TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI HUZURIDAGI ILMIY DARAJALAR BERUVCHI DSc.13/30.12.2019.T.07.02 RAQAMLI ILMIY KENGASH

TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI

HAYDAROV ELSHOD DILSHOD O‘G‘LI

ELEKTRON POCHTA XABARLARINI FILTRLASH VA SPAM XABARLARNI ANIQLASH ALGORITMLARINI ISHLAB CHIQISH

05.01.05 – Axborotlarni himoyalash usullari va tizimlari. Axborot xavfsizligi

TEXNIKA FANLARI BO‘YICHA FALSAFA DOKTORI (PhD) DISSERTATSIYASI AVTOREFERATI

Toshkent-2023

UDK: 004.056.57.

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

Оглавление автореферата диссертации доктора философии (PhD) по техническим наукам

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

TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI HUZURIDAGI ILMIY DARAJALAR BERUVCHI DSc.13/30.12.2019.T.07.02 RAQAMLI ILMIY KENGASH

TOSHKENT AXBOROT TEXNOLOGIYALARI UNIVERSITETI

HAYDAROV ELSHOD DILSHOD O‘G‘LI

ELEKTRON POCHTA XABARLARINI FILTRLASH VA SPAM XABARLARNI ANIQLASH ALGORITMLARINI ISHLAB CHIQISH

05.01.05 – Axborotlarni himoyalash usullari va tizimlari. Axborot xavfsizligi

TEXNIKA FANLARI BO‘YICHA FALSAFA DOKTORI (PhD) DISSERTATSIYASI AVTOREFERATI

Toshkent-2023

Texnika fanlari bo‘yicha falsafa doktori (PhD) dissertatsiyasi mavzusi O‘zbekiston Respublikasi Vazirlar Mahkamasi huzuridagi Oliy attestatsiya komissiyasida B2022.4.PhD/T3425 raqam bilan ro‘yxatga olingan.

Dissertatsiya Toshkent axborot texnologiyalari universitetida bajarilgan.

Dissertatsiya avtoreferati uch tilda (o‘zbek, rus, ingliz (rezume)) Ilmiy kengash veb-sahifasida (www.tuit.uz) va «Ziyonet» Axborot ta’lim portalida (www.ziyonet.uz) joylashtirilgan.

Ilmiy rahbar:	**Xamdamov Rustam Xamdamovich** texnika fanlari doktori, professor
Rasmiy opponentlar:	**Kerimov Kamil Fikratovich** texnika fanlari doktori, dotsent
	Nasrullayev Nurbek Baxtiyorovich texnika fanlari bo‘yicha falsafa doktori, dotsent
Yetakchi tashkilot:	**«UNICON.UZ» - fan – texnika va marketing tadqiqotlar markazi**

Dissertatsiya himoyasi Toshkent axborot texnologiyalari universiteti huzuridagi DSc.13/30.12.2019.T.07.02 raqamli Ilmiy kengashning 2023-yil «___» _____ soat __ dagi majlisida bo‘lib o‘tadi. (Manzil: 100084, Toshkent shahri, Amir Temur ko‘chasi, 108-uy. Tel.: (99871) 238-64-43, e-mail: tuit@tuit.uz).

Dissertatsiya bilan Toshkent axborot texnologiyalari universiteti Axborot-resurs markazida tanishish mumkin (_____ raqam bilan ro‘yxatga olingan.). (Manzil: 100084, Toshkent shahri, Amir Temur ko‘chasi, 108-uy. Tel.: (99871) 238-64-70).

Dissertatsiya avtoreferati 2023-yil «___» _____ da tarqatildi.
(2023-yil «___» _____ dagi ___ raqamli reestr bayonnomasi.)

B.SH. Maxkamov
Ilmiy darajalar beruvchi ilmiy kengash raisi, i.f.d., professor

E.SH. Nazirova
Ilmiy darajalar beruvchi ilmiy kengash ilmiy kotibi, t.f.d., professor

S.K. Ganiyev
Ilmiy darajalar beruvchi ilmiy kengash qoshidagi ilmiy seminar raisi, t.f.d., professor

KIRISH (falsafa doktori (PhD) dissertatsiyasining annotatsiyasi)

Dissertatsiya mavzusining dolzarbligi va zarurati. Dunyodagi barcha mamlakatlar tomonidan tan olingan yagona elektron xat almashish tizimi bu elektron pochta hisoblanadi. Shuning uchun ham elektron pochtaga qaratilgan hujumlar soni yildan yilga ortib bormoqda. Bu hujumlarni asosiy tashkil etuvchilari spam xabarlardir. Kaspersky kompaniyasi ma’lumotlariga ko‘ra, 2021-yilning birinchi choragida davlat tashkilotlariga bo‘lgan noma’lum qo‘ng‘iroqlar va spam xabarlar soni barcha kiruvchi qo‘ng‘iroqlar va xabarlarning 70% ni tashkil etgan. Bu yo‘nalishda xorijiy mamlakatlarda, jumladan, Rossiya Federatsiyasi, AQSH, Janubiy Koreya, Yaponiya, Malayziya, Xitoy va boshqa davlatlarda elektron pochta xabarlarini filtrlash va ulardan spam xabarlarni aniqlab olish usullari va algoritmlarini ishlab chiqish hamda korxonaning elektron pochta serverini himoyalash tizimlarini takomillashtirish muhim ahamiyat kasb etmoqda.

Jahonda elektron pochta xabarlarini filtrlash va spam xabarlarni aniqlashning usul va algoritmlarini takomillashtirishga hamda intellektual usullar yordamida spam xabarlarni aniqlashga qaratilgan ilmiy-tadqiqot ishlari olib borilmoqda. Bu borada, jumladan ro‘yxatga va lingvistikaga asoslangan spam xabarlarni aniqlash usullarini ishlab chiqish muhim vazifalardan biri hisoblanmoqda. Shu bilan birga, elektron pochta xabarlarini filtrlash jarayonida yangi xabarlar to‘g‘risidagi bilimlar to‘liq bo‘lmaganda yoki spam xabarlarning joriy ma’lumotlar bazasida noaniqliklar bilan bir qatorda, spam xabarlar signaturasi mavjud bo‘lmaganda elektron pochtani himoyalash uchun intellektual usullardan foydalanib, himoya mexanizmini ishlab chiqish zarur bo‘lmoqda.

Respublikamizda davlat va xo‘jalik boshqaruv organlarida elektron pochta xabarlarini filtrlash va ulardan spam xabarlarni aniqlab olishga qaratilgan keng qamrovli chora-tadbirlar amalga oshirilmoqda.

2022-2026-yillarga mo‘ljallangan Yangi O‘zbekistonning Taraqqiyot strategiyasi to‘g‘risida, jumladan «... «UZ» domen zonasi Internet-makonining kiberxavfsizligini ta’minlashning asosiy yo‘nalishlarini hamda elektron hukumat, energetika, raqamli iqtisodiyot tizimlarini va muhim axborot infratuzilmasiga taalluqli boshqa yo‘nalishlarni himoya qilish bo‘yicha kompleks vazifalarni belgilash» rejalashtirilgan. Mazkur vazifalarni amalga oshirishda, elektron pochtadagi xabarlarni filtrlash hamda ulardan spam xabarlarni aniqlab olishning intellektual usullarini qo‘llash va elektron pochtani himoyalashda zamonaviy tizimlardan foydalanish muhim vazifalardan biri hisoblanadi.

O‘zbekiston Respublikasi Prezidentining 2022-yil 28-yanvardagi PF-60-son «2022-2026-yillarga mo‘ljallangan Yangi O‘zbekistonning Taraqqiyot strategiyasi to‘g‘risida»gi, 2018-yil 19-fevraldagi PF-5349-son «Axborot texnologiyalari va kommunikatsiyalari sohasini yanada takomillashtirish chora-tadbirlari to‘g‘risida»gi farmonlari, 2018-yil 21-noyabrdagi PQ-4024-son «Axborot texnologiyalari va kommunikatsiyalarining joriy etilishini nazorat qilish, ularni himoya qilish tizimini takomillashtirish chora-tadbirlari to‘g‘risida»gi va 2019-yil 14-sentabrdagi PQ-4452-son «Axborot texnologiyalari va kommunikatsiyalarining joriy etilishini nazorat qilish, ularni himoya qilish tizimini takomillashtirishga oid

qo‘shimcha chora-tadbirlar to‘g‘risida»gi qarorlari hamda mazkur faoliyatga tegishli boshqa meyoriy-huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishga mazkur dissertatsiya tadqiqoti ma’lum darajada xizmat qiladi.

Tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor yo‘nalishlariga mosligi. Mazkur tadqiqot respublika fan va texnologiyalar rivojlanishining IV. «Axborotlashtirish va axborot-kommunikasiya texnologiyalarini rivojlantirish» ustuvor yo‘nalishi doirasida bajarilgan.

Muammoning o‘rganilganlik darajasi. Elektron pochta xavfsizligini ta’minlashda xabarlarni filtrlash, tayanch vektorlar usuli va shunga o‘xshash intellektual usullarni qo‘llagan holda spam xabarlarni aniqlash uchun ishlab chiqilgan usullar va algoritmlarni axborot xavfsizligini ta’minlashda qo‘llash bo‘yicha E.Dada, J.Bassi, T.S. Guzella, V. Zorkadis, K. Tretyakov, Vaughan-Nikols, A.Kumari va boshqa chet ellik olimlar tomonidan ilmiy izlanishlar olib borilmoqda. Elektron pochta xavfsizligini ta’minlashda tayanch vektorlarni qo‘llagan holda S. Lin, Y. Zhou, K. Yang, L.Chen, R.Krestel ilmiy izlanishlar olib borishgan. Bundan tashqari, Kaspersky, ESET, AGAVA SpamProtexx va Trend Micro tashkilotlari tomonidan intellektual usullarni qo‘llash orqali elektron pochta xabarlarini filtrlash va ulardan spam xabarlarni aniqlashning dasturiy-apparat vositalarini ishlab chiqish bo‘yicha muxandislik-tadqiqot ishlari olib borilmoqda.

O‘zbekistonda akademik T.F.Bekmuratov tomonidan elektron pochta xabarlarini filtrlash va ulardan spam xabarlarni aniqlashning usul va algoritmlari ishlab chiqish bo‘yicha ilmiy izlanishlar olib borilgan va hozirda S.K.Ganiyev, R.X.Xamdamov, M.M.Karimovlar boshchiligidagi ilmiy jamoalar tomonidan ilmiy izlanishlar olib borilmoqda.

Shu bilan birga, tayanch vektorlar, neyron tarmoq, mashinali o‘qitish (machine learning) va chuqur o‘qitish (deep learning) asosida spam xabarlarni aniqlash, bartaraf etish usullari va algoritmlari yetarlicha tadqiq etilmagan.

Dissertatsiya tadqiqotining dissertatsiya bajarilgan oliy ta’lim muassasasining ilmiy-tadqiqot ishlari rejalari bilan bog‘liqligi. Dissertatsiya tadqiqoti Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universitetining ilmiy-tadqiqot ishlari rejasining 598661-EPP-1-2018-1-RO-EPPKA2-CBHE-JP «Developing Services for Individuals with Disabilities» mavzusidagi loyiha doirasida bajarilgan.

Tadqiqotning maqsadi elektron pochtani himoyalash tizimining samaradorligini oshirishga imkon beruvchi spam xabarlarni aniqlash usul va algoritmlarini ishlab chiqishdan iborat.

