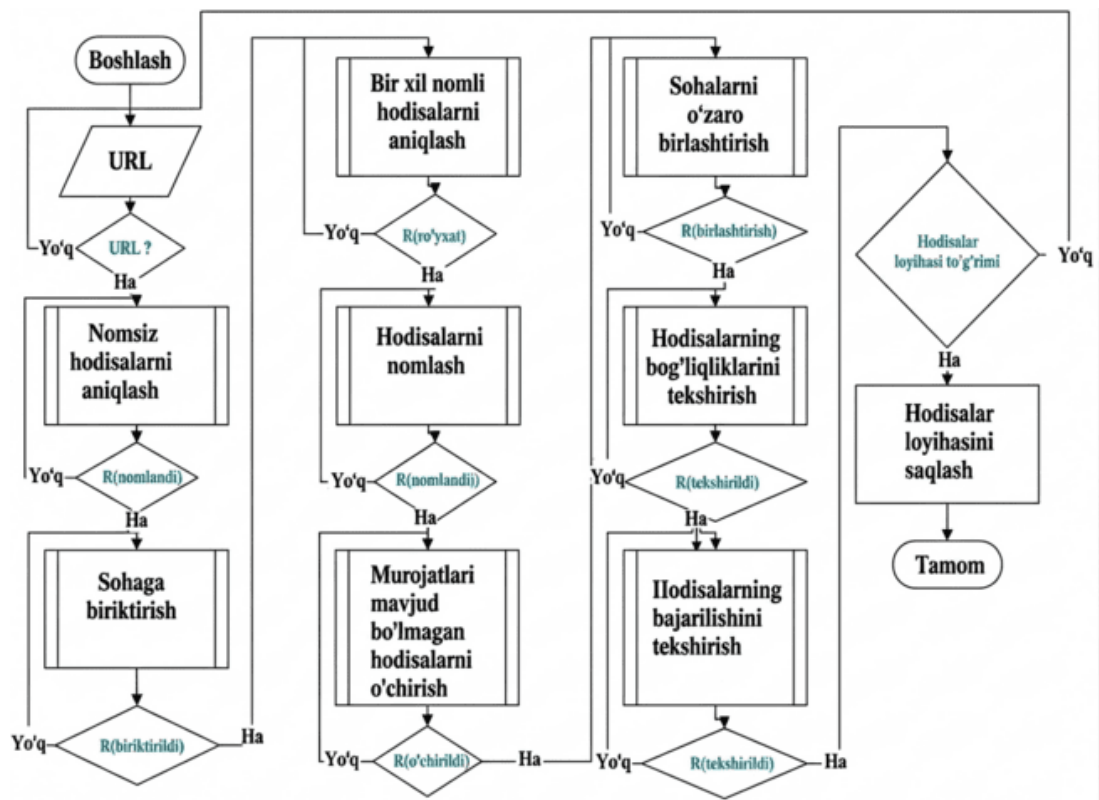
1.3-paragrafda tarmoq ma’lumotlariga intellektual ishlov berish, mashinaviy o‘qitish, tabiiy tilni qayta ishlash (NLP), graf nazariyasiga asoslangan sun’iy intellekt modellaridan foydalanish hamda mashinaviy o‘qitish algoritmlari yordamida veb hujjatlar o‘rtasidagi yashirin bog‘liqliklarni aniqlash, ularni avtomatik sinflashtirish va klasterlash imkoniyati, nazoratsiz o‘qitish algoritmlari, xususan klasterlash usullari to‘g‘risidagi ma’lumotlar keltirilgan.

Dissertatsiyaning **“Veb hujjat tuzilishi va tahliliga yo‘naltirilgan modellar va algoritmlar”** deb nomlanuvchi ikkinchi bobi 5 ta paragrafdan iborat bo‘lib, unda Veb hujjating tuzilishi yaxshilash va tahlilni amalga oshirish uchun veb-metodlar yordamida saytni o‘rganish, rejalashtirilgan tartib-protsedura asosida sayt tuzilishi bo‘yicha ma’lumotlarni to‘plash, veb-grafikani qurish va tahlil o‘tkazish, ehtiyoj mezonini ishlab chiqish, mezon bo‘yicha tuzilmani yaxshilaydigan nazorat qoidalarini aniqlash va sayt tuzilmasini takomillashtirish bo‘yicha tavsiya va takliflarni ishlab chiqishni o‘z ichiga oladi.

Natijada, sun’iy intellekt texnologiyalariga asoslangan vebometrika modellari veb hujjatlarning tuzilishini takomillashtirish, ularni avtomatik baholash va tarmoq ma’lumotlarini chuqur tahlil qilishda samarali ilmiy-amaliy yechim sifatida asoslab berilgan.

Kompyuter tarmoqlarida foydalaniladigan veb hujjatlarda 8 turdagi hodisalar loyihalandi.

Bundan veb hujjatdagi barcha hodisalar o‘zaro bog‘langan va maxsus loyihalashtirishni nazoratga olish algoritmi yoki yondashuvini talab qiladi. Veb hujjatlarda hodisalarni to‘g‘ri loyihalash uning ish sifatini oshiradi va kontent ma’lumotlarining to‘g‘ri saralanishini ta’minlab beradi. Hodisalarni loyihalash nazorati amalga oshirish quyidagicha algoritm quydagi 1-rasmda tavsiya qilinadi.



1-rasm. Hodisalarni loyihasini nazorat qilishning umumiy algoritmi.

Veb saytlar va veb hujjatlarning tarmoqdagi ahamiyatini aniqlashga yo'naltirilgan PageRank reyting algoritmini takomillashtirish masalalari ko'rib chiqilgan. Internet muhitida veb sahifalar o'zaro giperhavolalar orqali bog'langan murakkab tarmoqni hosil qiladi. Bunday tarmoqni samarali tahlil qilish va sahifalarning nufuzini aniqlash graf nazariyasiga hamda iterativ hisoblashga asoslangan intellektual yondashuvlarni talab etadi.

Tadqiqotda veb sahifalar tugunlar, ularning o'zaro havolalari esa yo'naltirilgan qirralar sifatida qaralib, veb muhit yo'naltirilgan graf modeli ko'rinishida ifodalangan. Ushbu graf asosida sahifalarning nisbiy ahamiyatini aniqlashda PageRank algoritmi qo'llanilgan. Algoritmning asosiy g'oyasi shundan iboratki, sahifaning nufuzi unga ishora qilayotgan sahifalar soni va ularning o'z nufuziga bog'liq holda aniqlanadi.

PageRank algoritmining an'anaviy modeli tahlil qilinib, uning hisoblash murakkabligi va katta hajmdagi tarmoq ma'lumotlari bilan ishlashdagi cheklavlari ko'rsatib berilgan. Shu asosda, PageRank reytingini samarali hisoblashga yo'naltirilgan takomillashtirilgan algoritmik yondashuv taklif etilgan. Taklif etilgan usul iterativ hisoblash jarayonini optimallashtirish, ortiqcha hisob-kitoblarni kamaytirish hamda katta hajmdagi veb sahifalar to'plamida tezkor va barqaror natijalarga erishishni ta'minlaydi.

Natijada, taklif etilgan PageRank reytingini samarali hisoblash algoritmi veb saytlar va veb hujjatlarni tarmoq muhitida yanada aniq baholash, qidiruv tizimlari, veb analitika va tarmoq ma'lumotlarini tahlil qilish masalalarida qo'llashda ilmiy-amaliy jihatdan asoslab berilgan.

Veb hujjatning tuzilishi va havollalari asosida klasterlash modellarini tanlash algoritmini yaratishni ikkita alohida algoritm asosida tadqiq qilinadi. Chunki veb

hujjatning tuzilish har bir sahifada mavjud va uni klasterlashda alohida yondashuvni amalga oshirish lozim. Veb hujjatning havolalari har bir sahifa bilan bog‘liq.

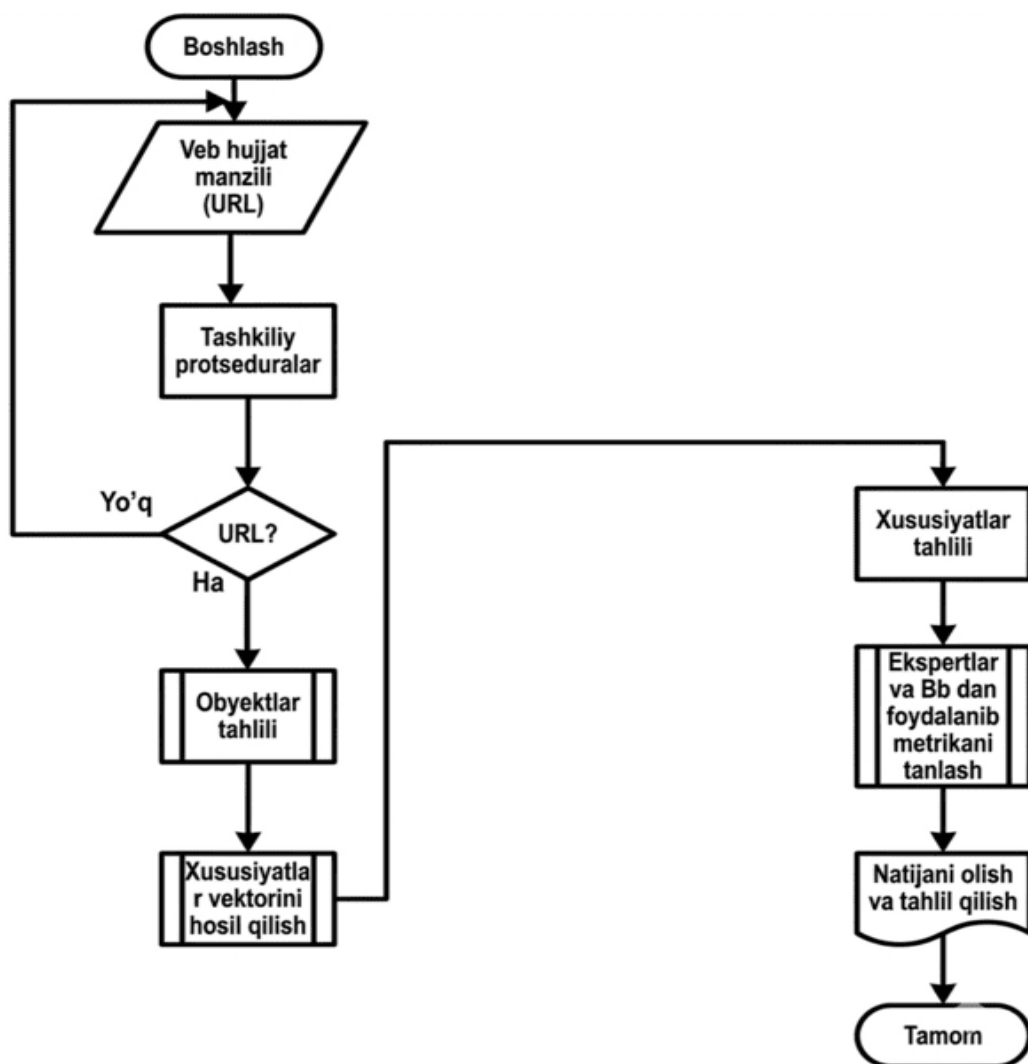
Birinchi yondashuv. Berilgan obyektlarning mos ravishda ikkitasining xos hususiyatlari bo‘yicha o‘zaro o‘xshashlikni o‘lchash metrikalari. Bunda maxsus *BiDist* kutubxonasi yaratiladi va uning asosiy funksiyasi quyidagicha:

$$d = BiDist(x, y, metric_{name}, outType, rw)$$

Bunda x, y – obyektning hususiyatlar vetori, $metric_{name}$ - metrika turi, $outType$ -o‘xshashlik chiqish tipi, rw - sozlamalar.

Foydalanishda metrika usullari yoki masofa funksiyasi sifatida quyidagilarni tanlab olish mumkin: 'braycurtis', 'kanberra', 'chebyshev', 'cityblock', 'korrelyatsiya', 'kosinus', 'dice', 'evklid', 'hamming', 'jaccard', 'jensenshannon', 'kulczynski', 'mahalanobis', 'match', 'minkowski', 'rogerstanimoto', 'russellrao', 'seuclidean', 'sokalmichener', 'sokalsneath', 'sqeuclidean', 'yule'.

Ushbu yuqorida keltirilgan masofa o‘lchash metrikalari kengaytirilgan modellar keng tarqalgan. Veb hujjatlardagi obyektlarni o‘zaro munosabati ularning o‘xshashligini belgilaydi va bu metrikalar asosda aniqlanishi maqsadga muvofiq. Bu matematik ifodalarning hammasini birlashtirib, instrumental kutubxona yaratish lozim.



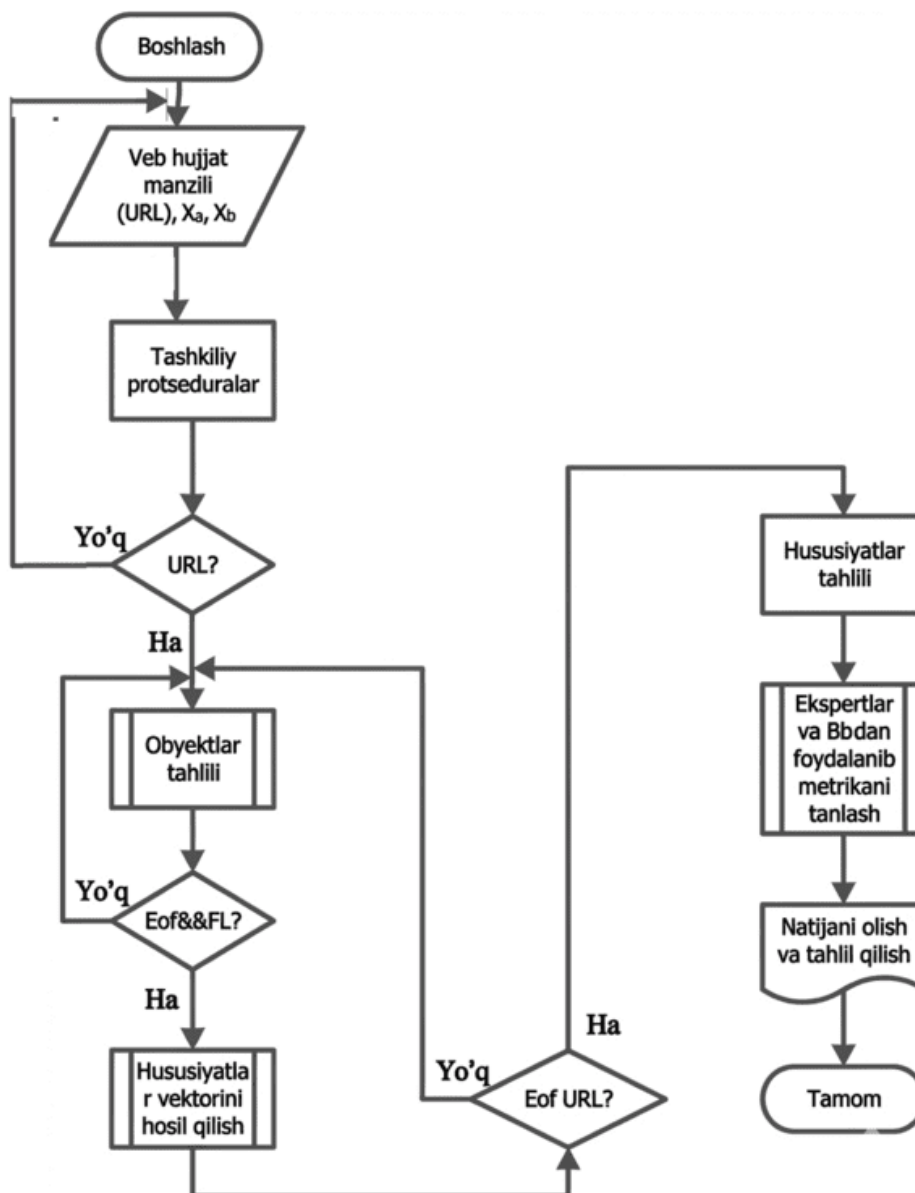
2-rasm. Obyektlarning o‘zaro o‘xshashligi uchun klasterlash algoritmi.