Tadqiqotning vazifalari:

elektron pochta xabarlarini filtrlash usullarini qiyosiy tahlil qilish va filtrlashda zarur bo‘lgan mezonlarni tanlash;

filtrlash tizimining bilimlar bazasini to‘ldirish algoritmini ishlab chiqish;

elektron pochtadagi hujjatlarni filtrlash algoritmini ishlab chiqish;

spam xabarlarni aniqlashning modifikatsiyalangan tayanch vektor usuli va algoritmini ishlab chiqish;

logistik regressiya asosida spam hujjatlarni aniqlash algoritmini ishlab chiqish.

Tadqiqotning obyekti sifatida elektron pochtadagi spam xabarlar olingan.

Tadqiqotning predmetini elektron pochtani filtrlash va spam xabarlarni aniqlash usul va algoritmlari tashkil etadi.

Tadqiqotning usullari. Tadqiqot jarayonida elektron pochtani filtrlash va spam xabarlarni aniqlash usullari, neyron tarmoqlar, mashinali o‘qitish, ehtimollik nazariyasi, diskret matematika va obyektga yo‘naltirilgan dasturlash usullaridan foydalanilgan.

Tadqiqotning ilmiy yangiligi quyidagilardan iborat:

barcha turdagi spam xabarlari uchun umumiy bo‘lgan belgilarni ajratib olish natijasida keng imkoniyatli va ko‘p xususiyatli elektron pochta xabarlari uchun umumiy bo‘lgan mezonlar shakllantirilgan;

tashkilotdagi mavjud bilimlar bazasidagi bilimlardan foydalanish chastotasini hisobga olgan holda eng ko‘p foydalaniladigan bilimlar guruhi shaklantirilgan, natijada real vaqt rejimida yangi turdagi spam xabarlarni aniqlash imkonini beradigan bilimlar bazasini to‘ldirish algoritmi ishlab chiqilgan;

elektron pochtadagi hujjat turlarini hisobga olgan holda ushbu hujjatlar orasidan fayl kengaytmasi asosida ko‘p darajali bilimlar bazasi shakllantirilgan, natijada spam xabarlarni aniqlash tizimining ko‘p darajali bilimlar bazasi yordamida uch bosqichdan iborat elektron pochtadagi hujjatlarni filtrlash algoritmi ishlab chiqilgan;

o‘quv tanlanma bilan o‘qitish usuliga asoslangan holda spam xabarlarni aniqlashning modifikatsiyalangan tayanch vektor usuli va algoritmi ishlab chiqilgan;

logistik regressiya usulini moslashuvchan tasodifiy qidiruv usuli bilan umumlashtirgan holda xabarlar orasidan eng ko‘p spam xabar bo‘lish ehtimolligini topish orqali spam hujjatlarni aniqlash algoritmi ishlab chiqilgan.

Tadqiqotning amaliy natijasi quyidagilardan iborat:

elektron pochta xabarlarini filtrlash usullari asosida spam xabarlarni aniqlashning dasturiy vositasi ishlab chiqilgan;

o‘quv tanlanma bilan o‘qitish asosida spam xabarlarni aniqlashning tayanch vektor usuli takomillashtirilgan.

Tadqiqot natijalarining ishonchliligi. Tadqiqot natijalarining ishonchliligi tashkilot elektron pochtasini filtrlash, bilimlar bazasini to‘ldirish, spam xabarlarni aniqlash usullaridan turli sharoitlarda olingan real hamda tajribaviy tahlil natijalari bilan izohlanadi.

Tadqiqot natijalarining ilmiy va amaliy ahamiyati. Tadqiqot natijalarining ilmiy ahamiyati ishlab chiqilgan yangi turdagi spam xabarlarni aniqlash imkonini beradigan bilimlar bazasini to‘ldirish algoritmi, uch bosqichdan iborat elektron pochtadagi hujjatlarni filtrlash algoritmi va spam xabarlarni aniqlashning modifikatsiyalangan tayanch vektor usuli va algoritmi hamda spam hujjatlarni aniqlash algoritmi bilan izohlanadi.

Tadqiqot natijalarining amaliy ahamiyati taklif etilgan usullar va algoritmlar asosida ishlab chiqilgan dasturiy vositaning tashkilotdagi elektron pochta xabarlarini filtrlash va spam xabarlarni aniqlash jarayonini samaradorligini ortishiga ko‘maklashishi bilan izohlanadi.

Tadqiqot natijalarining joriy qilinishi. Elektron pochta xabarlarini filtrlash va spam xabarlarni aniqlash algoritmlari hamda dasturiy vositalari bo‘yicha olingan ilmiy natijalar asosida:

elektron pochta xabarlarini filtrlash va spam xabarlarni aniqlash algoritmlari asosida ishlab chiqilgan dasturiy vosita «Kiberxavfsizlik markazi» davlat unitar korxonasining amaliy faoliyatiga joriy qilingan (Axborot texnologiyalari va kommunikatsiyalarini rivojlantirish vazirligining 2022-yil 12-oktabrdagi 33-8/6761-son ma’lumotnomasi). Ilmiy tadqiqot natijasi korporativ tarmoqdagi 4673 ta elektron pochta xabarlarini filtrlab, ulardagi 1364 ta spam xabardan 1327 tasini 97.3% aniqlik qayd etgan holda aniqlash imkonini bergan;

elektron pochta xabarlarini filtrlash va spam xabarlarni aniqlash algoritmlari asosida ishlab chiqilgan dasturiy vosita «UZINFOCOM Davlat axborot tizimlarini yaratish va qo‘llab-quvvatlash bo‘yicha yagona integrator» ma’suliyati cheklangan jamiyatining amaliy faoliyatiga joriy qilingan (Axborot texnologiyalari va kommunikatsiyalarini rivojlantirish vazirligining 2022-yil 12-oktabrdagi 33-8/6761-son ma’lumotnomasi). Ilmiy tadqiqot natijasida tashkilotning lokal tarmog‘idagi 3624 ta elektron pochta xabarlarini samarali filtrlab, ulardagi 1465 ta spam xabardan 1431 tasini 97.6% aniqlik qayd etgan holda aniqlash imkonini bergan;

elektron pochta xabarlarini filtrlash va spam xabarlarni aniqlash algoritmlari asosida ishlab chiqilgan dasturiy vosita «ASR KABEL» ma’suliyati cheklangan jamiyatining amaliy faoliyatiga joriy qilingan (Axborot texnologiyalari va kommunikatsiyalarini rivojlantirish vazirligining 2022-yil 12-oktabrdagi 33-8/6761-son ma’lumotnomasi). Ilmiy tadqiqot natijasida korporativ tarmoqdagi 2526 ta elektron pochta xabarlarini samarali filtrlab, ulardagi 1280 ta spam xabardan 1248 tasini 97.5% aniqlik qayd etgan holda aniqlab ularni muvaffaqiyatli bloklash imkonini bergan.

Tadqiqot natijalarining aprobatsiyasi. Mazkur tadqiqot natijalari 2 ta xalqaro va 7 ta respublika ilmiy-amaliy anjumanlarida muhokamadan o‘tkazilgan.

Tadqiqot natijalarining e’lon qilinganligi. Dissertatsiyaning mavzusi bo‘yicha jami 24 ta ilmiy ish chop etilgan, jumladan, O‘zbekiston Respublikasi Oliy attestasiya komissiyasining dissertatsiyalarning asosiy ilmiy natijalarini chop etish tavsiya etilgan ilmiy nashrlarida 6 ta maqola, shulardan, 3 tasi xorijiy va 3 tasi respublika jurnallarida nashr etilgan hamda 4 ta EHM uchun yaratilgan dasturiy vositalarni qaydlash guvohnomalari olingan.

Dissertatsiyaning tuzilishi va hajmi. Dissertatsiya tarkibi kirish, to‘rtta bob, xulosa, foydalanilgan adabiyotlar ro‘yxati va ilovalardan iborat. Dissertatsiya hajmi 112 betni tashkil etadi.

DISSERTATSIYANING ASOSIY MAZMUNI

Kirish qismida dissertatsiya mavzusining dolzarbligi va zaruriyati asoslangan, tadqiqotning O‘zbekiston Respublika fan va texnologiyalari rivojlanishining ustuvor yo‘nalishlariga mosligi ko‘rsatilgan, maqsad va vazifalari belgilab olingan hamda tadqiqot obyekti va predmeti aniqlangan, olingan natijalarning ishonchliligi asoslab

berilgan, ularning nazariy va amaliy ahamiyati, tadqiqot natijalarini amalda joriy qilish holati, nashr etilgan ishlar va dissertatsiyaning tuzilishi bo‘yicha ma’lumotlar keltirilgan.

Dissertatsiyaning **«Elektron pochta xabarlarini filtrlash va spam xabarlarni aniqlash muammolari»** deb nomlangan birinchi bobida elektron pochta xabarlarini filtrlash usullari va ularning qiyosiy tahlili hamda elektron pochta xabarlarini filtrlashda zarur bo‘lgan mezonlarni tanlash va elektron pochta xabarlaridan spam xabarlarni aniqlab olishdagi muammolar hamda ularni yechimlari bo‘yicha tavsiyalar keltirilgan.

Birinchi paragrafda elektron pochta xizmatlari turlari, spam xabarlarning ko‘rinishlari, elektron pochtani filtrlash usullarini turli xususiyatlar, parametrlar va muhitlar bo‘yicha qiyosiy tahlili amalga oshirilgan va keng imkoniyatli va ko‘p xususiyatli ekanligini hisobga olgan holda qulayligi, universalligi, amalga oshirilishi, to‘planuvchanligi, vaqt sarfi, takrorlanmasligi va aniqligi bo‘yicha tayanch vektorlar usulining samaradorligi yuqori ekanligi aniqlangan.

Ikkinchi paragrafda spam xabarlarining barcha ko‘rinishlari uchun umumiy bo‘lgan, ya’ni spam xabarni qaysidir bir ko‘rinishida mavjud boshqa bir turi uchun mavjud bo‘lmagan yoki spam xabar turlari, ularning kelib chiqishi va tarqalishi hamda ularni aniqlashdagi zarur omillarni hisobga olgan holda filtrlash tizimining ortiqcha yuklanishlar soni oshib ketmasligini ta’minlash va yuqori samaradorlikga erishish uchun kerakli mezonlar tanlab olingan. Shuningdek spam omillari va ularni antispam usullari bilan o‘zaro bog‘liqligi keltirilgan.