Ikkinchi yondashuv. Berilgan ikki veb hujjatning obyektlarning mos ravishda xos ikki xususiyatlari bo'yicha o'zaro o'xshashlikni o'lchash metrikalari. Bunda maxsus $cDist$ kutubxonasi yaratiladi va uning asosiy funksiyasi quyidagicha:

$$d = cDist(x_a, x_b, metric_{name}, outType, rw)$$

Bunda x_a, x_b – mos ikki obyektning xususiyatlar vektori, $metric_{name}$ - metrika turi, $outType$ - o'xshashlik chiqish tipi, rw - turli sozlamalar.

Ushbu yuqorida keltirilgan masofa o'lchash metrikalari kengaytirilgan modellar keng tarqalgan. Bu matematik ifodalarning hammasini birlashtirib, instrumental kutubxona yaratish lozim.



3-rasm. Ikki veb hujjatning o'zaro o'xshashligi uchun klasterlash algoritmi.

Ikki veb hujjatning o'zaro obyektlarining hususiyatlarini o'xshashligini klasterlash modeli tanlash algoritmi ikki va undan ortiq veb hujjatlar tuzilmasidagi obyektlarning o'xshashligini hisoblash keltirilgan. Unda ikki veb hujjatning to'liq yo'li ko'rsatiladi va tashkiliy protseduralarga o'tiladi. Agar tashkiliy protseduralarda kamchilik bo'lsa, qaytadan boshlaydi, aks holda veb hujjatdagi obyektlarni aniqlash qism algoritmi ishga tushadi. Obyektlarning taxlilidan so'ng mos ravishda

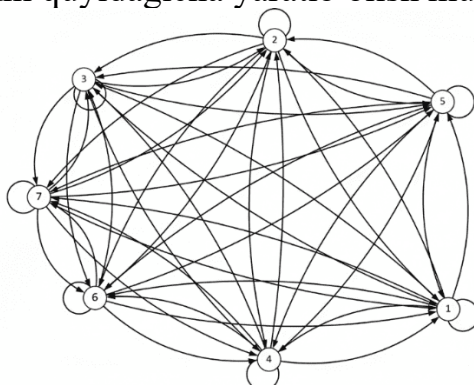
obyektlarga mos xususiyatlar vektori yaratiladi. Bu jarayon veb hujjatlarning soniga mos holda takrorlanadi. Ikki vektordagi hususiyatlar taxlil qilinib, mos mantiqiy vektor hosil qidinadi. Ekspertlar va bilimlar bazasidan foydalanib, hususiyat tipiga mos metrika tanlash amalga oshiriladi. Natijalar ikkilangan matritsa ko'rinishida mos ravishda ikki veb hujjatdagi obyektlarning mos hususiyatlari asosida o'xshashlik qiymatlari hisoblanadi.

Veb hujjat obektlari o'xshashligini hisoblash yuqoridagi paragrafda keltirilganidek, ikkiga bo'lib obyektlarning o'xshashligi asosida baholash modellari alogoritmlarini taklif etiladi.

Bir veb hujjat obektlari o'xshashliklarini Q matritsi ko'rinishida berilgan bo'lsin $Q = (q_{i,j})$, mos ravishda veb sahifaning i va j obyektlarining o'xshashligidir. Obyektning har bir hususiyati bo'yicha yo'naltirilgan grafning bir tuguni sifatida qaralsa, veb hujjatning obyektlari o'xshashligining graf model qurish mumkin. Yo'naltirilgan grafni $D = (V, E)$ kabi ifodalasak, unda V grafning uchi va bu holda obyektlari grafning uchlari bo'ladi. E – esa tartiblangan to'plam bo'lib, $e_i \in V, i = 1, \dots, n, n$ –obyektlar soni.

Graf uchiga kiriuvchi yoylarni $e_i^I \in V$, chiquvchi yoylarni $e_i^O \in V, i = 1, \dots, n, n$ –obyektlar soni bo'lsin. Bu yoylarning uchi kirish yoki chiqish vaznlari q_{ij} deb, i -obyektning j -obyektiga o'xshashligi olinadi.

Yo'naltirilgan grafni quyidagicha yaratib olish mumkin



4-rasm. Veb hujjat obyektlarining o'xshashligiga asoslangan graf modeli.

Veb hujjat obyektlarining o'xshashligiga asoslangan graf modelda o'xshashlik vaznlari keltirilmagan. Bu vaznlarni matritsasi ko'rinishida quyidagicha ifodalanadi

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} & q_{1,3} & \dots & q_{1,n} \\ q_{2,1} & q_{2,2} & q_{2,2} & \dots & q_{2,n} \\ q_{3,1} & q_{3,2} & q_{3,3} & \dots & q_{3,n} \\ \dots & \dots & \dots & \dots & \dots \\ q_{n,1} & q_{n,2} & q_{n,3} & \dots & q_{n,n} \end{pmatrix}$$

Veb hujjat obyektlarining graf modeli asosida o'xshashlikgi asosida baholash mumkin.

Veb hujjat obyektlarining o'xshashligi matritsasi bu to'liq simmetrik matritsa hisoblanmaydi, ya'ni qiymatlari bir xil emas. Matritsaning qiymatlari asosida e_i^I kiruvchi Q ning qator, ya'ni $(q_{i,1}, q_{i,2}, \dots, q_{i,n})$ va e_j^O chiquvchi Q ning ustuni, ya'ni $(q_{1,i}, q_{2,i}, \dots, q_{n,i})$ yoylarning qiymatlari qabul qilinsin.

Veb sahifadagi obyektning faolligini baholashda uning o'xshashligi matritsasidan quyidagi ifoda asosida xulosalash taklif qilinadi.

$$W_j = \sum_{i=1}^n q_{i,j} - \sum_{i=1}^n q_{j,i}, j = 1..m \quad (1)$$

Yuqoridagi (1) ifodaning qiymatidan quyidagi xossani kiritish lozim

1. Agar $W_j > 0$ bo'lsa, q_j obyektidan boshqa obyektga o'tilgan va bu obyekt veb sahifa uchun o'ta muhim hisoblanadi.

2. Agar $W_j = 0$ bo'lsa, q_j obyektidan boshqa obyektga o'tilgan va boshqa obyektidan bu obyektga kirishlar soni teng, veb sahifa uchun yaxshi (qoniqarli) obyekt hisoblanadi.

3. Agar $W_j < 0$ bo'lsa, boshqa obyektidan q_j obyektga o'tilgan va bu obyekt veb sahifa uchun o'ta qoniqarsiz hisoblanadi. Bu obyektни tezroq takomillashtirish yoki olib tashlash lozim.

Tashkilotning oldiga qo'ygan reja tadbirlariga o'zaro o'xshash obyektlar orqali indekslashtirish va bir vektorga joylashtirish mumkin. Bu esa o'zaro o'xshash obyektlarning veb hujjatdagi reytinglarini aniqlash orqali amalga oshirish mumkin.

Veb hujjatdagi obektlarning reytingini aniqlashda foydalanuvchi shu obyektga murojaatlar soni, ya'ni kelish va ketishlar sonini aniqlash lozim bo'ladi. Bunda veb sahifada kursorning harakatini boshqarish yoki sichqonchaning harakatini boshqarish va uning obyektlari qadamlarini sanashdan iborat. Bunda sichqonchaning quyidagi 5 ta asosiy hodisasiga tayanish mumkin Ularga:

1. MuoseDown – obyektning ta'sir doirasida sichqonchaning x kordinatisi bo'yicha kelish harakati;

2. MouseUp - obyektning ta'sir doirasida sichqonchaning y kordinatisi bo'yicha kelish harakati;

3. MouseMove – obyektning ta'sir doirasida sichqonchaning chiqib ketish xarakati;

4. Click (mouseClick) – obyektning ta'sir doirasida sichqonchaning tugmasi bir marta bosish orqali, tanlanganlik harakati;

5. DbClick (mouseDbClick) - obyektning ta'sir doirasida sichqonchaning tugmasi ikki marta bosish orqali, tanlanganlik harakati, agar bu hodisa bo'lsa, boshqa obyektga o'tish harakatini beradi.

Yuqorida sanab o'tilgan barcha, sichqinchaning harakatlari barcha dasturlash tillari va skript tillari mavjud va loyihaltirilgan. Veb hujjatning har bir sahifasida yozilgan obyektlar bu hodisalarni boshqarish mumkin. Bunda quyidagicha yondashuvlar taklif etiladi.

1. MuoseDown va MouseUp hodisasi har bir obyektga yagona bir identifikator i ($i=0$)e'lon qilinadi. Bu $i++$ amalni real vaqt davomida obyektga bo'lgan hodisada bir marta amalga oshiradi.

2. MuoseMove hodisasi har bir obyektga yagona bir identifikator j ($j=0$)e'lon qilinadi. Bu $j++$ amalni real vaqt davomida obyektga bo'lgan hodisada bir marta amalga oshiradi.

3. Click hodisasi har bir obyektga yagona bir identifikator c ($c=0$) e'lon qilinadi. Bu $c++$ amalni real vaqt davomida obyektga bo'lgan hodisada bir marta amalga oshiradi.

4. DbClick hodisasi har bir obyektga yagona bir identifikator dc ($dc=0$) e'lon qilinadi. Bu $dc++$ amalni real vaqt davomida obyektga bo'lgan hodisada bir marta amalga oshiradi.

Bu o'zgaruvchilar o'zlari veb sahifaning ishga tushirgandan boshlab, to statistik ma'lumotlar olguncha o'z vazifasini bajaradi. Har bir obyektga bu 4 ta hodisalar doimiy ishlab turadi va quyidagi ifoda bilan hisoblanadi.

$$E_I = i + c, \quad E_O = j + dc \quad (2)$$

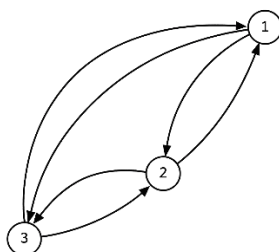
E_I – bu obyektga kiruvchi murojaatlar va E_O – obyektga chiquvchi murojaatlar. Hodisalarning xususiyatlari inobatga olinib ikkiga ajratilgan.

Hodisalarga asoslangan kirish va chiqish murojaatlarni graf model ko'rinishida ifodalash orqali axborot modelini qurish mumkin. Bunda qaraliyotgan joriy obyektning grafning uchi va yo'ylari sifatida kiruvchi, chiquvchi murojaatlarni olish mumkin.

Yo'naltirilgan grafni $G = (O, E)$ kabi ifodalasak, unda O grafning uchi va bu holda obyektlar grafning uchlari bo'ladi. E – esa tartiblangan to'plam bo'lib, mos holda kiruvchi va chiquvchi murojaatlar, $e_i \in V, i = 1, \dots, n, n$ –obyektlar soni.

Graf uchiga kiruvchi murojaat yo'ylarni $e_i^I \in V$, chiquvchi murojaat yo'ylarni $e_i^O \in V, i = 1, \dots, n, n$ –obyektlar soni bo'lsin. Bu yo'ylarning uchi kirish yoki chiqish vaznlari q_{ij} deb, i -obyektning j -obyektiga kirish yoki chiqish murojaati soni olinadi.

Yo'naltirilgan grafni quyidagicha yaratib olish mumkin



5-rasm. Veb hujjat obyektlarining murojaatlariga asoslangan graf modeli.

Umuman olganda bir veb hujjatda n ta veb sahifa va bu sahifalarning har birida m ta obyekt bo'lsin. Jami obyektlar soni $N = nm$ bo'ladi. Veb hujjatning kirish va chiqish murojaatlarini M matritsiya ko'rinishida ifodalash mumkin. Bunda quyidagilar o'rinli bo'lsin:

1. M matritsaning har bir mos qatori mos holda chiqish murojaatlarini bersin.
2. M matritsaning har bir mos ustuni mos holda kirish murojaatlarini bersin.
3. Bu holda M matritsaning diagonal elementlari qiymati $E_D = (E_I + E_O)/2$ ifoda bilan hisoblanadi.

M matritsaning umumiy ko'rinishi quyidagicha ifodalanadi:

$$M = \begin{pmatrix} E_{D_{1,1}} & E_{O_{1,2}} & E_{O_{1,3}} & \dots & E_{O_{1,n}} \\ E_{I_{2,1}} & E_{D_{2,2}} & E_{O_{2,3}} & \dots & E_{O_{2,n}} \\ E_{I_{3,1}} & E_{I_{3,1}} & E_{D_{3,3}} & \dots & E_{O_{3,n}} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ E_{I_{n,1}} & E_{I_{n,2}} & E_{I_{n,3}} & \dots & E_{D_{n,n}} \end{pmatrix}$$

M matritsiyadan har bir E_I^k, E_O^k obyekt uchun quyidagi ifoda o‘rinli

$$E_I^k = \sum_{i=1, i \neq k}^N E_{I_{k,i}}, E_O^k = \sum_{i=1, i \neq k}^N E_{O_{k,i}} \quad k = 1 \dots N \quad (3)$$

(2) ifodadan quyidagi xulosalarga kelish mumkin:

1. Agar har bir k uchun $E_I^k - E_O^k < 0$ bo‘lsa, bu obyektни yangilash yoki o‘zgartirish kiritish kerak.

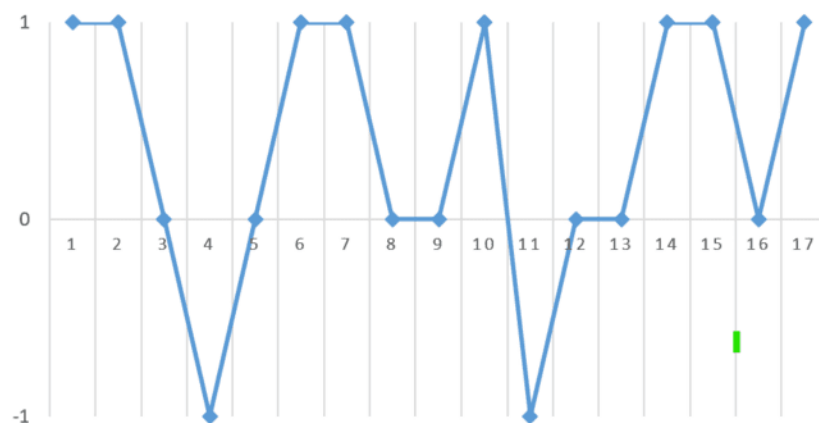
2. Agar har bir k uchun $E_I^k - E_O^k = 0$ bo‘lsa, bu obyekt yaxshi qoniqarli hisoblanadi.

3. Agar har bir k uchun $E_I^k - E_O^k > 0$ bo‘lsa, bu obyekt juda yaxshi va bu kabi obyektlarni ko‘paytirish lozimligini bildiradi.