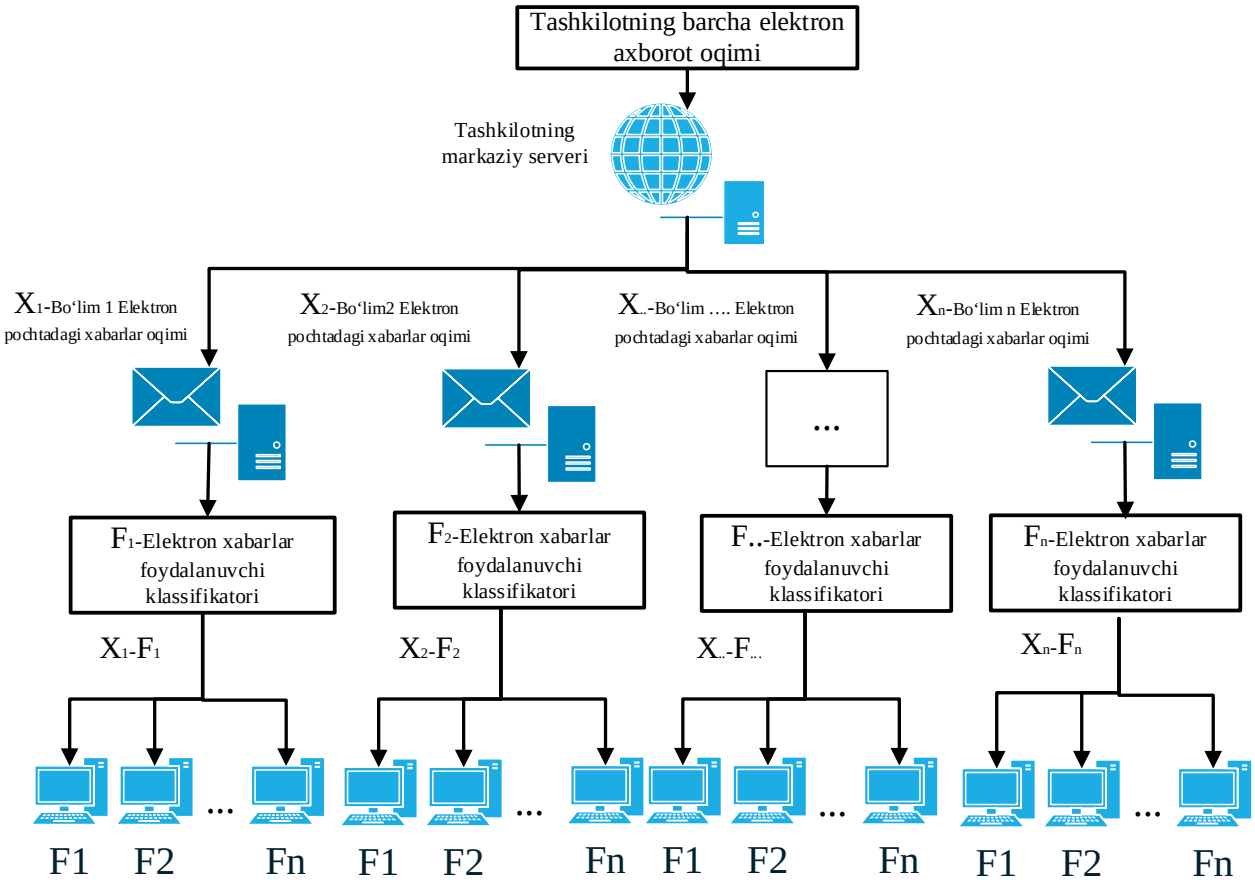
Uchinchi paragrafda elektron pochta xizmatida filtrlash tizimini amalga oshirish bosqichlari va elektron pochta xabarlarni filtrlash va ulardan spam xabarlarni aniqlab olishning amaldagi tizimlaridagi mavjud kamchiliklar va filtrlash jarayonidagi muammolar aniqlangan, DNS yozuvlarini tekshirish va domenlarni diagnostika qilish uchun ishlab chiqilgan himoya mexanizmlari, spamga qarshi kurashish bo‘yicha ishlab chiqilgan qoidalar keltirilgan hamda ushbu muammolarni bartaraf etish bo‘yicha tavsiyalar berilgan.

Dissertatsiyaning **«Elektron pochta tizimining bilimlar bazasini to‘ldirish va hujjatlarni filtrlash algoritmlari»** deb nomlangan ikkinchi bobida elektron pochta xabarlaridagi spam xabarlarni filtrlash iyerarxiyasi hamda ushbu iyerarxiyaga mos holda elektron pochta xabarlarini filtrlash tizimining bilimlar bazasini to‘ldirish va elektron pochtadagi hujjatlarni filtrlash algoritmlari ishlab chiqilgan.

Mazkur bobning *birinchi paragrafida* tashkilotning axborot kanali orqali keladigan barcha axborotni qabul qilish, qayta ishlash, uzatish va saqlash xizmatlarini amalga oshiruvchi tizim va shu tizimning aloqa kanali orqali kirib kelgan barcha elektron pochta xabarlarini filtrlash va spam xabarlarni aniqlab olish hamda tizimga kiruvchi va chiquvchi xabarlarni guruhlarga ajratgan holda ishlash imkonini beradigan elektron pochta xabarlaridagi spam xabarlarni filtrlash iyerarxiyasi taklif etilgan.

Ikkinchi paragrafda tashkilotdagi mavjud bilimlar bazasini taqsimlash orqali yangi turdagi spam xabarlarni aniqlash imkonini beradigan bilimlar bazasiga yangi bilimlarni real vaqt rejimida qo‘shish, ya’ni bilimlar bazasini to‘ldirish algoritmi

ishlab chiqilgan. Bilimlar bazasida bilimlar soni keskin ortib ketgan holda yuborilgan so‘rovlarga javob qaytarish vaqti ortadi. Bu muammoni oldini olish maqsadida spam xabarlarni aniqlash tizimining ko‘p darajali tizimini bilimlar bazasining iyerarxiyasi shakllantirilgan. 1, 2, 3 formulalarda bilimlar bazasiga yangi bilimlarni qo‘shish keltirilgan.



1-rasm. Spam filtrlarni tarmoqni turli tugunlarida iyerarxik joylashgan tizimlaridagi axborot oqimi

$$B_1 = b_{11} \cup b_{12} \cup \ldots \cup b_{1n} = \cup_{k=1}^{n} b_{1k} \quad (1)$$

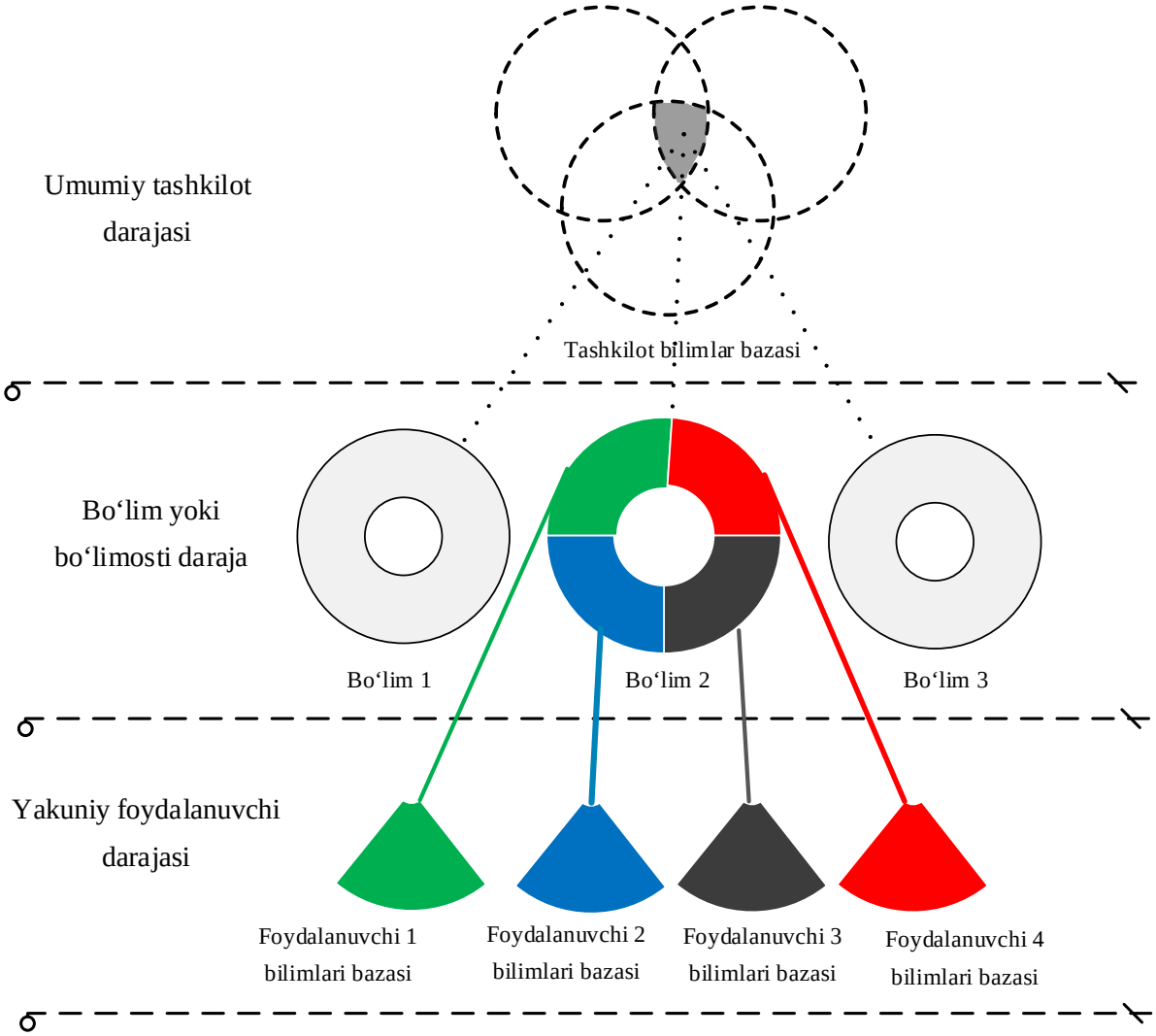
$$B_B = B_1 \cap B_2 \cap \ldots \cap B_p = \cap_{k=1}^{p} B_k \quad (2)$$

$$B_B = (b_{11} \cup b_{12} \cup \ldots \cup b_{1n}) \cap (b_{21} \cup b_{22} \cup \ldots \cup b_{2m}) \cap \ldots \cap (b_{p1} \cup b_{p2} \cup \ldots \cup b_{po}) = \cap_{k=1}^{p} \left(\cup_{j=1, j \neq k}^{r} b_{k,j} \right) \quad (3)$$

2-rasmda spam xabarlarni aniqlash ko‘p darajali tizimining bilimlar bazasini shakllantirish iyerarxiyasi keltirilgan.

Uchinchi paragrafda spam xabarlarni aniqlash tizimining ko‘p darajali tizimini bilimlar bazasi iyerarxiyasiga mos holda elektron pochtadagi hujjatlarni filtrlash algoritmi ishlab chiqilgan (3-rasm). Buning natijasida spam xabarlarni filtrlash tizimi elektron xabarni spam xabarga tegishli ekanlik holatini yakuniy ehtimolligini guruhga ajratgan holda baholash imkonini beradi. XOB<0,5 0,5<=XOB<= 0,9 XOB> 0,9.

Dissertatsiya ishining «**Spam xabarlarni aniqlashning intellektual usul va algoritmlari**» nomli uchinchi bobida mashinali o‘qitishga asoslangan spam xabarlarni aniqlash sxemasi, spam xabarlarni aniqlashning modifikatsiyalangan tayanch vektorlar usul va algoritmi hamda logistik regressiya asosida spam hujjatlarni aniqlash algoritmi ishlab chiqilgan.



2-rasm. Spam xabarlarni aniqlash ko‘p darajali tizimining bilimlar bazasini shakllantirish iyerarxiyasi

Birinchi paragrafda amaldagi mavjud usullar asosida spam xabarlarni aniqlashda foydalaniladigan neyron tarmoqning kirish parametrlariga statistik bo‘lgan va statistik bo‘lmagan belgilarni kiritish orqali elektron pochta xabarlarini tasniflash taklif qilingan. Ushbu tasniflashga mos holda elektron pochta xabarlarini filtrlashning intellektual tizimi va mashinali o‘qitishga asoslangan spam xabarlarni aniqlash sxemasi shakllantirilgan (4-rasm).