Yuqoridagi taklif asosida veb hujjat uchun M matritsiyani o‘zgartirilgan M^* varianti kelib chiqadi.

$$M^* = \begin{cases} 1, & E_I^k - E_O^k > 0 \\ 0, & E_I^k - E_O^k = 0 \\ -1, & E_I^k - E_O^k < 0 \end{cases} \quad (4)$$

M^* vektorning qiymatlari asosida veb hujjatning o‘zgartirish talab qilinayotgan obyektlarini vizual ko‘rish imkoniyati yaratiladi.



6-rasm. Veb hujjatning obyektlariga murojaatlar grafigi.

Takomillashtirilgan PageRank reytingini samarali hisoblashda qavariq optimallashtirish va Frank-Vulfaning shartli gradienti usullaridan foydalanilgan. Dasturiy vositasini loyihalashda obyekt va predmetini aniqlab olinadi.

Dasturiy vositaning obyekti: veb-graf va undagi giperhavolalar to‘ri.

Dasturiy vositaning predmeti: veb-sahifalarning PageRank reytingini hisoblashning qavariq optimallashtirish va Frank-Vulfaning shartli gradienti usullari asosida tokomillashtirilgan algoritmlar.

Dasturiy vositaning maqsadi: veb-sahifalarning reytingini tezkor va aniq hisoblash va natijalarni kerakli ma'lumotlar bazasida, mos holda integratsiyani ta'minlashdan iborat.

Dasturiy vositani loyihalashda oldingi paragraflarda keltirilgan algoritmlar va yordamichi algoritmlardan foydalaniladi. Bu quyidagi 4 bosqichda amalga oshiriladi:

1-bosqich. PageRankning modelini tanlash va loyihalashtirish.

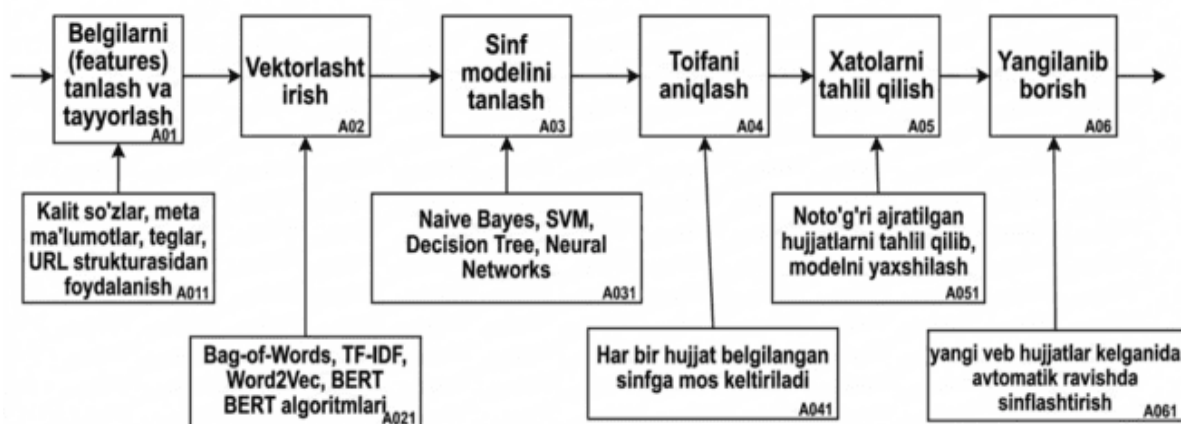
2-bosqich. Iteratsion usullarni (Power Method, Jacobi, Gauss-Seidel) solishtirish.

3-bosqich. Mos parallel algoritm ishlab chiqish.

4-bosqich. Dasturiy vositani zaruriyat bor (Python/Java/C++) muhitda ishlab chiqish.

Har bir bosqichni o'ziga xos yondashuvlar asosida amalga oshirish mumkin.

Tadqiq qilinayotgan veb hujjatlarda tuzilishi asosida sinflashtirishning asosiy maqsadi veb hujjatning turli qismlari keltirilgan. Ularga teglar (title, meta, headings), navigatsiya, menyu, asosiy matn (body), havolalar, DOM daraxti, URL hamda SEO teglar, so'rovlar, tasvirlar va boshqalar kiradi. DOM daraxti asosida sahifaning qismlari chiqarilishi va ularning munosabati asosida sinflashtirish ancha ishonchli bo'ladi.



7-rasm. Sinflashtirishning funksional tuzilmasi.

Klasterlash-veb-hujjatlarni oldindan belgilangan sinflarsiz, o'xshashligi va tuzilishiga ko'ra guruhlarga ajratish jarayoni. Bu orqali bir xil HTML tuzilishga ega sahifalarda yangiliklar, bloglar, reklama, kontakt sahifalari kabi teglarni olish mumkin. Shuningdek bu ma'lumotlarni oldindan belgilanmagan sinflarsiz, o'xshashligi asosida guruhlarga ajratish usuli. Bu usul intellektual ma'lumotlar tahlilida Data Mining, Text Mining, Web Mininglarda keng qo'llaniladi.

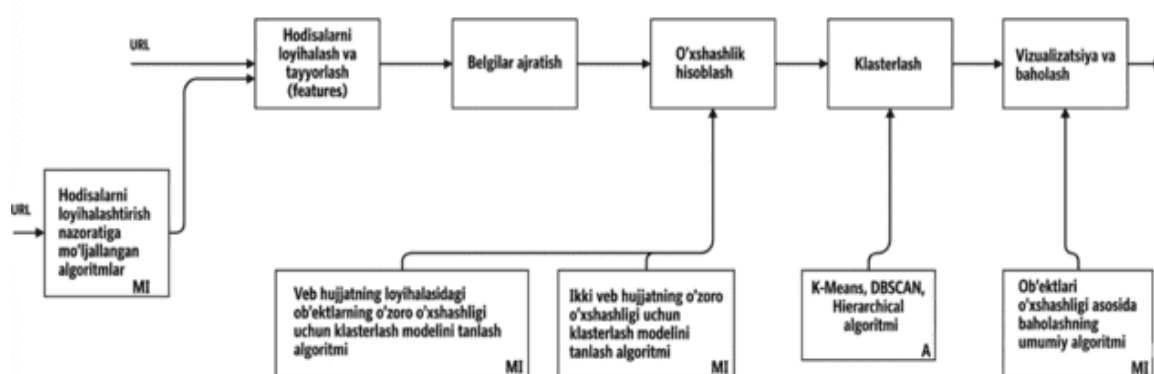
Veb hujjat obektlari o'xshashligi asosida baholash algoritmi hamda veb hujjat obektlari o'xshashligi asosida baholashning umumiy algoritmini ishlab chiqilgan.

Unda obyektlarning o'xshashligi, obyektlarga murojaatlar soni, obyektlarga murojaatlar soni bo'yicha o'xshashligi birlashtirilgan va umumiy baholash algoritmi ishlab chiqilgan.

Veb hujjatlarga intellektual ishlov berishning dasturiy vositasini loyihalashda tahlil va belgilarni ajratish amallarini kamaytirishda yangi veb hujjatlar hodisalarni loyihalashtirish nazoratiga mo'ljallangan algoritmlar ishlab chiqildi.

Hodisalarni loyihasini nazorat qilishning umumiy algoritmi 8 ta qism algoritmlaridan iborat.

Veb hujjatlarga intellektual ishlov berish dasturiy vositasini mantiqiy funksional tuzilmasini yuqoridagi elementlarni inobatga olib quyidagicha ishlab chiqildi.



8-rasm. Veb hujjatlarga intellektual ishlov berish dasturiy vositasining mantiqiy funksional tuzilmasi.

Veb hujjatlarni sinflashtirish va klasterlash asosida tahlil qilish va baholashda bilimlar va ma'lumotlar bazasi kerak. Bunda avvalo bilimlarni ifodalash modelini tanlash lozim. Tadqiqot doirasida Graf nazariyasi, RDF (Resource Description Framework) modeli, WOL (Web Ontology Language) modeli, Neo4j modeli va Rule-based modeli tadqiq qilindi. Tadqiqotning maqsadiga asosan Rule-based modeli asosida bilimlarni shakllantirish va boshqarishni loyihashtirish samarali bo'ladi deb hisoblandi.

Rule-based (qoidalarga asoslangan) bilimlar bazasi - bilimlarni "IF - THEN" qoidalari ko'rinishida ifodalaydi. Unda IF (shart) - ma'lum holat, vaziyat yoki belgilar yig'indisi va THEN (natija) - avtomatik chiqariladigan xulosa yoki hodisa. Uning ifodalanishi quyidagicha:

Rule: IF Condition(X) THEN Action(Y)

Matematik ifodalanishi esa $IF A \rightarrow B, A TRUE \Rightarrow B TRUE$.

Qoidalarga asoslangan bilimlar bazasining daraxt (ierarxik) tuzilishi ham bor. Bunda qoidalar va bilimlar daraxt shaklida tashkil etiladi.

Qoidalar [IF - THEN] ko'rinishida yozishda quyidagi shablon qoidalarini yaratib olinadi. Buni to'g'ridan-to'g'ri qoidalar bazasiga kiritish mumkin.

Umumiy holdagi shablon

IF (shartlar/alomatlar) THEN (xulosa/amal).

Formal (pseudokod) shabloni:

IF condition1 AND/OR condition2 AND ... THEN conclusion/action

Veb-hujjatlarni sinflashtirish shabloni:

IF ("terim" so'zi chastotasi > 5) OR (terim+mavjud) THEN class = "terimSinf"

IF (meta_keywords ichida "terim

+ " bor) AND (title da terimbor) THEN class = "terimSinf"

IF (h1 soni ≥ 1) AND (a – teglar soni > 20) THEN document_type = "news_portal"

Veb-hujjat tuzilmasiga (DOM) mos shabloni:

*IF (h1 mavjud) AND (img soni
≥ 3) AND (caption teglari mavjud) THEN layout
= "article_with_images"*

*IF (nav menyu topildi) AND (footer mavjud) AND (menu_items
≥ 5) THEN site_section = "main_page"*

Tavsiya (recommendation) qoidalari shabloni

*IF user_interest CONTAINS "machine learning" THEN recommend_category
= "AI/ML articles"*

*IF (user_read_time > 10 min) AND (shares ≥ 1) THEN recommend
= "long_read_content"*

Sifat nazorati yoki validatsiya uchun shabloni:

*IF (title yo'q) OR (meta_description yo'q) THEN quality_flag
= "low"; action = "needs_edit"*

IF (page_load_time > 3s) THEN performance_status = "slow"

Kirishni boshqarish (access control) uchun shabloni:

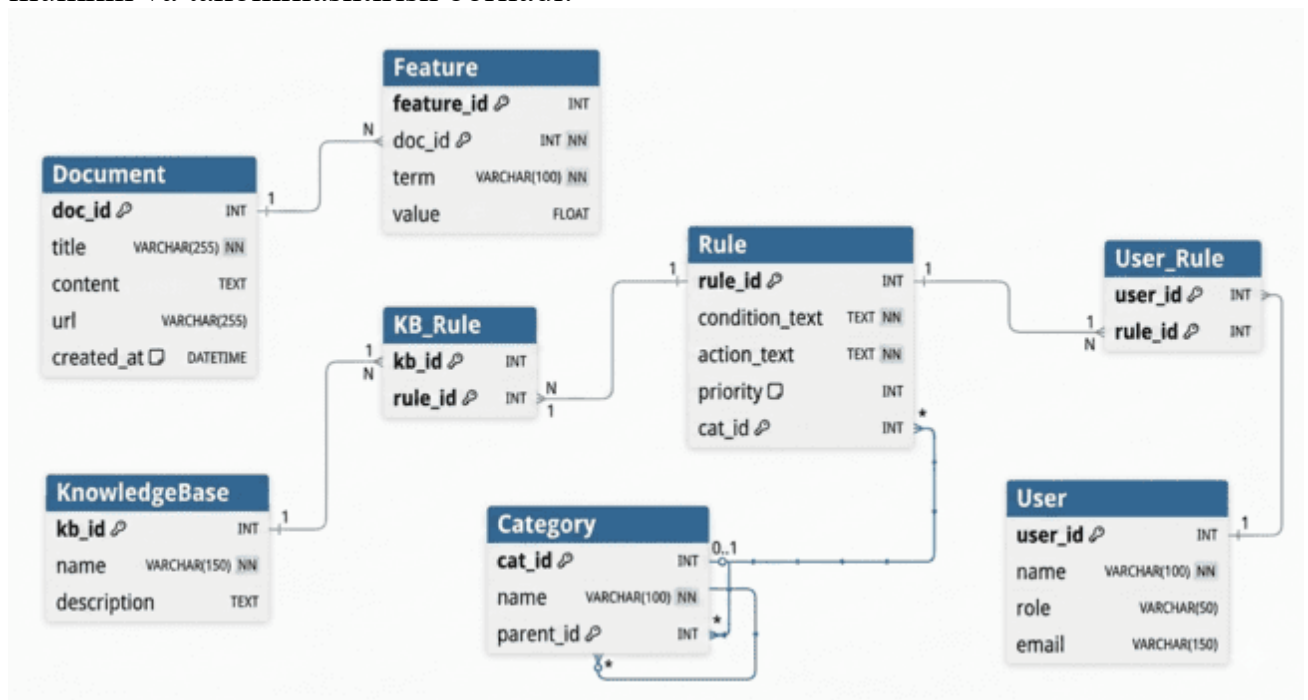
IF (user_role = "admin") THEN permission = "write, delete"
*IF (user_role = "editor") AND (section = "news") THEN permission
= "write"*

Bir necha qoida zanjiri (nested rules) shabloni:

IF "terim" belgisi TRUE THEN topic = "terimSinf"

*IF (topic = "terimSinf") AND (" + terim" exists) THEN subtopic
= " + terimSinf"*

Bu shablonlar asosida veb hujjatlarni tahlil qilishda ixtiyoriy bilimlarni yaratish
mumkin va takomillashtirish boriladi.



9-rasm. Bilimlar bazasining obyekt-aloqa modelining diagrammasi.

Tarmoq ma'lumotlari tahlili dasturi ta'minotlari va talablari aniqlashtirishdan
avval uning nazariy asoslarini inobatga olish lozim. Tarmoq nazariyasi asosida

tarmoq ma'lumotlari, tugunlar va ularning bog'lanishlari (yoylari)dan iborat bo'lib, Graf nazariyasi asosida tahlil qilinadi. Bunda markaz darajasi, markaz o'rtasi va klasterlash koeffitsienti kabi metrikalardan foydalaniladi. Axborot nazariyasidan foydalanib ma'lumot oqimlarini o'lchash, Entropiya va tarmoqdagi ma'lumot turliligini tahlil qilishda amalga oshiriladi. Bunda ham intellektual texnologiyalar yordamida MapReduce modeli - katta hajmli ma'lumotlarni taqsimlangan muhitida ishlov berish usuli, oqimli algoritmlar - ma'lumot oqimini uzluksiz va cheklangan xotiradan foydalangan holda ishlov beradigan algoritmlar, parallel hisoblash bir vaqtning o'zida bir nechta protsessor yoki yadro orqali hisob-kitoblarni bajarish usulidan foydalaniladi. Mashinali o'rgatish va sun'iy intellekt texnologiyalaridan tarmoq trafik tahlilida klasterlash, prognozlash qo'llaniladi, bunda graph neural network modelidan foydalanish maqsadga muvofiq.

Tarmoq ma'lumotlari tahliliga yo'naltirilgan dasturiy vosita rivojlantirish ilmiy va amaliy jihatdan juda muhimdir. Bu yo'nalishning asosi - graf nazariyasi, axborot nazariyasi, Big Data, machine learning/deep learning va graph neural network tarmoqlaridir. Yaqin kelajakda bunday dasturiy vositalar kiberxavfsizlik, internet buyumlari va intellektual tizimlarni rivojlantirishda hal qiluvchi rol o'ynaydi degan g'oyalar ham mavjud.

Veb hujjatlarni tahlil qiladigan dasturiy vositaning ta'minotlari, talablari va rivojlantirish, qo'llab quvvatlash yondashuvlar asosida Toshkent davlat iqtisodiyot universiteti, Toshkent davlat tibbiyot universiteti, Raqamli hukumat loyihalarini boshqarish markazi va Raqamli ta'lim texnologiyalarini rivojlantirish markazi kabi tashkilotlarga joriy qilidi. Joriy qilishda tashkilotlarning oldingi paragrafda keltirilgan talablari asosida texnik, tashkiliy, xodimlar bo'yicha barcha talablarga rioya qilindi. Bunda 4 ta tashkilotdan 4 ta administrator, 8 ta tajribali tashkilot tarmog'iga xizmat ko'rsatuvchi xodimlar, rahbariyat va professor-o'qituvchilar, tashkilot saytlaridan foydalanuvchilar ishtirok etib, jami 2000 dan ortiq bo'ldi.

Yuqorida keltirilgan 4 ta tashkilotlarda dasturiy vositaning samaradorligini aniqlashda uning parametrlarini aniqlab olish lozim. Talablarni tashkilotlarning ehtiyojlari va resurslaridan kelib chiqqan holda aniqlab olindi.

Asosiy samaradorlik parametrlariga quyidagilar kiradi:

1. T - Texnik samaradorlik parametrlari.
2. F - Funktsional samaradorlik parametrlari.
3. E - Iqtisodiy samaradorlik parametrlari.
4. U - Foydalanuvchi samaradorligi parametrlari.

Bu samaradorlik parametrlarining o'ziga xos tomonlarini quyidagicha aniqlab olindi. Bunda joriy qilinayotgan tashkilotlarning xodimlarining fikrlari va tashkilot strategiyasi inobatga olindi.

Dasturiy vositaning umumiy samaradorligi (5) ifoda bilan hisoblanadi.

$$Ea_{umumiy} = \frac{1}{4}(w_1T + w_2F + w_3E + w_4U) \quad (5)$$

bunda T - texnik ko'rsatkichlar, F - funksional ko'rsatkichlar, E - iqtisodiy ko'rsatkichlar, U - foydalanuvchi ko'rsatkichlar, w_i - salmoq koeffitsientlari (ekspert bahosiga ko'ra aniqlanadi).

T - Texnik samaradorlik parametrlari quyidagi ifoda bo'yicha belgilanadi:

$$T_{umumiy} = \frac{1}{5}(wt_1T_s + wt_2T_s + wt_3T_s + wt_4T_s + wt_5T_m) \quad (6)$$

bunda, wt_i - salmoq koeffitsientlari (ekspert bahosiga ko'ra aniqlanadi).

F - Funktsional samaradorlik parametrlari quyidagi ifoda bo'yicha belgilanadi:

$$F_{umumiy} = \frac{1}{5}(wf_1F_a + wf_2F_r + wf_3F_p + wf_4F_{AI} + wf_5F_A) \quad (7)$$

bunda, wf_i - salmoq koeffitsientlari (ekspert bahosiga ko'ra aniqlanadi).

E - Iqtisodiy samaradorlik parametrlari quyidagi ifoda bo'yicha belgilanadi:

$$E_{umumiy} = \frac{1}{4}(we_1E_t + we_2E_m + we_3E_e + we_4E_c) \quad (8)$$

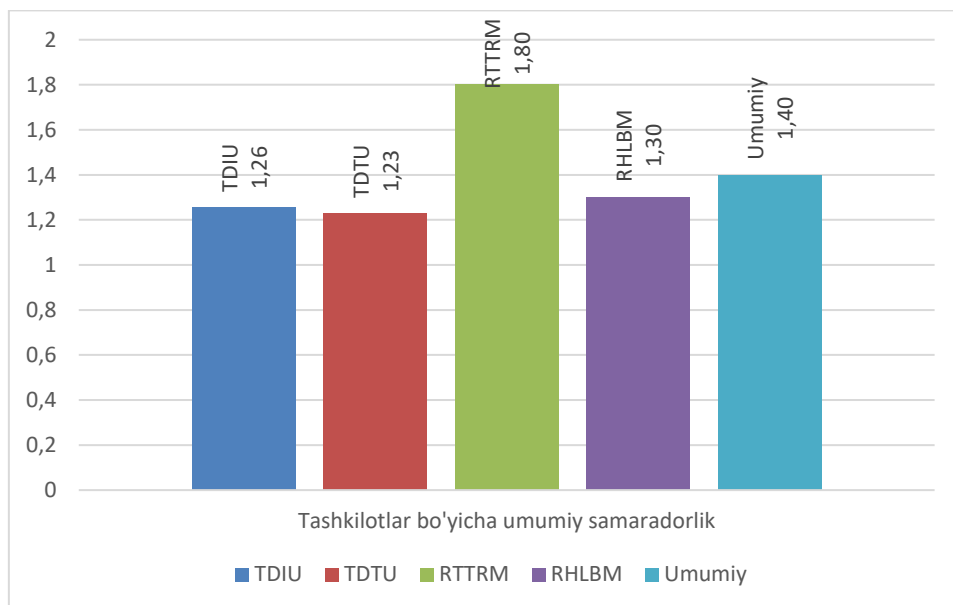
bunda we_i – salmoq koeffitsientlari (ekspert bahosiga ko'ra aniqlanadi).

U - Foydalanuvchi samaradorligi parametrlari quyidagi ifoda bo'yicha belgilanadi:

$$U_{umumiy} = \frac{1}{4}(wu_1U_i + wu_2U_l + wu_3U_q + wu_4U_u) \quad (9)$$

bunda wu_i – salmoq koeffitsientlari (ekspert bahosiga ko'ra aniqlanadi).

Yuqoridagi barcha olingan tahlililar natijasida dasturiy vositani joriy qilishdan oldin va keyingi holatni solishtirish orqali quyidagi samaradorlik natijalari aniqlanadi.



10-rasm. Umumiy samaradorlik

Samaradorlik natijalaridan Toshkent davlat iqtisodiyot universitetida 1,26 barobar, Toshkent davlat tibbiyot universitetida 1,23 barobar, Raqamli ta'lim texnologiyalarini rivojlantirish markazida 1,8 barobar, Raqamli hukumat loyihalarini boshqarish markazida esa 1,3 barobar oshganligini ko'rsatdi. Agar tadqiqotasosida ishlab chiqilgan dasturiy vositaning tanlangan tashkilotlari asosidagi samaradorligi 1.4 barobar ekanligini ko'rish mumkin.

XULOSA

“Sun’iy intellekt texnologiyalari asosida tarmoq ma’lumotlarini tahlil qilish model va algoritmlari” mavzusidagi dissertatsiya ishi bo’yicha olib borilgan tadqiqotlar natijasida quyidagi xulosalar taqdim etildi.

1. Veb hujjatlar tuzilishiga asoslangan holda obyektlar o’rtasidagi o’zaro o’xshashlikni aniqlash uchun metrikalarni tanlash algoritmlari ishlab chiqildi va ular document object model elementlarning xususiyatlarini samarali baholash hamda obyektlarni to’g’ri guruhlash imkonini berdi.

2. Veb hujjat obyektlarining o’zaro o’xshashligini baholash uchun matematik model va algoritmlar ishlab chiqilib, ular obyektlar orasidagi munosabatlarni aniq ifodalash hamda ularni samarali baholash imkonini berdi.

3. Produksion qoidalar modeli asosida bilimlarni loyihalash, obyekt-aloqa modeli va boshqarish algoritmi ishlab chiqildi. Taklif etilgan yondashuvda bilimlar IF–THEN qoidalari orqali formal ko’rinishda ifodalanib, ular orasidagi munosabatlar obyekt-aloqa modeli yordamida tizimli ravishda strukturallashtirildi. Bu esa bilimlar bazasini mantiqiy va semantik jihatdan yaxlit holda tashkil etish imkonini berdi.

4. Joriy etish bosqichida dasturiy vosita Toshkent davlat iqtisodiyot universiteti, Toshkent tibbiyot akademiyasi, Raqamli hukumat loyihalarini boshqarish markazi va Raqamli ta’lim texnologiyalarini boshqarish markazi kabi tashkilotlarda sinovdan o’tkazildi. Sinov natijalariga ko’ra, tizimning texnik, funksional, iqtisodiy va foydalanuvchi samaradorligi bo’yicha ijobiy o’sish qayd etildi. Texnik samaradorlik 1,25 barobar, funksional samaradorlik 0,25-0,30 barobar oralig’ida, iqtisodiy samaradorlik tashkilotlarda 2,6 barobargacha hamda foydalanuvchi samaradorligi o’rtacha 1,4 barobarga o’sdi.

Umuman olganda tadqiqotda veb hujjatlar tuzilishi, obyektlari va hodisalarini intellektual tahlil qilishga yo’naltirilgan matematik modellar, algoritmlar hamda qoidaviy bilimlar bazasini boshqarish yondashuvlari ishlab chiqildi. Natijada veb hujjatlarni samarali tahlil qilish, o’xshash obyektlarni aniqlash va tizimlarni optimallashtirish uchun yagona ilmiy asoslangan yechim taklif etildi.

**НАУЧНЫЙ СОВЕТ DSc.09/2025.27.12.T.01.03 ПО ПРИСУЖДЕНИЮ
УЧЕНЫХ СТЕПЕНЕЙ ПРИ ТАШКЕНТСКОМ УНИВЕРСИТЕТЕ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

**ТАШКЕНТСКОГО УНИВЕРСИТЕТА ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ ИМЕНИ МУХАММАДА АЛ-ХОРАЗМИЙ**

ХУСАНОВ ШЕРЗОД АБДУМАННОНОВИЧ

**МОДЕЛИ И АЛГОРИТМЫ АНАЛИЗА СЕТЕВЫХ ДАННЫХ НА
ОСНОВЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

05.01.11 – Цифровые технологии и искусственный интеллект

**АВТОРЕФЕРАТ ДИССЕРТАЦИИ ДОКТОРА ФИЛОСОФИИ (PhD) ПО
ТЕХНИЧЕСКИМ НАУКАМ**

Ташкент – 2026

Тема диссертации доктора философии (PhD) по техническим наукам зарегистрирована в Высшей аттестационной комиссии при Министерстве высшего образования, науки и инноваций Республики Узбекистан под номером B2025.3.PhD/T5888.

Диссертация выполнена в Ташкентском университете информационных технологий имени Мухаммада ал-Хоразмий.

Аннотация диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-странице Научного совета (www.tuit.uz) и на Информационно-образовательном портале «Ziyonet» (www.ziyonet.uz).

Научный руководитель:

Муминов Баходир Болтаевич
доктор технических наук, профессор

Официальные оппоненты:

Носиров Хабибулло Хикматулло угли
доктор технических наук, доцент

Хамроев Алишер Шодмонкулович
доктор философии по техническим наукам, ст.науч. сотр.

Ведущая организация:

Нукусский государственный технический университет

Защита диссертации состоится «9» сентября 2026 года в 14⁰⁰ часов на заседании Научного совета DSc.09/2025.27.12.T.01.03 при Ташкентском университете информационных технологий. (Адрес: 100084, г. Ташкент, ул. Амира Темура, д. 108. Тел.: (99871) 207-59-42, e-mail: ilmiy_kengash@tuit.uz).

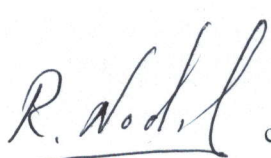
С диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского университета информационных технологий имени Мухаммада ал-Хоразмий (регистрационный номер № 445). (Адрес: 100084, г. Ташкент, ул. Амира Темура, д. 108. Тел.: (99871) 238-64-15).

Аннотация диссертации разослан «28» август 2026 года.
(Протокол реестра рассылки № 24 от «28» август 2026 года.)




М.М.Камилов
Председатель научного совета по присуждению ученых степеней, доктор технических наук, академик АН РУз


Н.А.Эгамбердиев
Ученый секретарь научного совета по присуждению ученых степеней, доктор философии по техническим наукам


Н.О.Рахимов
Председатель научного семинара при научном совете по присуждению ученых степеней, доктор технических наук, доцент

ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))

Актуальность и необходимость темы диссертации. В мировом масштабе основную часть информационных ресурсов составляют документы, формирующиеся в веб-среде. В глобальной сети Интернет ежедневно создаются миллиарды новых веб-страниц, а существующие обновляются или модифицируются. В результате данного процесса веб-среда превращается в сложную информационную систему огромного объема, характеризующуюся разнообразием форматов и структур данных.

Веб-документы по своей природе представляют собой не только текстовую информацию, но и включают изображения, видео, аудио, гиперссылки, скрипты, а также метаданные, формируя многослойную структуру. В связи с этим их комплексный анализ на основе технологий искусственного интеллекта, выявление взаимосвязей между ними и оценка степени сходства являются одной из актуальных научно-практических задач.

В мировой практике ведутся научно-исследовательские работы, направленные на разработку теоретических основ информационных технологий различного назначения, а также моделей и алгоритмов анализа веб-сайтов и страниц сетевых данных с применением технологий искусственного интеллекта. Данные подходы широко используются в системах кластеризации данных и интеллектуального анализа. В частности, поисковые системы при предоставлении пользователю наиболее релевантных результатов опираются на определение семантической близости между веб-страницами. В то же время в системах обнаружения плагиата важную роль играет не поверхностное сходство текста, а степень его соответствия с точки зрения семантики и структуры. Актуальность темы обусловлена тем, что в существующих подходах многие веб-документы оцениваются преимущественно на основе текстовых элементов, при этом семантические и объектно-ориентированные компоненты структуры веб-страниц (HTML-теги, атрибуты, DOM-деревья, сети гиперссылок, графические объекты) остаются без должного внимания. В результате принимаемые решения о сходстве веб-документов оказываются недостаточно точными и надёжными. С этой точки зрения особое внимание уделяется разработке новых алгоритмов и математических моделей, позволяющих оценивать сходство веб-документов не только на основе текста, но и с учётом их структурных, семантических и объектно-ориентированных характеристик в комплексе.

В Республике также реализуются широкомасштабные меры по обработке сетевых и веб-документных данных на основе технологий искусственного интеллекта, а также по их внедрению в социально-экономические сферы. В стране уделяется особое внимание повышению эффективности в различных отраслях национальной экономики и поддержке инноваций. В частности, в рамках стратегии «Цифровой Узбекистан - 2030» развитие технологий искусственного интеллекта включено в национальные программы цифровизации, где определены задачи по изучению возможностей использования и практическому внедрению таких технологий, как

виртуальная и дополненная реальность, искусственный интеллект, криптография, машинное обучение, анализ больших данных и облачные вычисления в различных секторах экономики. Кроме того, в постановлении «Об утверждении стратегии развития технологий искусственного интеллекта до 2030 года» определена задача по обработке и анализу «больших данных» на основе технологий искусственного интеллекта. Реализация данных задач, в частности внедрение технологий искусственного интеллекта при анализе веб-сайтов организаций и образовательных учреждений, имеет важное значение.