Elektron pochta xabarlarini filtrlashning neyron tarmog‘ining samarali modelini (neyron tarmoq turini) to‘g‘ridan-to‘g‘ri qurish quyidagi bosqichlarni o‘z ichiga olgan ma’lumotlar bazasida bilimlarni aniqlash texnologiyasidan foydalanish orqali amalga oshiriladi:

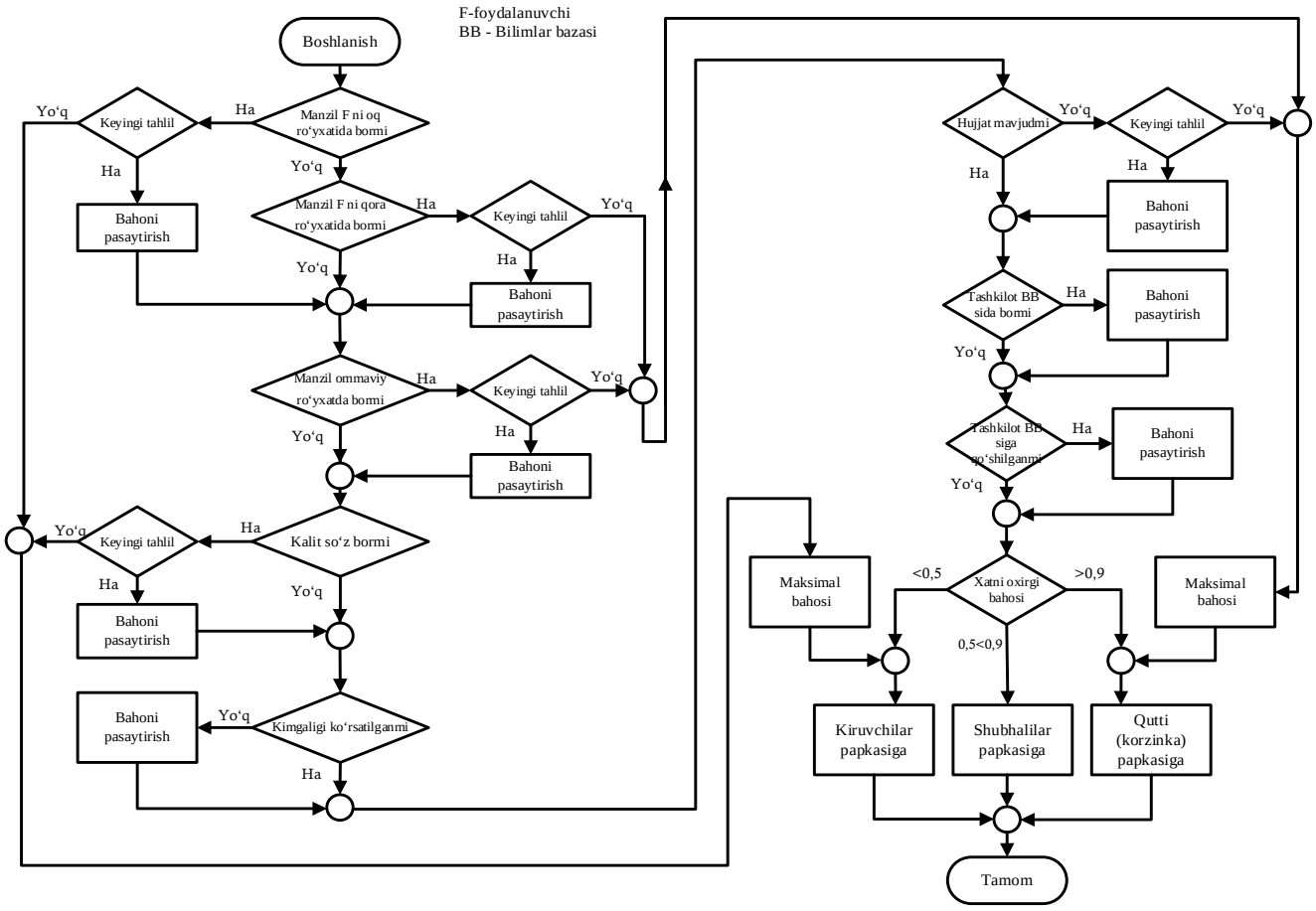
1. Elektron pochta xabarlarining dastlabki ma’lumotlarini, shu jumladan spam va spam bo‘lmagan elektron pochta xabarlarining namunalarini olish;

2. Dastlabki ma’lumotlarni oldindan qayta ishlash va neyron tarmoqni o‘qitish uchun o‘quv namunasini shakllantirish;

3. Neyron tarmoq tuzilmasini ishlab chiqish: har bir qatlamda kirishlar, chiqishlar, tarmoq qatlamlari va neyronlarning sonini belgilash;

4. Spam xabarlarni filtrlash modelini yaratish uchun neyron tarmoqni o‘rgatish;

5. Spam xabarlarni filtrlashning neyron tarmoqli modelini sinab ko‘rish va baholash.



3-rasm. Hujjatlarni filtrlash algoritmining blok sxemasi

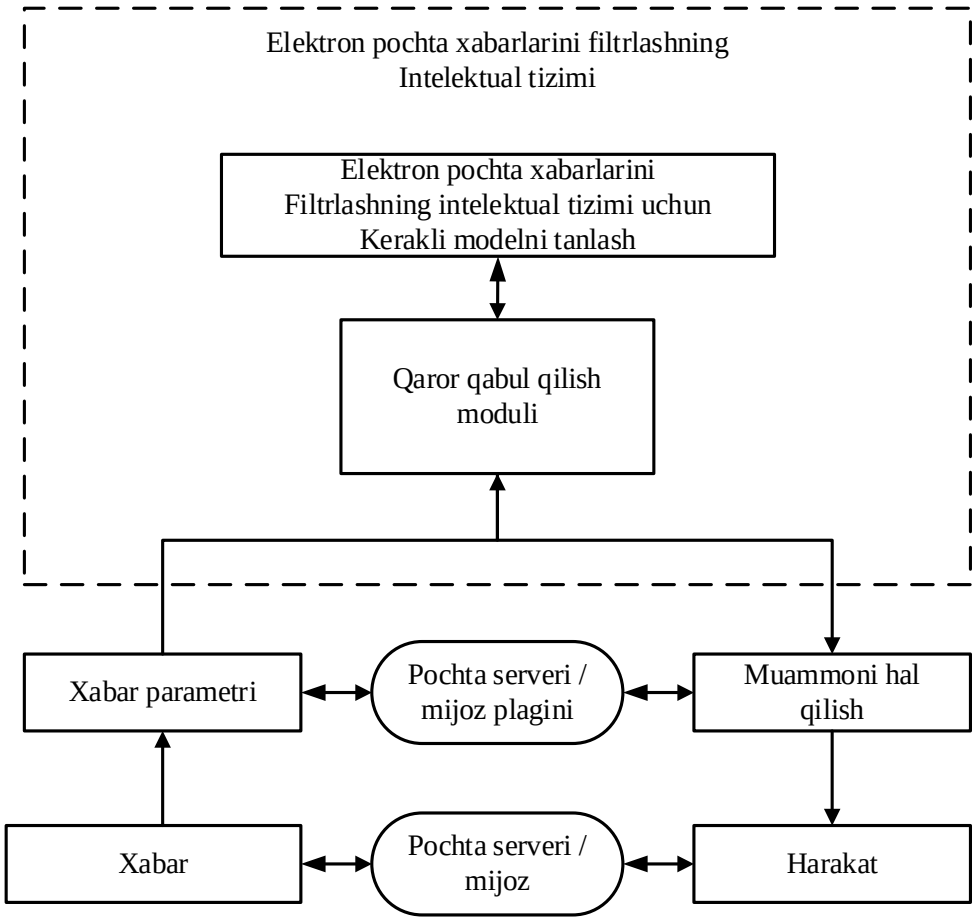
Ikkinchi paragrafda spam xabarlarni aniqlashning tayanch vektorlar usuli yordamida o‘quv tanlanma bilan o‘qitish, ya’ni har bir kategoriya uchun qaysi kategoriyada ekanligi oldindan ma’lum bo‘lgan namunalar hisobga olingan holda o‘qitish va hujjatlarni aniqlangan tavsifi asosida induktiv tarzda o‘qitish natijasida klassifikator yordamida ikki toifaga ajratib olish imkonini beradigan spam xabarlarni aniqlashning modifikatsiyalangan tayanch vektorlar usuli va algoritmi ishlab chiqilgan(5-rasm).

Elektron pochtaning ixtiyoriy x xabari(spam yoki ham)ni n-o‘lchovli belgilar tizimi (vektori) yordamida ifodalaniladi:

$$x=(x^1,x^2,\dots,x^n). \quad (4)$$

Elektron pochtaning ikki sinf xabarlari(spam yoki ham)ni quyidagi *n*-o‘lchovli belgilar tizimi bilan ifodalangan obyektlar (xabarlar) to‘plami ko‘rinishida ifodalab olinadi:

$$x_{pi} = (x_{pi}^1, x_{pi}^2, \ldots, x_{pi}^n), \quad p = 1, 2, \ldots, m; \; i = 1, 2, \ldots, k_p \tag{5}$$



4-rasm. Elektron pochta xabarlarini filtrlashning intellektual tizimi

bu yerda x_{pi}^j – *r*-sinfdagi *i*-obyektga tegishli *j*-belgining son qiymati, *m*-berilgan sinflar soni, k_p – *r*-sinfdagi obyektlar soni.

$$x_i = \begin{cases} \text{spam}, & agar \quad \sum_{j=1}^{6} x_i^j \geq b_q \; yoki \; (x_i^7 > 0.06 \; yoki \; x_i^8 > 0.03\;) \\ \text{ham}, & boshqa \; hollarda \end{cases} \tag{6}$$

Klassifikator sifatida tayanch vektorlar usulida ikkala sinfni optimal tarzda R^n fazoda ajratuvchi gipertekislikning

$$\omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n + \omega_0 = 0 \tag{7}$$

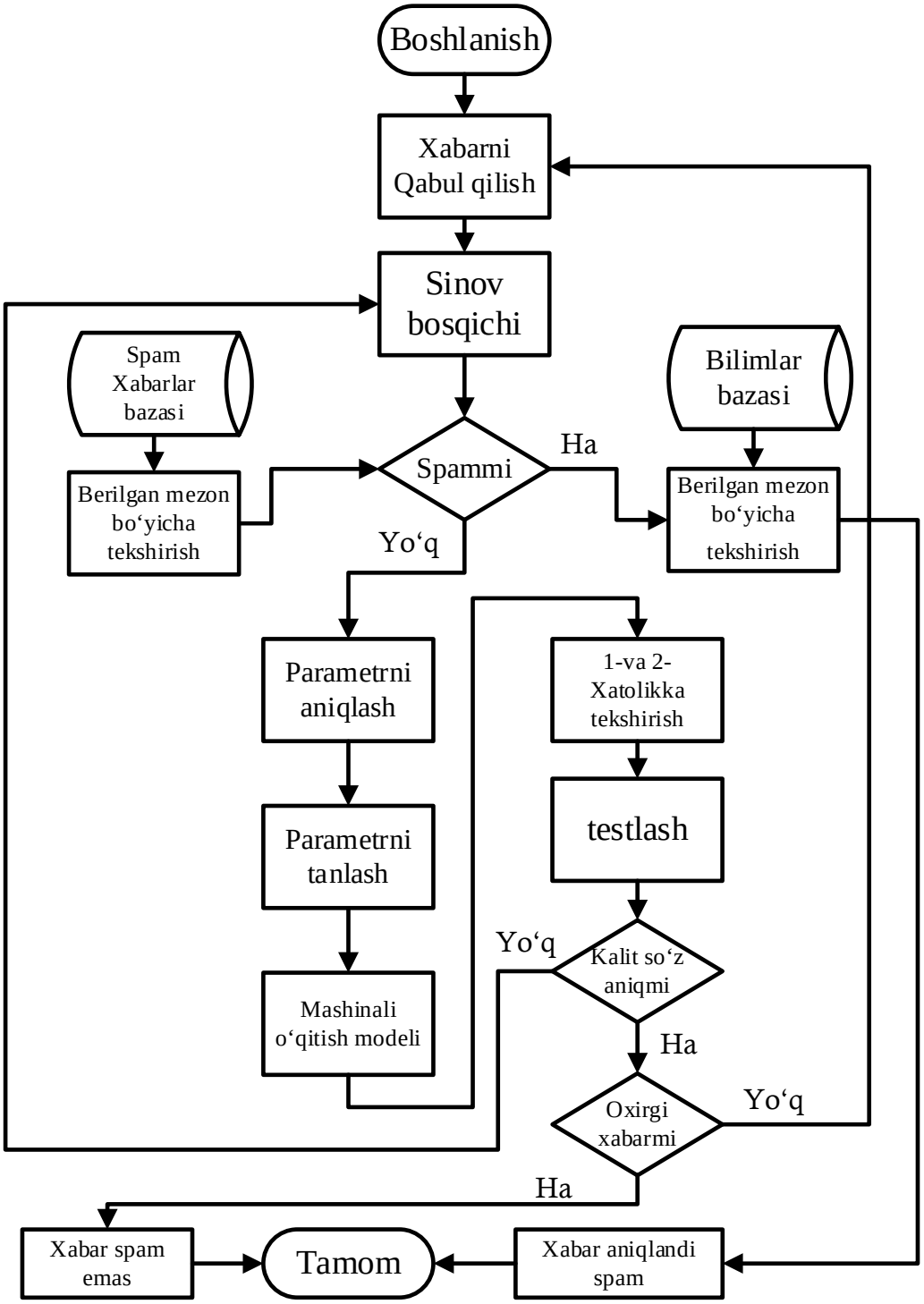
(7) tenglamasini topish hisoblanadi. *x* obyektining *F* funksiyasini *Y* sinf yorlig‘iga aylantirishning umumiy ko‘rinishi:

$$F(x) = sign(\omega^T x - b). \tag{8}$$

Bu yerda

$$\omega = (\omega_1, \omega_2, \ldots, \omega_n), \quad b = -\omega_0 \tag{9}$$

belgilashlar kiritilgan. ω va b o‘qitish algoritm vazn koeffitsiyentlari (noma’lum parametrlari) sozlangandan so‘ng, qurilgan gipertekislikning bir tomoniga tushgan barcha obyektlar birinchi sinf, ikkinchi tomoniga tushgan obyektlar esa ikkinchi sinf sifatida belgilanadi.



5-rasm. Spam xabarlarni aniqlashning modifikatsiyalangan tayanch vektorlar algoritmining blok sxemasi

Uchinchi paragrafda logistik regressiya usulini L.A.Rastriginning moslashuvchan tasodifiy qidiruv usuli bilan qo‘llanilgan holda xabar xususiyatlari ichidan eng ko‘p spam bo‘lish ehtimolligini topish orqali tasniflash vazifasini yechish ishonchliligini va spam xabarni aniqlash tizimining tezligini oshirish imkoniga ega bo‘lingan.

Logistik regressiya (LR) modeli:

$$f_{w,b}(x) = \frac{1}{1+e^{(wx-b)}} \tag{10}$$

uchun optimizatsiyalash, ya’ni maksimallashtirish kriteriyasini quyidagi ko‘rinishda yozib olinadi:

$$q(w,b) = \sum_{i=1}^{N} [y_i \ln f_{wb}(x_i) + (1-y_i)\ln(1-f_{wb}(x_i))] \rightarrow \underset{w,b}{max} =>$$

$$(w*, b*) \tag{11}$$

Bu yerda

$$y_i = \begin{cases} 1, & i = 1,2,\dots,N_1 \\ 0, & i = N_1+1,\dots,N \end{cases} \tag{12}$$

(10),(12) shartlardan foydalanib, (11) optimallashtirish kriteriyasini quyidagi ko‘rinishda yozib olish mumkin:

$$q(w) = \sum_{i=1}^{N} \ln\left(1 + e^{\sum_{j=0}^{n} w_j x_i^j}\right) - \sum_{i=N_1+1}^{N} \sum_{j=0}^{n} w_j x_i^j \tag{13}$$

Bu yerda

$\sum_{j=0}^{n} w_j\, x_i^j = wx_i + b, i = 1,2,\dots,N;\ x_i^0 = 1,\ w_0 = \text{b}.$

Logistik regressiya modelining mohiyatini hisobga olgan holda optimallashtirish muammosi quyidagicha ifodalanishi mumkin:

$$q(w) \rightarrow \underset{w}{max} =>\ w^*, \tag{14}$$

$$f_w(x) = \frac{1}{1+e^{-\sum_{j=0}^{n} w_j x_i^j}} \geq \frac{1}{2},\ i = 1,2,\dots,N_1 \tag{15}$$

$$f_w(x) = \frac{1}{1+e^{-\sum_{j=0}^{n} w_j x_i^j}} < 1/2, \quad i = N_1+1,\dots,N \tag{16}$$

Bu yerda $w = (w_0, w_1, w_2, \dots, w_n)$.

Hosil bo‘lgan (14)-(16) optimallashtirish masalasini stoxastik tasodifiy qidiruv usullarining asoschisi L.A.Rastrigin tomonidan taklif etilgan moslashuvchan (adaptiv) tasodifiy qidiruv usulidan foydalanib yechish maqsadga muvofiq.

Moslashuvchan tasodifiy qidiruv usuli. Umumiy holda quyidagi optimallashtirish masalasi qo‘yilgan bo‘lsin

$$q(\omega) \rightarrow max => \omega^*, \tag{17}$$
$$\omega \in D$$

bu yerda *q(ω)* – umumiy holda nochiziqli ko‘p o‘zgaruvchili funksiya,

$\omega = (\omega_1, \omega_2, \dots, \omega_n)$,

$\omega^* = (\omega_1^*, \omega_2^*, \dots, \omega_n^*)$ – (17) masalaning yechimi.

D – optimallashtirish masalasining aniqlanish sohasi, odatda tenglik va tengsizliklar ko‘rinishida berilishi mumkin.

Moslashuvchan tasodifiy qidiruv usulida quyidagi rekkurent formuladan foydalaniladi:

$$\omega^{k+1} = \omega^k + \Delta\omega^{k+1}, \quad (18)$$

$$\Delta\omega^{k+1} = \begin{cases} a^{k+1}\Delta\omega^k, \ agarda \ q(\omega^k) > q(\omega^{k-1}) \\ a^{k+1} \cdot \xi^{k+1}, agarda \ q(\omega^k) \le q(\omega^{k-1}) \end{cases} \quad (19)$$

a^{k+1} – (k+1) qadamning uzunligini xarakterlovchi parametr.

Dissertatsiyaning «**Spam xabarlarni aniqlash jarayoni samaradorligini baholash va amaliyotga tadbiq etish natijalari**» nomli to‘rtinchi bobida elektron pochta xabarlaridan spam xabarlarni aniqlash usullarining samaradorligini baholash hamda spam xabarlarni aniqlashni dasturiy vositasining funksional strukturasi va ishlash prinsipi hamda joriy etishdan olingan tajriba-hisoblash natijalari keltirilgan.

Birinchi paragrafda Spam xabarlarni aniqlashda modifikatsiyalangan tayanch vektor usulining samaradorligi uchta asosiy ko‘rsatkich bo‘yicha baholangan (1-jadval). Bunda elektron pochta spami (EPS), ijtimoiy tarmoqlardagi spam (ITS), forumlardagi spam (FS), saytlardagi sharhlar orqali tarqatiladigan spam (SSHTS), kataloglar va byulletenlar ko‘rinishidagi spam (KBS), sms spamlar (SS).

1-jadval

Spam xabarlarni aniqlashning modifikatsiyalangan tayanch vektor usulini testlash natijalari

Spam xabar turlari	Jami spam xabar soni	Aniqlangan spam xabarlar %	Aniqlanmagan spam xabarlar %	Tizimning yolg‘on ishga tushishlari %
EPS	1364	97,3	2,1	0,6
ITS	1223	95,4	3,9	0,7
FS	1393	96,3	3,2	0,5
SSHTS	1676	95,1	4,1	0,8
KBS	1128	93,3	6,1	0,6
SS	1410	92,1	7,1	0,8

Spam xabarlarni aniqlashning modifikatsiyalangan tayanch vektor usuliga asoslangan tizimlarni boshqa spam xabarlarni aniqlash tizimlarining samaradorligi bilan solishtirganda olingan natijalar 2 – jadvalda keltirilgan. Bunda spam xabarlarni aniqlashda yuqorida ko‘rsatilgan 3 ta asosiy ko‘rsatgichlar bo‘yicha sinalgan. Shuningdek elektron pochta xabarlarini filtrlashning unumdorligi turlicha bo‘lgan protsessor va grafik protsessorli kompyuterlarda o‘tkazilgan testlash natijalari keltirilgan(3-jadval).

Ikkinchi paragrafda spam xabarlarni aniqlash dasturiy vositasining funksional strukturasi keltirilgan(6-rasm).

Axborot xavfsizligiga mas’ul xodimlariga ma’lumki, elektron pochta xabarlarini filtrlab ulardan spam xabarlarni aniqlab olish imkonini beradigan dasturiy vositalardan bugungi kunda deyarli barcha tashkilotlarda keng foydalanib

kelinmoqda. Ishlab chiqilgan algoritmlar asosida yaratilgan dasturiy vosita faqat spam xabarlarni aniqlash bilan cheklanib qolmasdan, bunga qo‘shimcha ravishda yana bir qancha vazifalarni bajarish imkonini beradi:

- spam xabarlarni aniqlash;
- spam xabarlarni karantinda saqlab turish;
- spam xabarlarni o‘chirib tashlash;
- xabar jo‘natish;
- xabarlarni qabul qilish va boshqalar.