Данное диссертационное исследование в определённой степени способствует реализации задач, предусмотренных в нормативно-правовых документах, регулирующих данную сферу, в частности: Указе Президента Республики Узбекистан от 22 октября 2025 года № ПФ-189 «О дополнительных мерах по дальнейшему развитию технологий искусственного интеллекта», Постановлении от 14 октября 2024 года № ПП-358 «Об утверждении стратегии развития технологий искусственного интеллекта до 2030 года», Постановлении от 17 февраля 2021 года № ПП-4996 «О мерах по созданию условий для ускоренного внедрения технологий искусственного интеллекта», а также Постановлении Кабинета Министров от 10 июля 2025 года № 425 «О мерах по реализации приоритетных проектов в сфере технологий искусственного интеллекта в 2025–2026 годах», и других нормативно-правовых актах, относящихся к данной деятельности.

Соответствие исследования приоритетным направлениям развития науки и технологий Республики. Данное исследование выполнено в рамках IV приоритетного направления развития науки и технологий Республики — «Информатизация и развитие информационно-коммуникационных технологий».

Степень изученности проблемы. За рубежом стремительно развиваются научные исследования в области обработки больших данных на основе технологий машинного и глубокого обучения, а также анализа и оценки веб-документов. Зарубежными учёными Larry Page и Sergey Brin был разработан алгоритм PageRank. Данный алгоритм был создан ими в период обучения в докторантуре Stanford University. Впоследствии алгоритм PageRank стал одним из ключевых механизмов функционирования поисковой системы Google. В области семантического анализа содержания веб-документов значительный вклад внесли научные работы Christopher D. Manning, Hinrich Schütze и Daniel Jurafsky. В результате их исследований были разработаны методы представления текстовой информации в векторном пространстве, определения семантического сходства и автоматического анализа.

Значительный вклад в развитие технологий машинного и глубокого обучения внесли такие учёные, как Geoffrey Hinton, Yoshua Bengio и Yann LeCun. Разработанные ими модели нейронных сетей (CNN, RNN, архитектура Transformer) существенно расширили возможности анализа текстовых, графических и других мультимодальных данных веб-документов. Ряд научных исследований также был проведён такими учёными, как Tim Berners-

Lee, Ricardo Baeza-Yates, Christopher D. Manning, Andrew Ng, а также рядом местных исследователей.

В Республике Узбекистан также проводятся научные исследования в области технологий машинного и глубокого обучения. Академики Академии наук Республики Узбекистан Т.Ф. Бекмуродов, М.М. Камилов, а также профессора Ш.Х. Фозилов, М.М. Мусаев, А. Нишанов, Н.С. Маматов, Д. Мухамедиева, М. Рахматуллаев, Б.Б. Муминов, Н. Рахимов и другие учёные вносят значительный вклад в развитие технологий искусственного интеллекта. Работы данных учёных служат важной научной основой для развития искусственного интеллекта и интеллектуальных вычислительных систем.

Несмотря на то, что в Республике Узбекистан в данном направлении проводятся определённые научные исследования, нельзя утверждать, что эта область достигла достаточного уровня развития. Во-первых, недостаточное количество научных лабораторий и центров, обладающих высокими вычислительными мощностями, ограничивает возможности работы с большими объёмами данных. Во-вторых, из-за нехватки локальных баз данных и национальных датасетов многие исследования проводятся на основе зарубежных данных, что снижает степень адаптации полученных результатов к местным условиям. В-третьих, наблюдается дефицит квалифицированных специалистов в данной области, вследствие чего научные исследования, направленные на разработку моделей и алгоритмов анализа веб-сайтов и страниц сетевых данных на основе технологий искусственного интеллекта, в Узбекистане ещё не полностью сформированы и недостаточно изучены.

Цель исследования. Разработка моделей и алгоритмов анализа веб-страниц сетевых информационных ресурсов на основе технологий искусственного интеллекта.

Задачи исследования:

провести анализ моделей и алгоритмов выполнения и совершенствования анализа структуры веб-документов на основе технологий искусственного интеллекта;

разработать гибридный алгоритм, ориентированный на анализ и проектирование событий веб-сайта на основе технологий искусственного интеллекта;

усовершенствовать алгоритмы кластеризации структуры веб-документов на основе технологий искусственного интеллекта;

разработать математическую модель и алгоритм оценки на основе сходства объектов веб-документа;

усовершенствовать объектно-реляционную модель и алгоритм управления на основе продукционных знаний;

спроектировать программное средство для интеллектуального анализа и оценки веб-документов.

Объект исследования. Веб-документы и их структурные компоненты (HTML-элементы, метаданные, гиперссылки, контентные блоки).

Предмет исследования. Математическое моделирование сходства объектов веб-документов, а также алгоритмы определения их семантической и структурной взаимосвязанности.

Методы исследования. В исследовании использованы методы математического моделирования, кластеризации, коэффициент сходства Жаккара, косинусное сходство, а также методы машинного обучения и обработки естественного языка (NLP). Кроме того, применялись язык программирования Python и библиотеки BeautifulSoup, Scikit-learn и TensorFlow.

Научная новизна диссертационного исследования заключается в следующем:

- разработан гибридный алгоритм, ориентированный на проектирование событий веб-документов на основе технологий искусственного интеллекта;

- разработаны алгоритмы выбора метрик, предназначенные для определения взаимного сходства объектов на основе структуры веб-документов;

- Разработаны математическая модель и алгоритм оценки взаимного сходства объектов веб-документов на основе анализа их событий.

- разработаны объектно-реляционная модель представления знаний и алгоритм управления на основе продукционных правил.

Практическая значимость исследования заключается в следующем:

- разработаны модели и алгоритмы классификации сетевых данных на основе структуры веб-документов, реализована классификация объектов по степени их взаимного сходства;

- разработаны алгоритмы оценки веб-документов на основе анализа гиперссылок, результаты их практического применения внедрены в процесс обработки данных;

- разработан гибридный алгоритм, предназначенный для управления проектированием событий; выполнены анализ и обработка событий, содержащихся в веб-документах;

- спроектировано программное средство для классификации и оценки веб-документов на основе анализа их структуры.

Достоверность результатов исследования. Достоверность результатов исследования обоснована тем, что исследования проводились с использованием методов и средств, основанных на технологиях искусственного интеллекта, а также согласованностью результатов теоретических и экспериментальных исследований, выполненных с применением аппаратно-программных средств. Кроме того, она подтверждается положительными результатами экспериментальных испытаний разработанных моделей и алгоритмов анализа веб-документов на основе технологий искусственного интеллекта, проведённых на веб-сайтах организаций, наличием соответствующих подтверждающих документов, а также внедрением результатов исследования в практику.

Научная и практическая значимость результатов исследования.

Научная значимость исследования заключается в том, что в его рамках, в пределах теоретических и практических основ использования технологий искусственного интеллекта при анализе веб-документов и сетевых данных, разработаны алгоритмы выбора метрик, направленные на определение взаимного сходства объектов на основе структуры веб-документов. Предложенные подходы, учитывающие сходство объектов, обеспечивают реализацию процесса кластеризации и позволяют получить новые научные решения. Таким образом, исследование обладает значимой научной ценностью в области теории кластеризации, методов машинного обучения и задач определения сходства. Практическая значимость результатов исследования выражается в том, что предложенный комплекс алгоритмов обеспечивает возможность анализа данных веб-документов. На основе разработанных алгоритмов создан программный комплекс, который служит методической основой для интеграции в прикладные аналитические системы, экономии вычислительных ресурсов и ускорения процесса принятия решений. Это подтверждает практическую направленность результатов исследования и их широкие возможности применения.

Внедрение результатов исследования. На основе разработанного в ходе исследования программного средства интеллектуальной обработки и оценки веб-документов:

разработанное на основе гибридного алгоритма, ориентированного на проектирование событий веб-документов с использованием технологий искусственного интеллекта, программное средство внедрено в практическую деятельность Ташкентского государственного экономического университета Министерства высшего образования, науки и инноваций Республики Узбекистан в целях анализа и совершенствования его официального веб-сайта (справка Министерства цифровых технологий Республики Узбекистан № 33-8/4770 от 2 июня 2026 года). В результате показатели эффективности официального веб-сайта университета по техническим, функциональным, экономическим и пользовательским параметрам повышены в 1,26 раза;

разработанное на основе алгоритмов выбора метрик, ориентированных на определение взаимного сходства объектов веб-документов по их структуре, а также алгоритмов классификации структуры веб-документов программное средство внедрено в практическую деятельность «Центра развития цифровых образовательных технологий» при Министерстве высшего образования, науки и инноваций Республики Узбекистан в целях анализа и совершенствования разработанных в центре информационных систем (справка Министерства цифровых технологий Республики Узбекистан № 33-8/4770 от 2 июня 2026 года). В результате показатели эффективности разработанных в центре информационных систем по техническим, функциональным, экономическим и пользовательским параметрам повышены в 1,8 раза;

программное средство, разработанное на основе математической модели и алгоритма оценки взаимного сходства объектов веб-документов на основе

их событий, внедрено в практическую деятельность «Центра управления проектами цифрового правительства» при Министерстве цифровых технологий Республики Узбекистан в целях анализа и совершенствования информационных систем (справка Министерства цифровых технологий Республики Узбекистан № 33-8/4770 от 2 июня 2026 года). В результате показатели эффективности разработанных в центре информационных систем по техническим, функциональным, экономическим и пользовательским параметрам повышены в 1,3 раза;

программное средство, разработанное на основе объектно-реляционной модели проектирования знаний и алгоритма управления на основе продукционных правил, а также алгоритмов классификации структуры веб-документов, внедрено в практическую деятельность Ташкентского государственного медицинского университета Министерства высшего образования, науки и инноваций Республики Узбекистан в целях анализа и совершенствования его официального веб-сайта (справка Министерства цифровых технологий Республики Узбекистан № 33-8/4770 от 2 июня 2026 года). В результате создана возможность повышения показателей эффективности официального веб-сайта университета по техническим, функциональным, экономическим и пользовательским параметрам в 1,62 раза.

Апробация результатов исследования. Результаты данного исследования были представлены и обсуждены на 2 международных и 2 республиканских научно-практических конференциях.

Публикация результатов исследования. Основные результаты исследования опубликованы в 13 научных работах, из них 5 в научных изданиях, рекомендованных Высшей аттестационной комиссией Республики Узбекистан для опубликования основных научных результатов докторских диссертаций, в том числе 2 в зарубежных научных журналах и 2 в республиканских научных журналах, 1 в материалах международной научной конференции. Кроме того, получены 2 свидетельства об официальной регистрации программ для ЭВМ.

Структура и объём диссертации. Диссертационная работа состоит из введения, четырёх глав, заключения и списка использованной литературы. Общий объём диссертации составляет 124 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Обоснованы актуальность и необходимость темы диссертации, показано соответствие исследования приоритетным направлениям развития науки и технологий Республики. Определены цель и задачи исследования, а также объект и предмет исследования. Обоснована достоверность полученных результатов, раскрыты их теоретическая и практическая значимость. Приведены сведения о внедрении результатов исследования в практику, опубликованных работах, а также о структуре диссертации.

В первой главе диссертации под названием «**Теоретические основы и средства анализа сетевых данных**» рассматривается анализ веб-документов как процесс технического изучения или оценки электронных документов.

Метод анализа веб-документов, как и многие другие методы исследования, направлен на изучение проблем анализа и интерпретации данных посредством их интеллектуальной или частично интеллектуальной обработки с целью раскрытия смысла, понимания и формирования выводов, а также на разработку путей их решения и постановку задачи исследования.

В параграфе 1.1 освещена значимость систем поддержки принятия решений в предметных областях, а также исследованы методы и алгоритмы, соответствующие каждому этапу их разработки и применения. Кроме того, приведены сведения об интеллектуальных системах, используемых в мировой практике для анализа веб-документов.

В параграфе 1.2 представлены основные понятия и определения моделей и алгоритмов веб-аналитики, ориентированных на контент веб-документов. В нём приведены подробные сведения о правилах, базах знаний, а также о файловых структурах, используемых на веб-сайтах. Кроме того, рассмотрены современные подходы к созданию интеллектуальных систем.

В параграфе 1.3 изложены вопросы интеллектуальной обработки сетевых данных, применения машинного обучения, обработки естественного языка (NLP), моделей искусственного интеллекта на основе теории графов, а также возможности выявления скрытых взаимосвязей между веб-документами с помощью алгоритмов машинного обучения, их автоматической классификации и кластеризации. Также приведены сведения об алгоритмах обучения без учителя, в частности о методах кластеризации.

Вторая глава диссертации под названием **«Модели и алгоритмы, ориентированные на структуру и анализ веб-документов»** состоит из 5 параграфов. В ней рассматриваются вопросы изучения веб-сайта с использованием веб-методов с целью улучшения структуры веб-документов и проведения анализа, сбора данных о структуре сайта на основе регламентированных процедур, построения веб-графа и проведения анализа, разработки критериев потребностей, определения контрольных правил, улучшающих структуру в соответствии с заданными критериями, а также разработки рекомендаций и предложений по совершенствованию структуры сайта.

В результате обосновано, что вебометрические модели, основанные на технологиях искусственного интеллекта, являются эффективным научно-практическим решением для совершенствования структуры веб-документов, их автоматической оценки и глубокого анализа сетевых данных.

В веб-документах, используемых в компьютерных сетях, были спроектированы 8 типов событий.

При этом все события веб-документа взаимосвязаны и требуют применения специального алгоритма или подхода для контроля процесса их проектирования. Корректное проектирование событий в веб-документах повышает качество их функционирования и обеспечивает корректную обработку и структурирование контентной информации. Для реализации контроля проектирования событий предлагается следующий алгоритм.

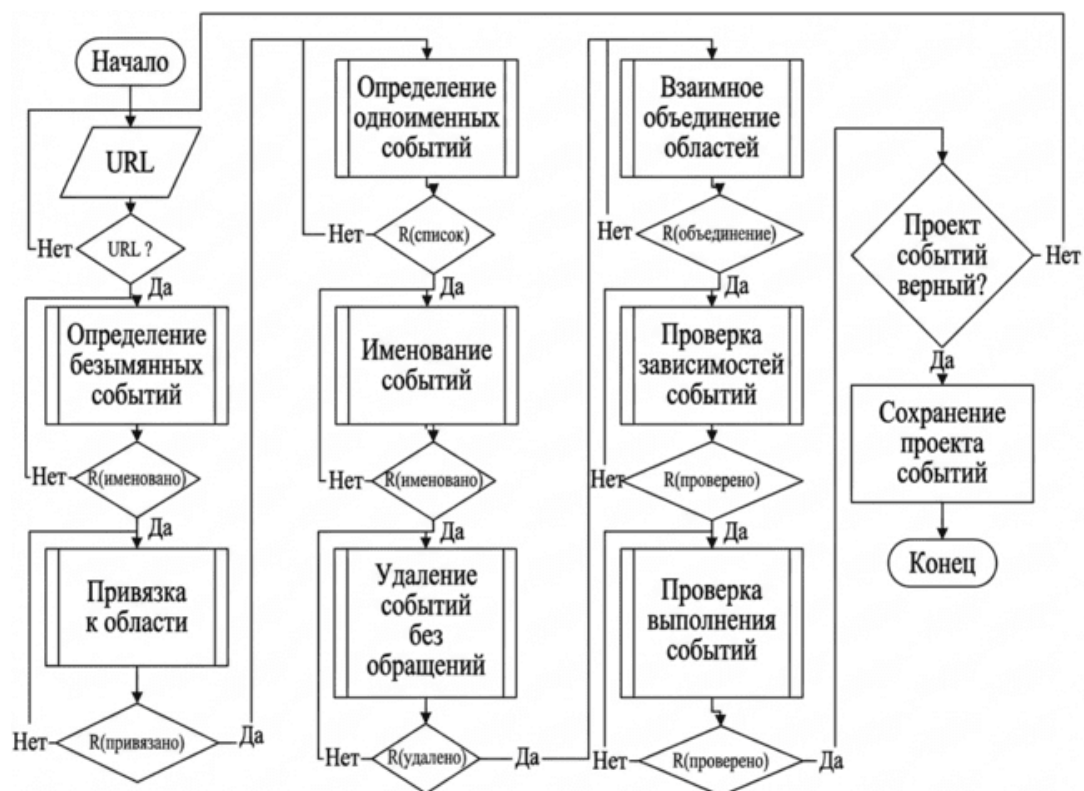


Рисунок 1. Общий алгоритм контроля проектирования событий.