2-jadval

Spam xabarlarni aniqlashning modifikatsiyalangan tayanch vektorlar usulini samaradorligini baholash natijalari

Spam xabarlarni aniqlash tizimlari	Jami spam xabar soni	Aniqlangan spam xabarlar %	Aniqlanmagan spam xabarlar %	Tizimning yolg‘on ishga tushishlari %
SXAD	1665	96,2	3,4	0,4
AntiSpam Sniper Pro.	1665	94,8	3,1	2,1
Malwarebytes Anti-Exploit Free	1665	93,4	4,4	2,2
Spybot Search & Destroy	1665	92,7	4,8	2,5
Spamhaus	1665	97,1	2,1	0,8

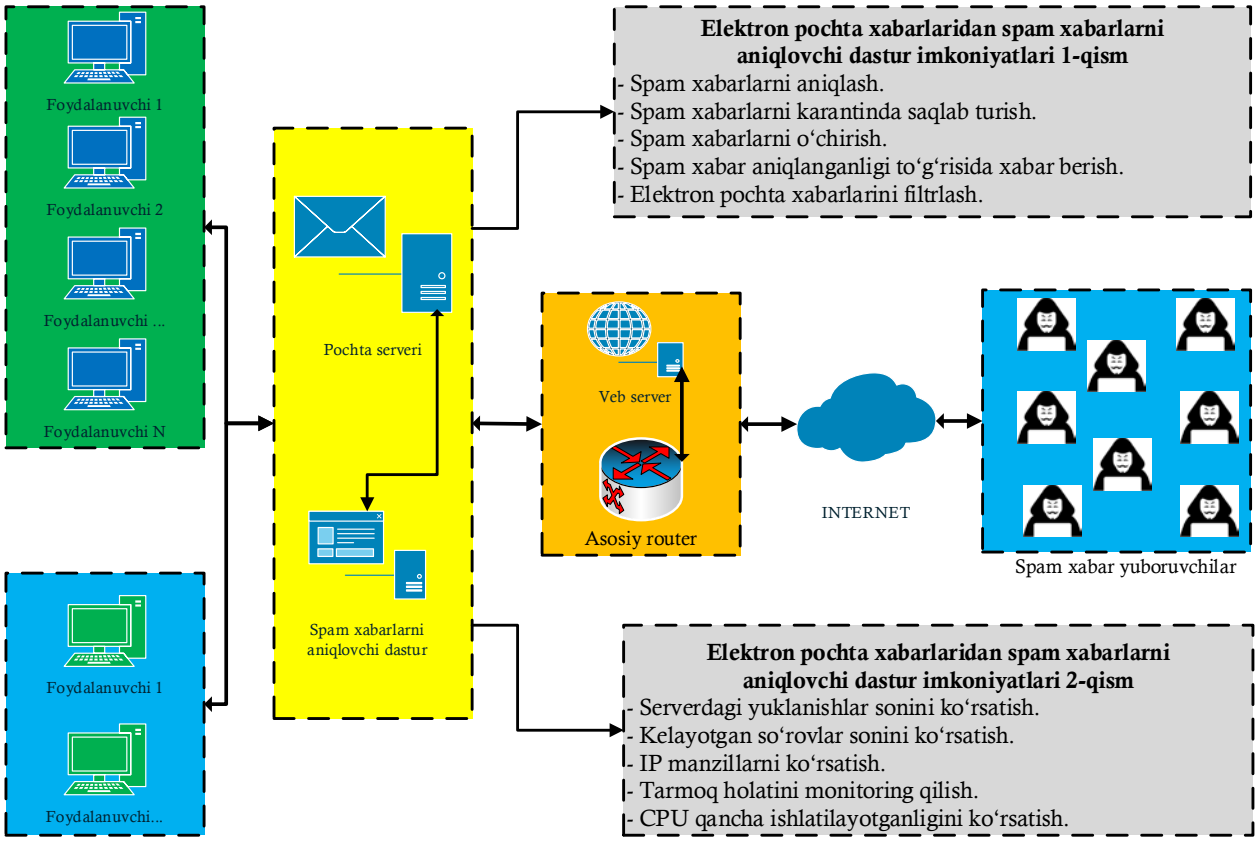
3 – jadval

Elektron pochta xabarlarini filtrlashning unumdorligi turlicha bo‘lgan protsessor va grafik protsessorli kompyuterlarda o‘tkazilgan testlash natijalari

Qurilmaga qo‘yilgan talablar / Elektron pochta xabarlarni filtrlash(EPXF)	Intel core i5 4,4GHz, 16Gb tezkor xotira	Intel Core i7 7900K, 64Gb tezkor xotira	Intel Core i9 7900K, 64Gb tezkor xotira, GTX 1080 TI grafik protsessor
SXAD (EPXF)	0,7 s	39 ms	11 ms
MailWasher Free (EPXF)	1,5 s	65 ms	26 ms
SpamBytes (EPXF)	0,9 s	43 ms	13 ms
Spamihilator (EPXF)	0,8 s	41 ms	12 ms
Spamfence (EPXF)	1,2 s	49 ms	15 ms

Uchinchi paragrafda tashkilot elektron pochta tizimini spam xabarlardan himoyalashning dasturiy vositasini amaliyotga tatbiq etish natijalari keltirilgan.

SXAD dasturining asosiy maqsadi viruslar, xakerlik hujumlari va spamlardan keng qamrovli himoya asosida tashkilotning elektron pochtasini himoya qilish bilan bir qatorda murakkab yondashuv yordamida kiruvchi xatlarni tahlil qilishdan iborat.



6-rasm Spam xabarlarni aniqlash dasturiy vositasining funksional strukturasi

Tashkilot va korxonalarni elektron pochta manzillarini filtrlash va ulardan spam xabarlarni aniqlab olishning dasturiy vositasi «Kiberxavfsizlik» markazi davlat unitar korxonasida, «UZINFOCOM Davlat axborot tizimlarini yaratish va qo‘llab-quvvatlash bo‘yicha yagona integrator» ma’suliyati cheklangan jamiyatida va «ASR KABEL» ma’suliyati cheklangan jamiyatida joriy etilgan va natijalar quyida keltirilgan.

«Kiberxavfsizlik» markazi davlat unitar korxonasida elektron pochta xabarlarini filtrlash, spam xabarlarni aniqlash va ularni bloklash uchun ishlab chiqilgan dasturiy vosita 4673 ta elektron pochta xabarlarini filtrlab, ulardagi 1364 ta spam xabardan 1327 tasini 97.3% aniqlik qayd etgan holda aniqlash imkonini berdi.

«UZINFOCOM Davlat axborot tizimlarini yaratish va qo‘llab-quvvatlash bo‘yicha yagona integrator» ma’suliyati cheklangan jamiyatida elektron pochta xabarlarini filtrlash va ulardagi spam xabarlarni aniqlash maqsadida tajribaviy sinovdan o‘tish jarayonida elektron pochta xabarlaridagi spam xabarlarni aniqlash dasturiy vositasi 3624 ta elektron pochta xabarlarini samarali filtrlab, ulardagi

1465 ta spam xabardan 1431 tasini 97.6% aniqlik qayd etgan holda aniqlash imkonini berdi.

«ASR KABEL» ma’suliyati cheklangan jamiyatida elektron pochta xabarlarini filtrlash va ulardagi spam xabarlarni aniqlash hamda bloklash maqsadida sinovdan o‘tish jarayonida elektron pochta xabarlaridagi spam xabarlarni aniqlash dasturiy vositasi 2526 ta elektron pochta xabarlarini samarali filtrlab, ulardagi 1280 ta spam xabardan 1248 tasini 97.5% aniqlik qayd etgan holda aniqlab ularni muvofaqqiyatli bloklash imkonini berdi.

Ishlab chiqilgan dasturiy vositani korxona va tashkilotlarda qo‘llash natijasida olingan natijalar asosida shuni aytish mumkinki, bitta foydalanuvchining etiborsizligi natijasida tashkilotning axborot tizimiga zarar yetishi va uning natijasida pochta serverini ish faoliyatiga salbiy tasir yuklanishi yoki tashkilotning ma’lumotlar bazasidan kerakli ma’lumotlarni chiqib ketishi, o‘chirilishi yoki o‘zgartirilishi kabi oqibatlarga sabab bo‘lishi mumkin. Bir qaraganda oddiy ko‘ringan spam xabarlarni vaqtida aniqlab bartaraf etilmasa, uning oqibatlari asosida tashkilotning boshqa tizimlariga kirishga yo‘l ochiladi, chunki biror tashkilot yo‘qki elektron pochta manziliga ega bo‘lmagan. Elektron pochta ana shunday barcha tashkilotlarda mavjud bo‘lgan va butun dunyo bo‘yicha ma’lumotlar almashish imkonini beradigan yagona texnologiyalaridan biridir. Shuning uchun ham ushbu texnologiyalardan samarali foydalanish uchun uning xavfsizligiga ham alohida e’tibor qaratish maqsadga muvofiq sanaladi.

XULOSA

«Elektron pochta xabarlarini filtrlash va spam xabarlarni aniqlash algoritmlarini ishlab chiqish» mavzusidagi dissertatsiya ishi bo‘yicha olib borilgan tadqiqotlar natijasida quyidagi xulosalar taqdim etildi:

1. Elektron pochtani filtrlash usullari turli xususiyatlar, parametrlar va muhitlar bo‘yicha qiyosiy tahlil qilindi va tayanch vektorlar usulining samaradorligi yuqori ekanligi aniqlandi. Spam xabarlarining barcha ko‘rinishlari uchun umumiy bo‘lgan kerakli mezonlarni tanlab olish natijasida filtrlash tizimining ortiqcha yuklanishlari sonining kamayishiga erishildi.
2. Tashkilotdagi mavjud bilimlar bazasini taqsimlash orqali yangi turdagi spam xabarlarni aniqlash imkonini beradigan yangi bilimlar asosida real vaqt rejimida bilimlar bazasini to‘ldirish algoritmi ishlab chiqildi. Natijada bilimlar bazasida bilimlar soni keskin ortib ketgan holda yuborilgan so‘rovlarga javob qaytarish vaqtining 1 ms kamayishiga erishildi.
3. Spam xabarlarni aniqlash tizimining, ko‘p darajali bilimlar bazasi iyerarxiyasiga mos holda, uch bosqichdan iborat elektron pochtadagi hujjatlarni filtrlash algoritmi ishlab chiqildi. Natijada elektron xabarni spam xabarga tegishli ekanligi holatining yakuniy ehtimolligini baholash imkoniga ega bo‘lindi.
4. Spam xabarlarni aniqlashning modifikatsiyalangan tayanch vektorlar usuli va algoritmi ishlab chiqildi. Natijada F klassifikator asosida hujjatlarni aniqlangan tavsifi asosida ikki toifaga ajratish imkoniyatiga erishildi.

5. Modifikatsiyalangan tayanch vektorlar usuli asosida ishlab chiqilgan spam xabarlarni aniqlashning dasturiy vositasi 97.3% aniqlikni ko‘rsatdi.

6. Ishlab chiqilgan spam xabarlarni aniqlash dasturi yordamida elektron pochta xabarlaridan spam xabarlarni aniqlash uchun ketgan vaqt 11 ms ni tashkil etdi. Bu boshqa dasturiy vositalarga qaraganda 1 ms tezroq ekanligi aniqlandi.

Taklif etilgan usullar asosida ishlab chiqilgan elektron pochta xabarlarini filtrlash va ulardan spam xabarlarni aniqlab olishning dasturiy vositasi tashkilotdagi axborotni himoyalash uchun foydalaniladigan himoya tizimining yolg‘on ishga tushish ehtimolini 1.4 barobarga kamaytirishga hamda ishlash tezligini 1.1 barobarga ortishiga va elektron pochta xabarlarini 24/7 rejimida himoyasini ta’minlash imkonini beradi. Spam xabarlarni aniqlashda bundanda yaxshiroq samaradorlik ko‘rsatkichiga erishish uchun spam xabarlar bazasini doimiy yangilab borish bo‘yicha tavsiyalar keltirildi.

**НАУЧНЫЙ СОВЕТ DSc. 13/30.12.2019.T.07.02
ПО ПРИСУЖДЕНИЮ УЧЕНЫХ СТЕПЕНЕЙ ПРИ ТАШКЕНТСКОМ УНИВЕРСИТЕТЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

ТАШКЕНТСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

ҲАЙДАРОВ ЭЛШОД ДИЛШОД УГЛИ

РАЗРАБОТКА АЛГОРИТМОВ ФИЛЬТРАЦИИ СООБЩЕНИЙ ЭЛЕКТРОННОЙ ПОЧТЫ И ОБНАРУЖЕНИЯ СПАМ СООБЩЕНИЙ

05.01.05 – Методы и системы защиты информации. Информационная безопасность.

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

Ташкент-2023

Тема диссертации доктора философии (PhD) по техническим наукам зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за № B2022.4.PhD/T3425.

Диссертация выполнена в Ташкентском университете информационных технологий.

Автореферат 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:	**Хамдамов Рустам Хамдамович** доктор технических наук, профессор
Официальные оппоненты:	**Керимов Камил Фикратович** доктор технических наук, доцент
	Насруллаев Нурбек Бахтиёрович доктор физико-математических наук, доцент
Ведущая организация:	**«UNICON.UZ» - центр научно-технических и маркетинговых исследований**

Защита диссертации состоится «__» _____2023 года в __ часов на заседании Научного совета DSc. 13/30.12.2019.T.07.02 при Ташкентском университете информационных технологий. (Адрес: 100084, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-43; e-mail: tuit@tuit.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского университета информационных технологий (регистрационный номер №___). (Адрес: 100084, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-70).

Автореферат диссертации разослан «___» _____ 2023 года.
(протокол рассылки №__от «___» _____ 2023 года.)

Б.Ш. Махкамов
Председатель научного совета по присуждению ученых степеней, д.э.н., профессор

Э.Ш. Назирова
Ученый секретарь научного совета по присуждению ученых степеней, д.т.н., профессор

С.К. Ганиев
Председатель научного семинара при научном совете по присуждению ученых степеней, д.т.н., профессор

ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))

Актуальность и востребованность темы диссертации. Электронная почта - единственная система обмена электронными сообщениями, признанная всеми странами мира. Именно поэтому количество атак на электронную почту с каждым годом увеличивается. Основными организаторами этих атак являются спам-сообщения. По данным Лаборатории Касперского, в первом квартале 2021 года количество неизвестных звонков и спам-сообщений в государственные организации составило 70% всех входящих звонков и сообщений. В этом направлении в зарубежных странах, в том числе в Российской Федерации, США, Южной Корее, Японии, Малайзии, Китае и других странах ведутся исследования по разработке методов и алгоритмов фильтрации сообщений электронной почты и выявлению из них спам-сообщений, а также совершенствованию систем защиты почтовых серверов компании.

В настоящее время весьма актуальными считаются исследования и разработки, направленные на совершенствование методов и алгоритмов фильтрации сообщений электронной почты и выявление спам-сообщений, а также выявление спам-сообщений с использованием интеллектуальных методов. В связи с этим одной из важных задач является разработка методов выявления спам-сообщений на основе списков и лингвистики. В то же время необходимо разработать механизм защиты с использованием интеллектуальных методов для защиты электронной почты, когда знание новых сообщений является неполным в процессе фильтрации электронной почты, или когда подпись спам-сообщений недоступна, наряду с неопределенностями в текущей базе спам-сообщений.

В нашей республике органами государственного и хозяйственного управления осуществляются комплексные мероприятия, направленные на фильтрацию сообщений электронной почты и выявление из них спам-сообщений.

В Стратегии развития Нового Узбекистана на 2022-2026 годы определено, что «…основные направления обеспечение кибербезопасности интернет-пространства доменной зоны «UZ» и комплексные задачи защиты систем электронного правительства, энергетики, цифровой экономики и другие направления, связанные с важной информационной инфраструктурой, отмечены значимыми». При реализации этих задач одной из важных является использование современных систем фильтрации сообщений электронной почты и выявление из них спам-сообщений, а также использование современных систем защиты электронной почты.

Реализация поставленных задач в Указах Президента Республики Узбекистана «О Стратегии развития Нового Узбекистана на 2022-2026 годы» №УП-60 от 28 января 2022 года, «О мерах по дальнейшему совершенствованию сферы информационных технологий и коммуникаций» №УП-5349 от 19 февраля 2018 года, Постановлениях Президента Республики Узбекистана «О мерах по совершенствованию системы контроля за

внедрением информационных технологий и коммуникаций, организации их защиты» № ПП-4024 от 21 ноября 2018 года, «О дополнительных мерах по совершенствованию системы контроля за внедрением информационных технологий и коммуникаций, организации их защиты» № ПП-4452 от 14 сентября 2019 года, и в других нормативно-правовых документах, касающихся данной деятельности, служат задачами для данного диссертационного исследования в определенной мере.

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетными направлениями развития науки и технологии IV. «Информатизация и развитие инфокоммуникационных технологий».

Степень изученности проблемы. E.Dada, J.Bassi, T.S. Guzella, V. Zorkadis, К. Третяков, Vaughan-Nikols, A.Kumari и другие зарубежные ученые ведут научные исследования по применению методов и алгоритмов, разработанных для обнаружения спам-сообщений с использованием фильтрации сообщений, метода опорных векторов и аналогичных интеллектуальных методов для обеспечения безопасности электронной почты в сфере информационной безопасности. S. Lin, Y. Zhou, K. Yang, L.Chen, R.Krestel проводили научные исследования с применением опорных векторов в защите электронной почты. Кроме того, организациями «Лаборатория Касперского», ESET, AGAVA SpamProtexx и Trend Micro ведутся инженерно-исследовательские работы по разработке программно-аппаратных средств фильтрации сообщений электронной почты и обнаружения спама с использованием интеллектуальных методов.

В Узбекистане академик Т.Ф.Бекмуратов проводил научные исследования по разработке методов и алгоритмов фильтрации сообщений электронной почты и выявления из них спам-сообщений, а в настоящее время исследования проводятся научными коллективами под руководством С.К.Ганиева, Р.Х.Хамдамова, М.М.Каримова.

В то же время методы и алгоритмы обнаружения и устранения спам-сообщений на основе опорных векторов, нейронных сетей, машинного обучения и глубокого обучения недостаточно исследованы.

Связь диссертационного исследования с планами научно-исследовательских работ высшего образовательного учреждения, где выполнена диссертация. Диссертационное исследование выполнено в рамках проекта 598661-EPP-1-2018-1-RO-EPPKA2-CBHE-JP «Развитие услуг для лиц с ограниченными возможностями» Ташкентского университета информационных технологий имени Мухаммада ал-Хоразмий.

Цель исследования заключается в разработке методов и алгоритмов обнаружения спам-сообщений, позволяющих повысить эффективность системы защиты электронной почты.

Задачи исследования:

сравнительный анализ методов фильтрации электронной почты и выбор критериев, необходимых для фильтрации;

разработка алгоритма пополнения базы знаний системы фильтрации;

разработка алгоритма фильтрации документов в электронной почте;

разработка модифицированного метода опорных векторов и алгоритма обнаружения спам-сообщений;

разработка алгоритма выявления спам-документов на основе логистической регрессии.

Объектом исследования является спам-сообщения в электронной почте.

Предмет исследования является методы и алгоритмы фильтрации электронной почты и обнаружения спама.

Методы исследования. В процессе исследования использовались методы фильтрации электронной почты и обнаружения спама, нейронные сети, машинное обучение, теория вероятностей, дискретная математика и объектно-ориентированное программирование.

Научная новизна исследования:

сформированы единые критерии для сообщений электронной почты с обширными возможностями и многими свойствами, в результате извлечения признаков, которые являются общими для спам сообщений всех видов;

с учетом частоты использования знаний в существующей базе знаний в организации была сформирована наиболее используемая группа знаний, в результате был разработан алгоритм пополнения базы знаний, позволяющий выявлять новые виды спам-сообщений в режиме реального времени;

с учетом типов документов в электронной почте была сформирована многоуровневая база знаний по расширению файла среди этих документов, в результате был разработан алгоритм фильтрации документов в электронной почте, состоящий из трех этапов, с использованием многоуровневой базы знаний системы обнаружения спам-сообщений;

разработаны модифицированный метод опорных векторов и алгоритм обнаружения спам-сообщений на основе метода обучения с обучающей выборкой;

обобщая метод логистической регрессии с гибким методом случайного поиска был разработан алгоритм обнаружения спам-документов путем нахождения наибольшей вероятности спам-сообщения среди сообщений.

Практические результаты исследования:

разработан программный инструмент для обнаружения спам-сообщений на основе методов фильтрации электронной почты;

усовершенствован метод опорных векторов обнаружения спама, основанный на обучении на обучающих выборках.

Достоверность результатов исследования. Достоверность результатов исследования объясняется результатами реального и экспериментального анализа, полученными в различных условиях методами фильтрации электронной почты организации, пополнения базы знаний, выявления спам-сообщений.

Научная и практическая значимость результатов исследования. Научная значимость результатов исследования объясняется разработанным алгоритмом пополнения базы знаний, позволяющим идентифицировать новый вид спам-сообщений, трехступенчатым алгоритмом фильтрации документов в электронной почте, модифицированным методом и алгоритмом опорных векторов для обнаружения спам-сообщений, а также алгоритмом обнаружения спам-документов.

Практическая значимость результатов исследования объясняется тем, что программный инструмент, разработанный на основе предложенных методов и алгоритмов, позволяет повысить эффективность процесса фильтрации сообщений электронной почты и выявления спам-сообщений в организации.

Внедрение результатов исследования. На основе научных результатов практического применения алгоритмов фильтрации электронной почты и обнаружения спама и программных средств:

программное средство, разработанное на основе алгоритмов фильтрации сообщений электронной почты и выявления спам-сообщений, внедрено в практическую деятельность ГУП «Центр кибербезопасности» (Справка Министерства развития информационных технологий и коммуникаций Республики Узбекистан от 12.10.2022 г. № 33-8/6761). В ходе научного исследования было отфильтровано 4673 сообщения электронной почты из корпоративной сети и обнаружено 1327 спам-сообщений из 1364. Это позволило идентифицировать спам-сообщения с точностью 97,3%;

программное средство, разработанное на основе алгоритмов фильтрации сообщений электронной почты и выявления спам-сообщений, внедрено в практическую деятельность общества с ограниченной ответственностью «Единый интегратор по созданию и сопровождению государственных информационных систем UZINFOCOM» (Справка Министерства развития информационных технологий и коммуникаций Республики Узбекистан от 12.10.2022 г. № 33-8/6761). В результате научных исследований удалось эффективно отфильтровать 3624 сообщения электронной почты в локальной сети организации и идентифицировать 1431 из 1465 спам-сообщений с точностью 97,6%;

в практическую деятельность общества с ограниченной ответственностью «ASR KABEL» внедрен программный инструмент, разработанный на основе алгоритмов фильтрации электронной почты и обнаружения спама (Справка Министерства развития информационных технологий и коммуникаций Республики Узбекистан от 12.10.2022 г. № 33-8/6761). В результате проведенного исследования было эффективно отфильтровано 2526 сообщений электронной почты в корпоративной сети и успешно заблокировано 1248 из 1280 спам-сообщений с точностью 97,5%.

Апробация результатов исследования. Результаты исследования обсуждались на 2 международных и 7 республиканских научно-практических конференциях.

Публикация результатов исследования. Всего по теме диссертации опубликовано 24 научные работы, в том числе 6 статей в научных изданиях

рекомендованных к публикации основных научных результатов диссертаций ВАК РУз, из них 3 опубликованы в иностранных и 3 республиканских журналах, а также получены 4 свидетельства о регистрации программных продуктов для ЭВМ.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав, заключения, списка литературы и приложений. Объем диссертации составляет 112 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во **введении** обоснована актуальность и необходимость темы диссертации, приведено соответствие исследования приоритетным направлениям развития науки и техники Республики Узбекистан, определены цели и задачи, объект и предмет исследования, обоснованы достоверность полученных результатов, их теоретическая и практическая значимость, состояние внедрения результатов исследования в практику, приведены опубликованные работы и информация о структуре диссертации.