Рассмотрены вопросы совершенствования рейтингового алгоритма PageRank, направленного на определение значимости веб-сайтов и веб-документов в сети. В интернет-среде веб-страницы образуют сложную сеть, взаимосвязанную посредством гиперссылок. Эффективный анализ такой сети и определение авторитетности страниц требуют применения интеллектуальных подходов, основанных на теории графов и итерационных вычислениях. В исследовании веб-страницы рассматриваются как вершины, а их взаимные ссылки - как ориентированные рёбра, при этом веб-среда представлена в виде модели ориентированного графа. На основе данного графа для определения относительной значимости страниц применяется алгоритм PageRank. Основная идея алгоритма заключается в том, что авторитет страницы определяется количеством страниц, ссылающихся на неё, а также их собственным уровнем значимости.

Проведён анализ классической модели алгоритма PageRank, выявлены его вычислительная сложность и ограничения при работе с большими объёмами сетевых данных. На этой основе предложен усовершенствованный алгоритмический подход, направленный на эффективное вычисление рейтинга PageRank. Предложенный метод обеспечивает оптимизацию итерационного процесса вычислений, снижение избыточных вычислений, а также достижение быстрых и устойчивых результатов при обработке больших наборов веб-страниц.

В результате предложенный алгоритм эффективного вычисления рейтинга PageRank научно и практически обоснован как средство более точной оценки веб-сайтов и веб-документов в сетевой среде, а также для применения в поисковых системах, веб-аналитике и задачах анализа сетевых

данных. Разработка алгоритма выбора моделей кластеризации на основе структуры и гиперссылок веб-документа исследуется на основе двух отдельных алгоритмов. Это обусловлено тем, что структура веб-документа присутствует на каждой странице и требует применения отдельного подхода при кластеризации. В свою очередь, гиперссылки веб-документа связаны с каждой страницей..

1. Метрические методы измерения взаимного сходства двух заданных объектов на основе их соответствующих характеристик. В данном случае разрабатывается специальная библиотека *BiDist*, основная функция которой имеет следующий вид:

$$d = BiDist(x, y, metric_{name}, outType, rw)$$

При этом x, y - вектор признаков объекта, $metric_{name}$ - тип метрики, $outType$ - тип выходного значения сходства, rw - параметры (настройки).

В качестве метрик или функций расстояния могут быть выбраны следующие: 'braycurtis', 'kanberra', 'chebyshev', 'cityblock', 'korrelyatsiya', 'kosinus', 'dice', 'evklid', 'hamming', 'jaccard', 'jensenshannon', 'kulczynski1', 'mahalanobis', 'match', 'minkowski', 'rogerstanimoto', 'russellrao', 'seuclidean', 'sokalmichener', 'sokalsneath', 'sqeuclidean', 'yule'.

Приведённые выше метрики измерения расстояния широко используются в расширенных моделях. Взаимосвязи между объектами веб-документов определяют степень их сходства, и целесообразно определять её на основе указанных метрик. Для объединения всех этих математических выражений необходимо разработать инструментальную библиотеку.

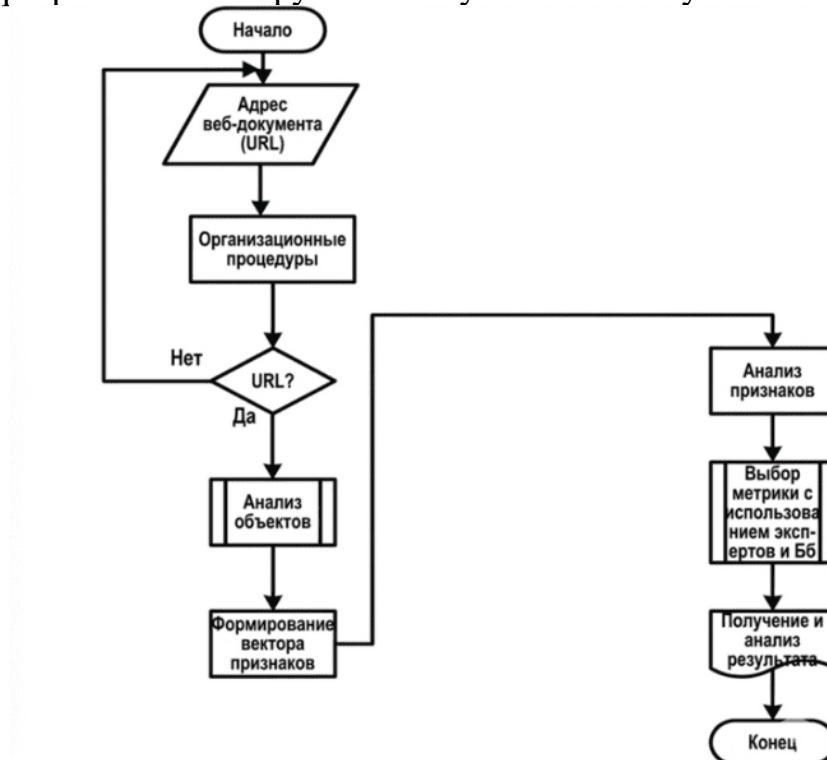


Рисунок 2. Алгоритм кластеризации для определения взаимного сходства объектов.

2. Метрические методы измерения взаимного сходства объектов двух заданных веб-документов на основе соответствующих пар их

характеристик. В данном случае разрабатывается специальная библиотека *cDist*, основная функция которой имеет следующий вид:

$$d = cDist(x_a, x_b, metric_{name}, outType, rw)$$

При этом x, y - вектор признаков объекта, *metric_name*- тип метрики, *outType*- тип выходного значения сходства, *rw*- параметры (настройки).

Приведённые выше метрики измерения расстояния широко применяются в расширенных моделях. Для объединения всех этих математических выражений необходимо разработать инструментальную библиотеку.

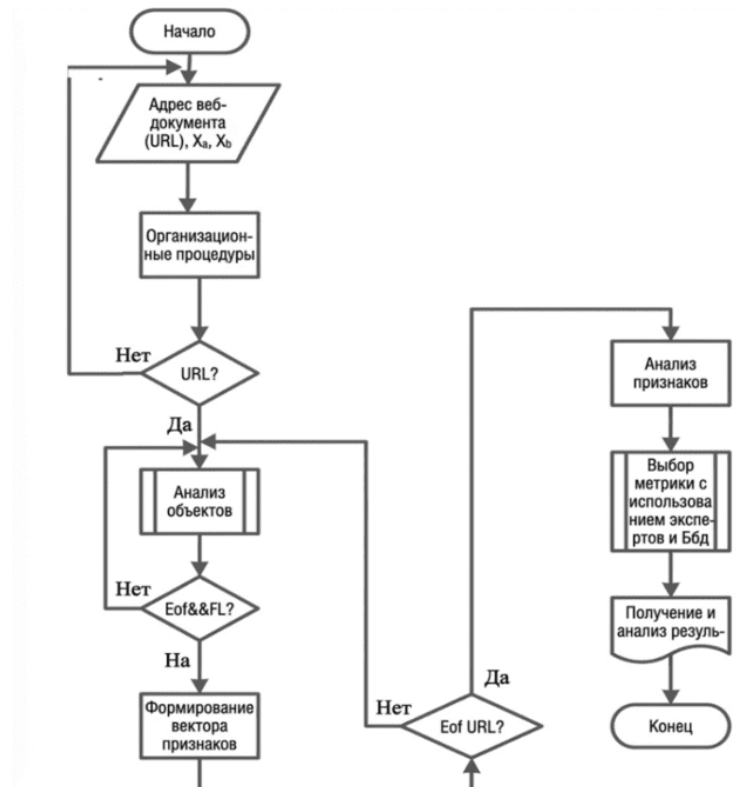


Рисунок 3. Алгоритм кластеризации для определения взаимного сходства двух веб-документов.

Алгоритм выбора модели кластеризации для определения сходства характеристик объектов двух веб-документов предусматривает вычисление сходства объектов в структуре двух и более веб-документов. На первом этапе указываются полные пути к двум веб-документам и осуществляется переход к организационным процедурам. В случае выявления недостатков на данном этапе процесс запускается повторно, в противном случае выполняется алгоритм определения объектов веб-документа. После анализа объектов формируются соответствующие векторы признаков для каждого объекта. Данный процесс повторяется в зависимости от количества веб-документов. Затем проводится анализ признаков двух векторов, в результате чего формируется соответствующий логический вектор. С использованием экспертных знаний и базы знаний осуществляется выбор метрики, соответствующей типу признаков. В результате вычисляются значения сходства на основе соответствующих признаков объектов двух веб-документов, представленные в виде матрицы.

Расчёт сходства объектов веб-документа, как указано в предыдущем параграфе, предлагается осуществлять, разделяя его на два подхода и используя модели и алгоритмы оценки на основе сходства объектов.

Пусть сходство объектов одного веб-документа задано в виде матрицы $Q = (q_{i,j})$, где $q_{i,j}$ - значение сходства между i -м и j -м объектами веб-страницы. Если рассматривать каждую характеристику объекта как вершину ориентированного графа, то можно построить графовую модель сходства объектов веб-документа.

Пусть ориентированный граф задан в виде $D = (V, E)$, где V - множество вершин графа, в данном случае соответствующих объектам, а E - упорядоченное множество рёбер. При этом $e_i \in V, i = 1, \dots, n$, где n — количество объектов.

Пусть входящие дуги вершины обозначаются как $e_i^I \in V$, а исходящие дуги как $e_i^O \in V, i = 1, \dots, n$, где n - количество объектов. Весами этих дуг принимаются значения входящих и исходящих связей q_{ij} , отражающие степень сходства i -го объекта с j -м объектом.

Ориентированный граф может быть построен следующим образом.

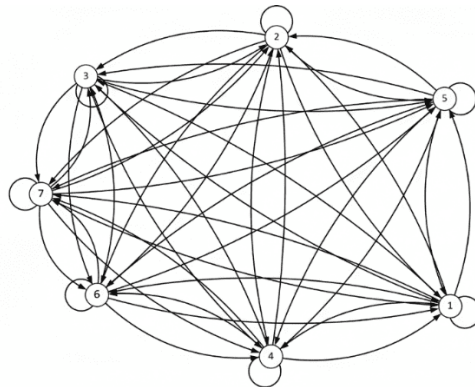


Рисунок 4. Графовая модель, основанная на сходстве объектов веб-документа.

В графовой модели, основанной на сходстве объектов веб-документа, веса сходства не заданы. Эти веса могут быть представлены в виде матрицы следующим образом.

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} & q_{1,3} & \dots & q_{1,n} \\ q_{2,1} & q_{2,2} & q_{2,2} & \dots & q_{2,n} \\ q_{3,1} & q_{3,2} & q_{3,3} & \dots & q_{3,n} \\ \dots & \dots & \dots & \dots & \dots \\ q_{n,1} & q_{n,2} & q_{n,3} & \dots & q_{n,n} \end{pmatrix}$$

На основе графовой модели объектов веб-документа возможно проведение оценки на основе их сходства. Матрица сходства объектов веб-документа не является полностью симметричной, то есть её значения не совпадают. На основе значений матрицы принимаются: для входящих дуг e_i^I - строка матрицы Q , то есть $(q_{i1}, q_{i2}, \dots, q_{in})$, а для исходящих дуг e_j^O - столбец матрицы Q , то есть $(q_{1j}, q_{2j}, \dots, q_{nj})$.

При оценке активности объекта веб-страницы предлагается делать вывод на основе его матрицы сходства с использованием следующего выражения:

$$W_j = \sum_{i=1}^n q_{i,j} - \sum_{i=1}^n q_{j,i} \quad , j = 1..m \quad (1)$$

Из значения выражения (1) необходимо ввести следующие свойства:

1. Если $W_j > 0$, то переход осуществляется от объекта q_j к другим объектам, и данный объект считается весьма важным для веб-страницы.
2. Если $W_j = 0$, то переход осуществляется от объекта q_j к другим объектам, и количество входящих переходов к данному объекту равно количеству исходящих, вследствие чего он считается хорошим (удовлетворительным) для веб-страницы.
3. Если $W_j < 0$, то переход осуществляется от других объектов к объекту q_j , и данный объект считается неудовлетворительным для веб-страницы. Такой объект необходимо как можно скорее усовершенствовать или удалить.

В целях реализации мероприятий, предусмотренных планом деятельности организации, предложен подход, основанный на индексировании и векторном представлении взаимно схожих объектов. Такой подход реализуется путём определения рейтингов объектов на основе оценки их взаимного сходства в структуре веб-документа.

При определении рейтинга объектов веб-документа необходимо учитывать количество обращений пользователей к данному объекту, то есть число входящих и исходящих переходов. При этом осуществляется отслеживание движения курсора или мыши на веб-странице и подсчёт взаимодействий с объектами. В данном случае можно опираться на следующие 5 основных событий мыши:

1. MuoseDown – перемещение курсора мыши в пределах области действия объекта по координате X;
2. MouseUp - перемещение курсора мыши в пределах области действия объекта по координате Y;
3. MouseMove - выход курсора мыши за пределы области действия объекта;
4. Click (mouseClick) - выбор объекта посредством однократного нажатия кнопки мыши в пределах области его действия;
5. DbClick (mouseDbClick) - выбор объекта посредством двойного щелчка кнопкой мыши в пределах области его действия; при возникновении данного события осуществляется переход к другому объекту.

Все перечисленные выше действия мыши реализованы во всех языках программирования и скриптовых языках. В каждом объекте, размещённом на веб-странице, имеется возможность управления этими событиями.

В связи с этим предлагаются следующие подходы:

1. Для событий MouseDown и MouseUp каждому объекту присваивается уникальный идентификатор i (начальное значение $i = 0$). Операция $i++$

выполняется один раз в реальном времени при возникновении соответствующего события для данного объекта.

2. Для события MouseMove каждому объекту присваивается уникальный идентификатор j (начальное значение $j = 0$). Операция $j++$ выполняется один раз в реальном времени при возникновении соответствующего события для данного объекта.

3. Для события Click каждому объекту присваивается уникальный идентификатор c (начальное значение $c = 0$). Операция $c++$ выполняется один раз в реальном времени при возникновении соответствующего события для данного объекта.

4. Для события DbClick каждому объекту присваивается уникальный идентификатор dc (начальное значение $dc = 0$). Операция $dc++$ выполняется один раз в реальном времени при возникновении соответствующего события для данного объекта.

Данные переменные функционируют с момента запуска веб-страницы и продолжают выполнять свои функции до получения статистических данных. Для каждого объекта эти четыре события работают непрерывно и рассчитываются по следующему выражению.

$$E_I = i + c, \quad E_O = j + dc \quad (2)$$

E_I – это входящие обращения к объекту а E_O – исходящие обращения от объекта. Характеристики событий учитываются и разделяются на две группы.

На основе входящих и исходящих обращений, определяемых событиями, можно построить информационную модель в виде графа. При этом рассматриваемый текущий объект представляется как вершина графа, а входящие и исходящие обращения - как его рёбра.

Пусть ориентированный граф задан в виде $G = (O, E)$, где O - множество вершин графа, в данном случае соответствующих объектам. - это упорядоченное множество, соответствующее входящим и исходящим обращениям, при этом $e_i \in O, i = 1, \dots, n$, где n - количество объектов.

Пусть входящие дуги обращений к вершине обозначаются как $e_i^I \in O$, а исходящие дуги обращений - как $e_i^O \in O, i = 1, \dots, n$, где n - количество объектов. Весами этих дуг принимаются значения q_{ij} , характеризующие количество входящих или исходящих обращений от i -го объекта к j -му объекту. Ориентированный граф может быть построен следующим образом.

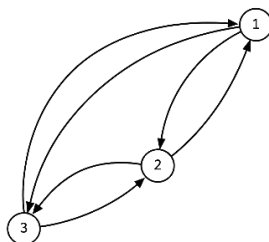


Рисунок 5. Графовая модель, основанная на обращениях к объектам веб-документа.

В общем случае пусть один веб-документ содержит n веб-страниц, и на каждой из этих страниц имеется по m объектов. Тогда общее количество объектов составляет $N = n \cdot m$. Входящие и исходящие обращения веб-документа могут быть представлены в виде матрицы M . При этом выполняются следующие условия:

1. Каждая соответствующая строка матрицы M представляет исходящие обращения.
2. Каждый соответствующий столбец матрицы M представляет входящие обращения.

3. В этом случае значения диагональных элементов матрицы M вычисляются по следующему выражению: $E_D = \frac{E_I + E_O}{2}$

Общий вид матрицы M выражается следующим образом:

$$M = \begin{pmatrix} E_{D_{1,1}} & E_{O_{1,2}} & E_{O_{1,3}} & \dots & E_{O_{1,n}} \\ E_{I_{2,1}} & E_{D_{2,2}} & E_{O_{2,3}} & \dots & E_{O_{2,n}} \\ E_{I_{3,1}} & E_{I_{3,2}} & E_{D_{3,3}} & \dots & E_{O_{3,n}} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ E_{I_{n,1}} & E_{I_{n,2}} & E_{I_{n,3}} & \dots & E_{D_{n,n}} \end{pmatrix}$$

Для каждого объекта из матрицы M величины E_I^k и E_O^k определяются следующим выражением

$$E_I^k = \sum_{i=1, i \neq k}^N E_{I_{k,i}}, E_O^k = \sum_{i=1, i \neq k}^N E_{O_{k,i}} \quad k = 1 \dots N \quad (3)$$

Из выражения (3) можно сделать следующие выводы:

1. Если для каждого k выполняется условие $E_I^k - E_O^k < 0$, то данный объект необходимо обновить или внести в него изменения.
2. Если для каждого k выполняется условие $E_I^k - E_O^k = 0$, то данный объект считается хорошим (удовлетворительным).
3. Если для каждого k выполняется условие $E_I^k - E_O^k > 0$, то данный объект считается высокоэффективным, и целесообразно увеличивать количество таких объектов.

На основе предложенного подхода для веб-документа формируется модифицированный вариант матрицы M^* .

$$M^* = \begin{cases} 1, & E_I^k - E_O^k > 0 \\ 0, & E_I^k - E_O^k = 0 \\ -1, & E_I^k - E_O^k < 0 \end{cases} \quad (4)$$

На основе значений вектора M^* обеспечивается возможность визуального выявления объектов веб-документа, требующих изменений.

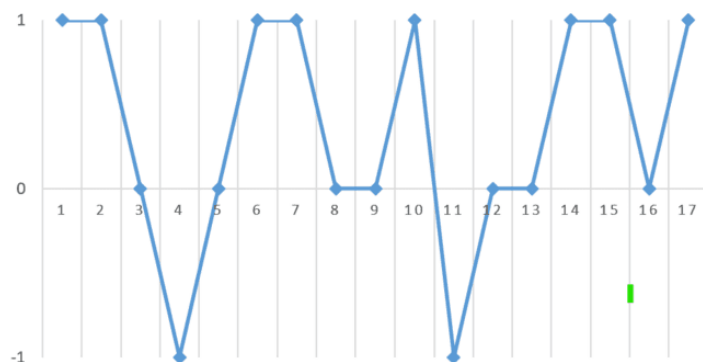


Рисунок 6. Граф обращений к объектам веб-документа.

При эффективном вычислении усовершенствованного рейтинга PageRank использованы методы выпуклой оптимизации и условного градиента Франка-Вульфа.

При проектировании программного средства определяются его объект и предмет.

Объект программного средства: веб-граф и сеть гиперссылок в нём.

Предмет программного средства: усовершенствованные алгоритмы вычисления рейтинга PageRank веб-страниц на основе методов выпуклой оптимизации и условного градиента Франка-Вульфа.

Цель программного средства: обеспечение быстрого и точного вычисления рейтинга веб-страниц, а также интеграции полученных результатов в соответствующую базу данных.

При проектировании программного средства используются алгоритмы и вспомогательные алгоритмы, представленные в предыдущих параграфах. Данный процесс осуществляется в следующих четырёх этапах:

1-й этап. Выбор и проектирование модели PageRank.

2-й этап. Сравнение итерационных методов (Power Method, Jacobi, Gauss–Seidel).

3-й этап. Разработка соответствующего параллельного алгоритма.

4-й этап. Разработка программного средства в необходимой среде (Python/Java/C++).

Каждый этап может быть реализован на основе соответствующих подходов.

Основной целью классификации исследуемых веб-документов на основе их структуры является выделение различных частей веб-документа. К ним относятся теги (title, meta, headings), навигация, меню, основной текст (body), гиперссылки, DOM-дерево, URL, а также SEO-теги, запросы, изображения и другие элементы. Классификация, основанная на выделении частей страницы с использованием DOM-дерева и анализе их взаимосвязей, обеспечивает более высокую надёжность

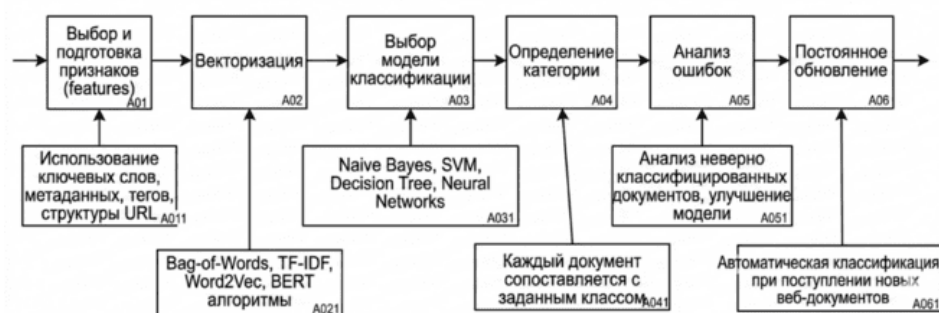


Рисунок 7. Функциональная структура классификации.

Кластеризация - это процесс группировки веб-документов на основе их сходства и структуры без предварительно заданных классов. Данный подход позволяет выделять такие категории, как новости, блоги, рекламные страницы, страницы контактов и другие, среди страниц с одинаковой HTML-структурой. Таким образом, это метод разделения данных на группы на основе их сходства без заранее определённых классов. Данный метод широко применяется в интеллектуальном анализе данных, в частности в Data Mining, Text Mining и Web Mining.

Разработан алгоритм оценки на основе сходства объектов веб-документов, а также общий алгоритм оценки, основанный на их сходстве. В данном алгоритме объединены показатели сходства объектов, количество обращений к объектам, а также сходство на основе количества обращений, в результате чего сформирован общий алгоритм оценки.

При проектировании программного средства интеллектуальной обработки веб-документов с целью сокращения операций анализа и выделения признаков были разработаны алгоритмы, предназначенные для контроля проектирования бытий веб-документов (см. главу 2). Общий алгоритм контроля проектирования событий состоит из 8 подалгоритмов. Логико-функциональная структура программного средства интеллектуальной обработки веб-документов разработана с учётом вышеуказанных элементов следующим образом.



Рисунок 8. Логико-функциональная структура программного средства интеллектуальной обработки веб-документов.

Для анализа и оценки веб-документов на основе их классификации и кластеризации необходимы базы знаний и данных. В первую очередь требуется выбор модели представления знаний. В рамках исследования были рассмотрены теория графов, модель RDF (Resource Description Framework), модель WOL (Web Ontology Language), модель Neo4j и rule-based модель. С

учётom цели исследования было определено, что наиболее эффективным является проектирование формирования и управления знаниями на основе rule-based модели

База знаний, основанная на правилах (rule-based), представляет знания в виде правил типа «IF - THEN». При этом IF (условие) - это определённое состояние, ситуация или совокупность признаков, а THEN (результат) - автоматически выводимое заключение или событие.

Её формализация представляется следующим образом:

Rule: IF Condition(X) THEN Action(Y)

Математическая формализация имеет следующий вид $IF A \rightarrow B, A \text{ TRUE} \Rightarrow B \text{ TRUE}$.

База знаний, основанная на правилах, также может иметь древовидную (иерархическую) структуру. В этом случае правила и знания организуются в виде дерева.

При записи правил в форме [IF – THEN] формируются следующие шаблонные правила. Их можно непосредственно включить в базу правил.

Общий шаблон *IF (shartlar/alomatlar) THEN (xulosa/amal)*.

Формальный (псевдокодovый) шаблон:

IF condition1 AND/OR condition2 AND ... THEN conclusion/action

Шаблон классификации веб-документов:

*IF ("terim" so'zi chastotasi > 5) OR (terim+mavjud) THEN class
= "terimSinf"*

*IF (meta_keywords ichida "terim
+ " bor) AND (title da terim bor) THEN class = "terimSinf"*
*IF (h1 soni ≥ 1) AND (a – teglar soni > 20) THEN document_type
= "news_portal"*

Шаблон, соответствующий структуре веб-документа (DOM):

*IF (h1 mavjud) AND (img soni
≥ 3) AND (caption teglari mavjud) THEN layout
= "article_with_images"*

*IF (nav menyu topildi) AND (footer mavjud) AND (menu_items
≥ 5) THEN site_section = "main_page"*

Шаблон рекомендательных правил

*IF user_interest CONTAINS "machine learning" THEN recommend_category
= "AI/ML articles"*

*IF (user_read_time > 10 min) AND (shares ≥ 1) THEN recommend
= "long_read_content"*

Шаблон для контроля качества или валидации:

*IF (title yo'q) OR (meta_description yo'q) THEN quality_flag
= "low"; action = "needs_edit"*

IF (page_load_time > 3s) THEN performance_status = "slow"

Шаблон для управления доступом (access control):

IF (user_role = "admin") THEN permission = "write, delete"

IF (user_role = "editor") AND (section = "news") THEN permission = "write"

Шаблон цепочки вложенных продукционных правил (nested rules):

IF "terim" belgisi TRUE THEN topic = "terimSinf"

IF (topic = "terimSinf") AND (" + terim" exists) THEN subtopic = " + terimSinf"

На основе данных шаблонов можно формировать произвольные знания для анализа веб-документов, а также осуществлять их дальнейшее совершенствование.

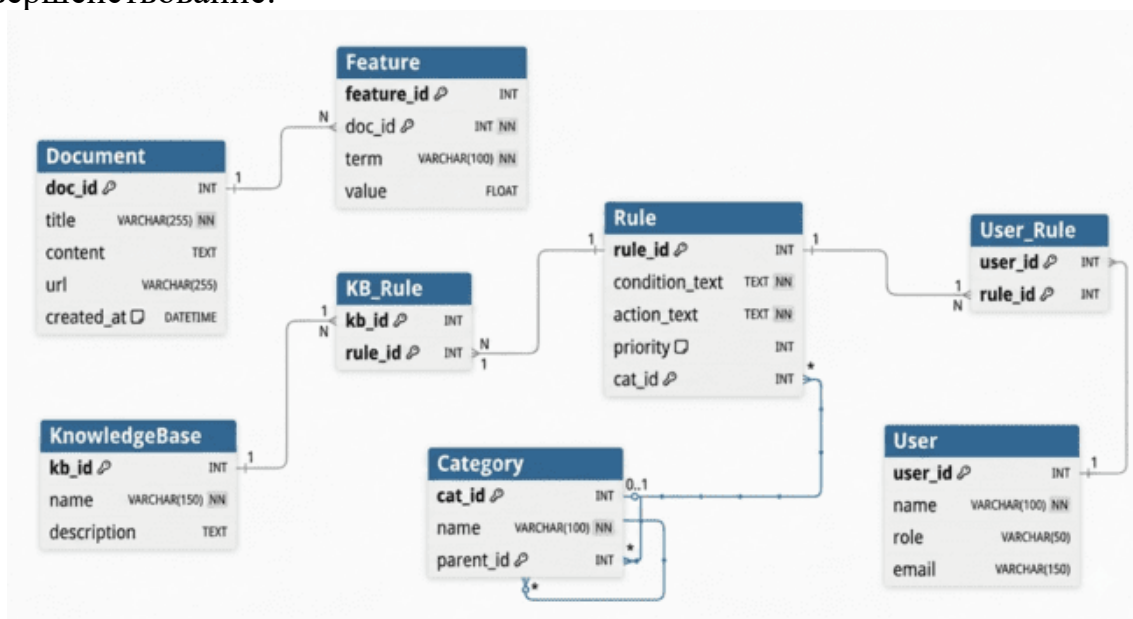


Рисунок 9. ER-диаграмма базы знаний.

Перед уточнением программного обеспечения и требований к анализу сетевых данных необходимо учитывать его теоретические основы. В рамках теории сетей сетевые данные представляются в виде узлов и их связей (рёбер) и анализируются на основе теории графов. При этом используются такие метрики, как степень центральности, посредническая центральность и коэффициент кластеризации. С применением теории информации осуществляется измерение потоков данных, а также анализ энтропии и разнообразия информации в сети. При этом с использованием интеллектуальных технологий применяются такие методы, как модель MapReduce - способ обработки больших объёмов данных в распределённой среде, потоковые алгоритмы, обеспечивающие непрерывную обработку потоков данных при ограниченной памяти, а также параллельные вычисления метод выполнения вычислений одновременно на нескольких процессорах или ядрах. Технологии машинного обучения и искусственного интеллекта применяются для анализа сетевого трафика, включая задачи кластеризации и прогнозирования, при этом целесообразно использование модели GNN (Graph Neural Network).

Разработка программного средства, ориентированного на анализ сетевых данных, имеет важное научное и практическое значение. Основу данного направления составляют теория графов, теория информации, технологии Big

Data, методы машинного и глубокого обучения (ML/DL), а также графовые нейронные сети (GNN). Существует мнение, что в ближайшем будущем подобные программные средства будут играть ключевую роль в развитии кибербезопасности, Интернета вещей (IoT) и интеллектуальных систем.

Программное средство для анализа веб-документов, включая его программное обеспечение, требования, а также подходы к развитию и поддержке, было внедрено в таких организациях, как Ташкентский государственный экономический университет, Ташкентский государственный медицинский университет, Центр управления проектами цифрового правительства и Центр развития цифровых образовательных технологий. При внедрении были соблюдены все технические, организационные и кадровые требования, основанные на критериях, приведённых в предыдущем параграфе. В процессе внедрения приняли участие 4 администратора из 4 организаций, 8 опытных специалистов, обслуживающих сетевую инфраструктуру организаций, представители руководства и профессорско-преподавательского состава, а также пользователи веб-сайтов организаций. Общее число участников превысило 2000 человек.

При определении эффективности программного средства в указанных выше 4 организациях необходимо предварительно определить его параметры. Требования были сформированы с учётом потребностей и ресурсов организаций.

К основным параметрам эффективности относятся:

- T*-параметры технической эффективности;
- F*-параметры функциональной эффективности;
- E*-параметры экономической эффективности;
- U*-параметры пользовательской эффективности.

Специфические характеристики данных параметров эффективности были определены следующим образом. При этом учитывались мнения сотрудников внедряемых организаций, а также стратегия развития организаций..

Общая эффективность программного средства рассчитывается по следующему выражению.

$$Ea_{umumiy} = \frac{1}{4}(w_1T + w_2F + w_3E + w_4U) \quad (5)$$

где *T*- технические показатели, *F*- функциональные показатели, *E*- экономические показатели, *U*- пользовательские показатели, w_i - весовые коэффициенты (определяются на основе экспертной оценки).

T-параметры технической эффективности определяются по следующему выражению:

$$T_{umumiy} = \frac{1}{5}(wt_1T_s + wt_2T_s + wt_3T_s + wt_4T_s + wt_5T_m) \quad (6)$$

где w_{t_i} - весовые коэффициенты (определяются на основе экспертной оценки).

F- параметры функциональной эффективности определяются по следующему выражению:

$$F_{umumiy} = \frac{1}{5}(wf_1F_a + wf_2F_r + wf_3F_p + wf_4F_{AI} + wf_5F_A) \quad (7)$$

где wf_i - весовые коэффициенты (определяются на основе экспертной оценки).

E -Параметры экономической эффективности определяются по следующему выражению:

$$E_{umumiy} = \frac{1}{4}(we_1E_t + we_2E_m + we_3E_e + we_4E_c) \quad (8)$$

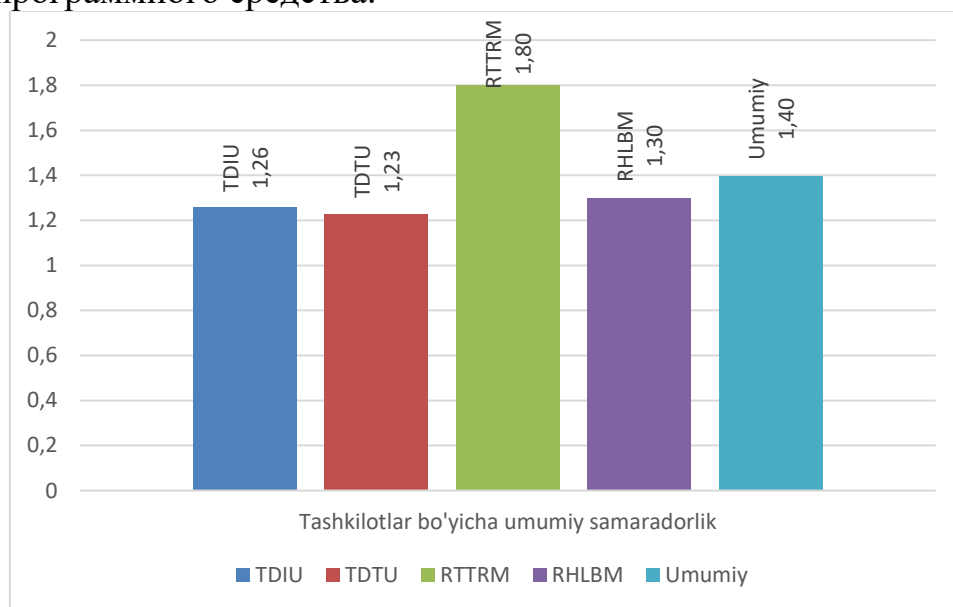
где we_i - весовые коэффициенты (определяются на основе экспертной оценки).

U -Параметры пользовательской эффективности определяются по следующему выражению:

$$U_{umumiy} = \frac{1}{4}(wu_1U_i + wu_2U_l + wu_3U_q + wu_4U_u) \quad (9)$$

где wu_i — весовые коэффициенты (определяются на основе экспертной оценки).

На основе всех вышеуказанных результатов анализа показатели эффективности определяются путём сравнения состояния до и после внедрения программного средства.



10-rasm. Umumiy samaradorlik

Результаты оценки эффективности показали её увеличение: в ТДИУ - в 1,26 раза, в ТДТУ - в 1,23 раза, в РТТРМ - в 1,8 раза, а в РХЛБМ - в 1,3 раза

ЗАКЛЮЧЕНИЕ

По результатам исследований, проведённых в рамках диссертационной работы на тему «Модели и алгоритмы анализа сетевых данных на основе технологий искусственного интеллекта», сформулированы следующие выводы:

1. Разработаны алгоритмы выбора метрик для определения взаимного сходства объектов на основе структуры веб-документов, что позволило эффективно оценивать характеристики DOM-элементов и корректно группировать объекты.

2. Разработаны математическая модель и алгоритмы оценки взаимного сходства объектов веб-документов, обеспечивающие точное представление взаимосвязей между объектами и их эффективную оценку.

3. На основе модели продукционных правил разработаны методы проектирования знаний, ER-модель и алгоритм управления. В предложенном подходе знания формализованы в виде правил IF–THEN, а взаимосвязи между ними системно структурированы с использованием ER-модели. Это позволило обеспечить логическую и семантическую целостность базы знаний.

4. На этапе внедрения программное средство было апробировано в таких организациях, как Ташкентский государственный экономический университет, Ташкентская медицинская академия, Центр управления проектами цифрового правительства и Центр управления цифровыми образовательными технологиями. По результатам испытаний зафиксирован положительный рост технической, функциональной, экономической и пользовательской эффективности системы. Техническая эффективность увеличилась в 1,25 раза, функциональная - в пределах 0,25–0,30 раза, экономическая - до 2,6 раза в организациях, а пользовательская эффективность в среднем увеличилась в 1,4 раза.

В целом в исследовании разработаны математические модели, алгоритмы и подходы к управлению базой знаний на основе правил, ориентированные на интеллектуальный анализ структуры, объектов и событий веб-документов. В результате предложено единое научно обоснованное решение для эффективного анализа веб-документов, выявления сходных объектов и оптимизации систем.

**SCIENTIFIC COUNCIL AWARDING SCIENTIFIC DEGREES
DSc.09/2025.27.12. T.01.03 AT TASHKENT UNIVERSITY OF
INFORMATION TECHNOLOGIES**

**TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES
NAMED AFTER MUHAMMAD AL-KHWARIZMI**

HUSANOV SHERZOD ABDIMANONOVICH

**MODELS AND ALGORITHMS FOR ANALYZING NETWORK DATA
BASED ON ARTIFICIAL INTELLIGENCE TECHNOLOGIES**

Specialty: 05.01.11 – Digital Technologies and Artificial Intelligence

**ABSTRACT OF THE DISSERTATION
SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)
IN TECHNICAL SCIENCES**

Tashkent – 2026

The theme of Doctor of Philosophy (PhD) dissertation in technical sciences was registered at the Supreme Attestation Commission at the Ministry of Higher Education, Science and Innovations of the Republic of Uzbekistan under number B2025.3.PhD/T5888.

The dissertation has been prepared at Tashkent University of Information Technologies named after Muhammad al-Khwarizmi.

The abstract of the dissertation is posted in three languages (Uzbek, Russian, English (resume)) on the website of the Scientific Council (www.tuit.uz) and on the "Ziyonet" Information and educational portal (www.ziyonet.uz).

Scientific adviser:

Muminov Bahodir Boltaevich
Doctor of Technical Sciences, Professor

Official opponents:

Nosirov Khabibullo Khikmatullo ugli
Doctor of Technical Sciences, Professor

Hamroyev Alisher Shodmonqulovich
Doctor of Philosophy in Technical Sciences

Leading organization:

Nukus State Technical University

The defense of dissertation will take place "9" september 2026 at 2^{pm} at the meeting of the Scientific Council No. DSc.09/2025.27.12.T.01.03 at Tashkent University of Information Technologies. (Address: 108, Amir Temur Street, Tashkent city, 100084. Tel.: (+99871) 207-59-42, e-mail: ilmiy_kengash@tuit.uz)

The dissertation is available at the Information Resource Centre of Tashkent University of Information Technologies named after Muhammad al-Khwarizmi (Registration number № 445). (Address: 108, Amir Temur Street, Tashkent city, 100084. Tel.: (+99871) 238-64-15)

The abstract of the dissertation was sent out on "28" august 2026.
(Mailing list report № 24 dated "28" august 2026.)



M.M.Kamilov

Chairman of the Scientific Council for
awarding scientific degrees, Doctor of
Technical Sciences, Academician of the ASU

N.A.Egamberdiev

Scientific Secretary of the Scientific
Council for awarding scientific degrees,
Doctor of Philosophy in Technical Sciences

N.O.Rakhimov

Chairman of the scientific seminar under the
Scientific Council for awarding scientific degrees,
Doctor of Technical Sciences, Docent

R. Nodir

Research Objective. The aim of the study is to develop models and algorithms for analyzing web sites and pages of network data based on artificial intelligence technologies.

Research Object. Web documents and their structural components (HTML elements, metadata, hyperlinks, and content blocks).

Research Subject. Mathematical modeling of the similarity of web document objects, as well as algorithms for determining their semantic and structural relationships.

Scientific Novelty of the Research consists of the following:

- a hybrid algorithm for web document event modeling based on artificial intelligence technologies has been developed;

- algorithms for selecting metrics to determine the similarity between web document objects based on their structural characteristics have been developed;

- a mathematical model and an algorithm for evaluating the similarity of web document objects based on their events have been developed;

- an object-relationship knowledge representation model and a control algorithm based on production rules have been developed.

Practical Results of the Research include:

- the development of models and algorithms for classifying network data based on the structure of web documents, enabling the classification of objects;

- the development and practical implementation of algorithms for evaluating web documents based on hyperlinks;

- the development of a hybrid algorithm designed for controlling the modeling of events, with all events in web documents analyzed and processed;

- the design of a software tool for classification and evaluation based on the structure of web documents.

Structure and Volume of the Dissertation. The dissertation consists of an introduction, four chapters, a conclusion, and a list of references. The total length of the dissertation is 124 pages.

E'LON QILINGAN ISHLAR TO'YXATI
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I bo'lim (I часть; part I)

1. Muminov B. B., Husanov Sh.A., Bekmurodov U.B. Analysis of Automatic Text Editing Methods // International Conference on “Information Science and Communications Technologies: Applications, Trends and Opportunities - ICISCT 2021”. Tashkent-2021. -5p. DOI: 10.1109/ICISCT52966.2021.9670343 (05.00.00, 30.10.2021 y.dagi №525-sonli OAK Rayosati qarori).

2. Muminov B. B., Husanov Sh.A., Tarmoq ma'lumotlariga intellektual shlov berish modellari va usullari // “Raqamli Transformatsiya va Sun'iy Intellekt” ilmiy jurnali. VOLUME 3, ISSUE 3, 2025. -B. 280-288. <https://dtai.uz/articles/tarmoq-malumotlariga-intellektual-ishlov-berish-modellari-va-usullari/> (05.00.00, 04.07.2023 y.dagi №340/3-sonli OAK Rayosati qarori).

3. Muminov B. B., Husanov Sh.A., Veb hujjat kontentlarga yo'naltirilgan veb analitika modellari va algoritmlari // “Raqamli Transformatsiya va Sun'iy Intellekt” ilmiy jurnali. VOLUME 3, ISSUE 4, 2025.-B. 137-140. <https://dtai.uz/articles/veb-hujjat-kontentlarga-yonaltirilgan-veb-analitika-modellari-va-algoritmlari/> (05.00.00, 04.07.2023 y.dagi №340/3-sonli OAK Rayosati qarori).

4. Muminov B. B., Husanov Sh.A., A Mathematical Model and Algorithm for Evaluation Based on the Similarity of Web Document Objects // Journal: “Bulletin of TUIT: Management and Communication Technologies”. № 3(29), 2025. -7p. <https://uzjurnal.uz/2/2025/3/index?issue=29> (05.00.00, 30.07.2020 y.dagi №283/7.1-sonli OAK Rayosati qarori).

5. Muminov B. B., Husanov Sh.A., DESIGNING KNOWLEDGE AND DATA BASES FOR WEBDOCUMENTS BASED ON INTELLECTUAL TECHNOLOGIES // Journal: “Bulletin of TUIT: Management and Communication Technologies”. Vol-4(34). 2026-B. 257-262. <https://uzjurnal.uz/2/2026/4/index?issue=34> (05.00.00, 30.07.2020 y.dagi №283/7.1-sonli OAK Rayosati qarori).

II bo'lim (II часть; part II)

6. Muminov B. B., Husanov Sh.A., Sun'iy intellekt texnologiyalar asosida veb hujjatlar o'xshashligini aniqlash metrikalari va klasterlash algoritmi // “International Conference on Adaptive Learning Technologies”. Vol. 24, Canary Island, Spain-2026. -B. 16-22. <https://top-conferences.us/index.php/ICALT/article/view/3061>

7. Muminov B. B., Husanov Sh.A., Artificial intellect technologies based on web document objects clustering progressive metric models and algorithms // “Modern problems in education and their scientific solutions”. Vol. 2, № 8, England-2026. -P. 226-229. <https://esiconf.org/index.php/MRIATS/article/view/6793>

8. Husanov Sh.A., Algorithms focused on event design based on artificial intelligence technologies // European science international conference: "Theory of scientific researches of whole world". Vol. 2, № 7, Switzerland-2026. -P. 91-94. <https://esiconf.org/index.php/TOSROF/article/view/6794>

9. Husanov Sh.A., Models and Methods of Intelligent Processing of Web Documents Based on Artificial Intelligence Technologies//American Journal of Applied Science and Technology, VOLUME 06, ISSUE 05, 2026. -B. 113-116. <https://theusajournals.com/index.php/ajast/article/view/11158/10235>

10. Husanov Sh.A., Models and Algorithms for Identifying Important Pages Based on User Behavior Analysis on Websites// Hamburg, Germany, International Journal of Advance Scientific Research, VOLUME 06, ISSUE 05, 2026. -B. 43-46. DOI: <https://doi.org/10.37547/ijasr-06-05-06>

11. Husanov Sh.A., Veb-hujjat obyektlari o'xshashligi asosida baholashning matematik modeli va algoritmi // "Yashil iqtisodiyot sharoitida moliya, bank ishi, buxgalteriya hisobi va audit sohasini rivojlantirish istiqbollari" mavzusidagi respublika ilmiy-amaliy konferensiya materiallari. Toshkent-2025. -B. 68-69.

12. Muminov B. B., Husanov Sh.A., Tarmoq ma'lumotlarini tahlil qilish uchun qo'llaniladigan metodlar va dasturiy ta'minot vositalari // "Davlat tili taraqqiyoti: muammo va yechimlar" mavzusidagi respublika an'anaviy III ilmiy-amaliy anjumani materiallari. Toshkent-2025. -B. 207-210.

13. Husanov Sh.A., Massive Open Online Courses // O'zbekiston Respublikasi Adliya vazirligi huzuridagi Intellektual mulk agentligi. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnoma № DGU 07393. 23.12.2019.

14. Husanov Sh.A., O'zbekiston Respublikasi Adliya vazirligi huzuridagi Intellektual mulk agentligi. Elektron hisoblash mashinalari uchun yaratilgan dasturning rasmiy ro'yxatdan o'tkazilganligi to'g'risidagi guvohnoma № DGU 56393. 14.11.2025