Первая глава диссертации под названием **«Проблемы фильтрации электронной почты и обнаружения спама»** содержит сведения о методах фильтрации электронных писем и их сравнительный анализ, сведения о проблемах выбора критериев, необходимых при фильтрации электронных писем и идентификации спам-сообщений из электронных писем, а также рекомендациях по их решению.

В первом параграфе приведены сравнительный анализ типов сервисов электронной почты, просмотров спам-сообщений, методов фильтрации электронной почты по различным характеристикам, параметрам и средам и, принимая во внимание широкую доступность и множественность, была установлена, эффективность метода базового вектора с точки зрения удобства, универсальности, внедрения, агрегирования, экономии времени, отсутствия дублирования и точности на высоком уровне.

Во втором параграфе приведены общие для всех типов проявления спам-сообщений (спам-сообщение существует в той или иной форме, несуществующей для другого типа спам-сообщений), принимая во внимание их происхождение и распространение, а также необходимые факторы при их идентификации, определены необходимые критерии для обеспечения количества перегрузок фильтрации, когда система не увеличивает и не достигает высокой эффективности. Также в параграфе представлены спам-факторы и их взаимосвязь с методами борьбы со спамом.

В третьем параграфе определены этапы внедрения системы фильтрации в почтовом сервисе и существующие недостатки в действующих системах фильтрации сообщений электронной почты и выявления из них спам-сообщений, определены проблемы в процессе фильтрации, разработаны механизмы защиты для проверки записей DNS и диагностики доменов, разработаны законы по борьбе со спамом, и даны рекомендации по устранению спам-сообщений.

Во второй главе диссертации под названием **«Алгоритмы пополнения базы знаний системы электронной почты и фильтрации документов»** разработаны иерархия фильтрации спам-сообщений в сообщениях электронной почты и алгоритмы пополнения базы знаний системы фильтрации сообщений электронной почты и фильтрации документов в сообщениях электронной почты в соответствии с этой иерархией.

В *первом параграфе* настоящей главы предложена система, осуществляющая прием, обработку, передачу и хранение всей информации, поступающей по информационному каналу организации, а также осуществляющая фильтрацию всех сообщений электронной почты, поступающих по каналу связи данной системы, идентификацию спам-сообщений. Помимо этого, предложена иерархия фильтрации спама в сообщениях электронной почты, позволяющая разделять входящие и исходящие-сообщения на группы.

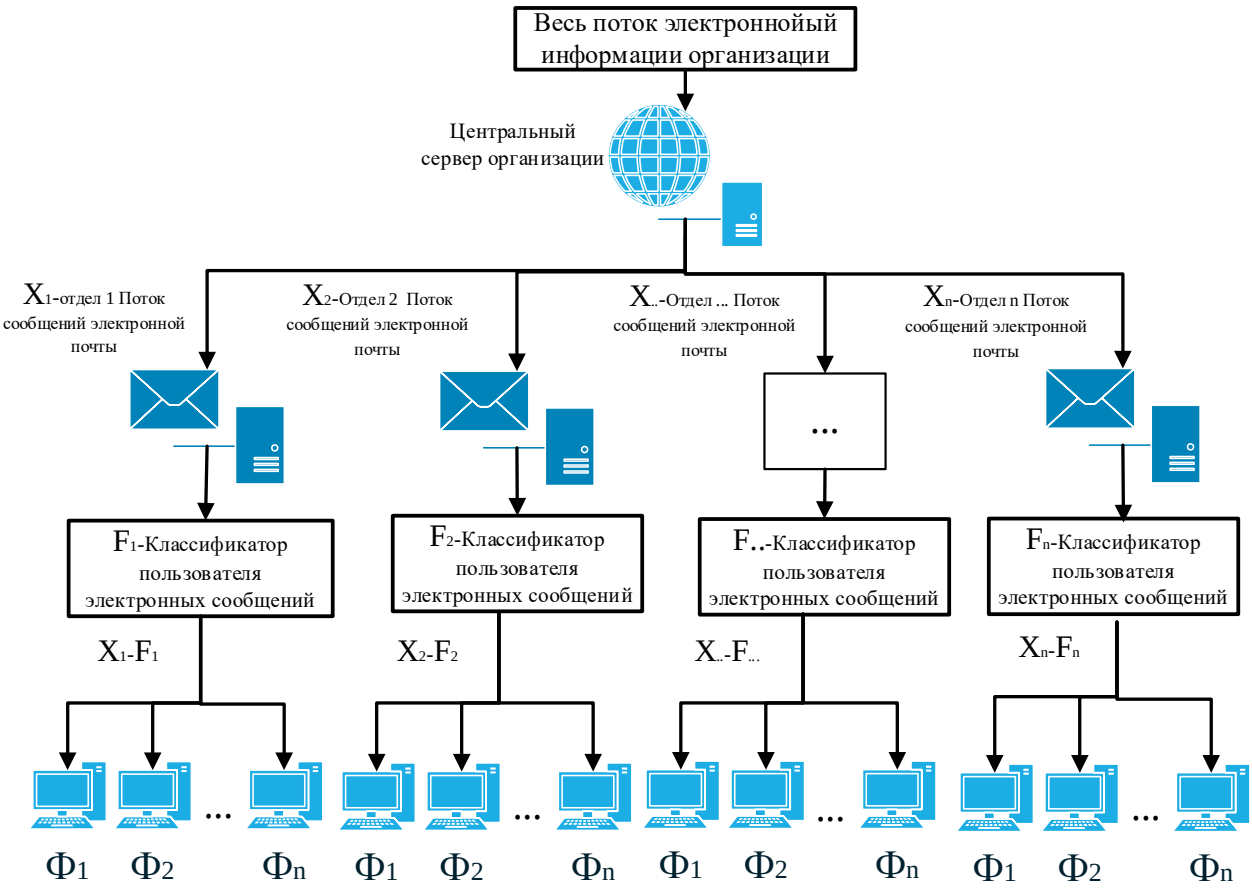


Рисунок 1. Поток информации иерархически расположенной в разных узлах сети системы спам-фильтров

Во втором параграфе разработан алгоритм добавления новых знаний в базу знаний в режиме реального времени, т.е. пополнения базы знаний, который позволяет выявлять новые виды спам-сообщений путем разделения существующей базы знаний в организации. Поскольку количество знаний в базе знаний резко увеличивается, время ответа на отправленные запросы увеличивается. Для предотвращения этой проблемы была сформирована

иерархия базы знаний многоуровневой системы обнаружения спам-сообщений. В формулах 1, 2, 3 приведено добавление новых знаний в базу знаний.

$$B_1 = b_{11} \cup b_{12} \cup \ldots \cup b_{1n} = \bigcup_{k=1}^{n} b_{1k} \quad (1)$$

$$B_B = B_1 \cap B_2 \cap \ldots \cap B_p = \bigcap_{k=1}^{p} B_k \quad (2)$$

$$B_B = (b_{11} \cup b_{12} \cup \ldots \cup b_{1n}) \cap (b_{21} \cup b_{22} \cup \ldots \cup b_{2m}) \cap \ldots \cap (b_{p1} \cup b_{p2} \cup \ldots \cup b_{po}) = \bigcap_{k=1}^{p} \left(\bigcup_{j=1, j \neq k}^{r} b_{k,j} \right) \quad (3)$$

На рисунке 2 представлена иерархия формирования базы знаний многоуровневой системы обнаружения спама.

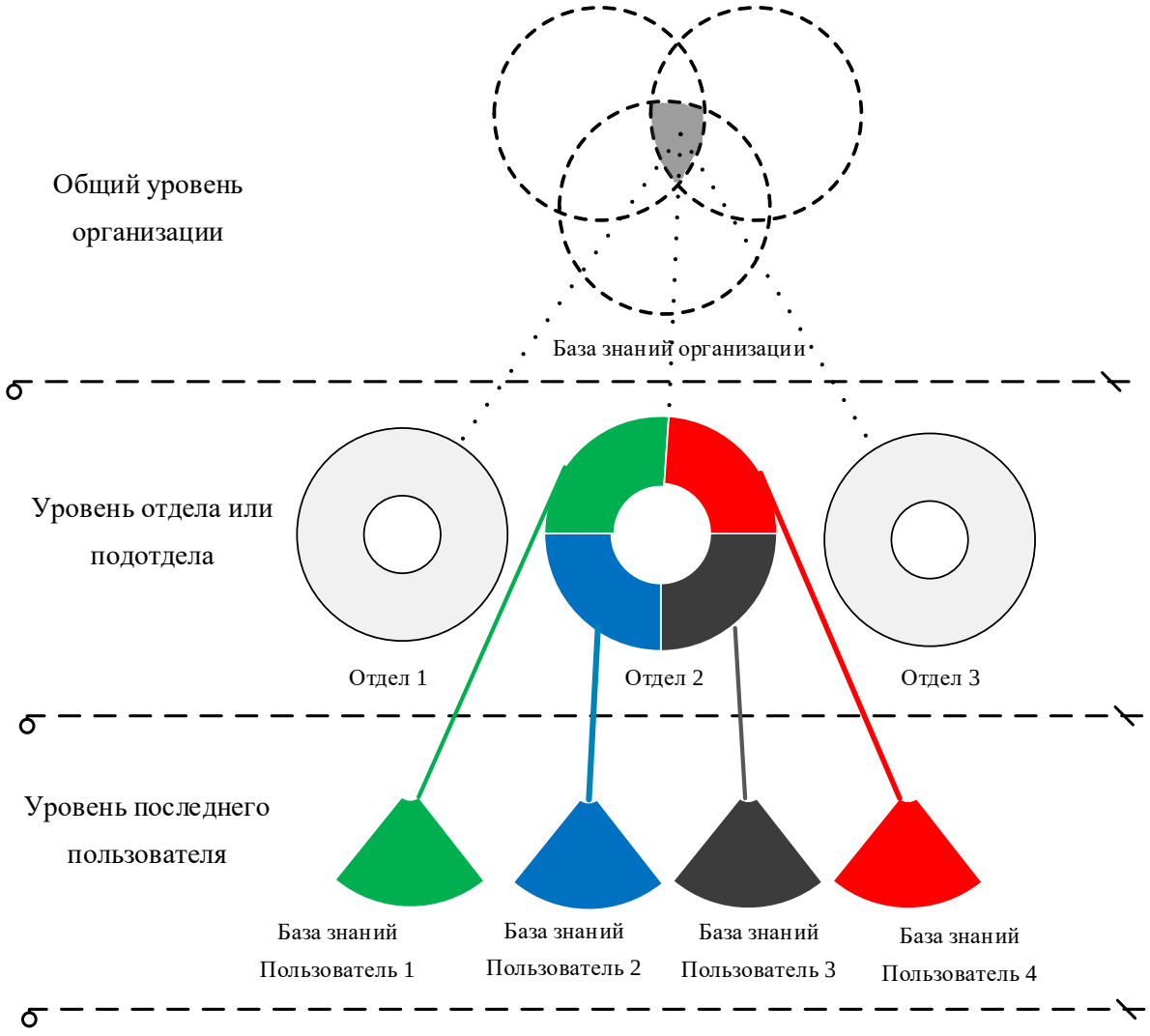


Рисунок 2. Иерархия формирования базы знаний многоуровневой системы обнаружения спама

В третьем параграфе разработан алгоритм фильтрации документов в электронной почте в соответствии с иерархией базы знаний многоуровневой системы обнаружения спама (рис. 3). В результате система фильтрации спама

позволяет оценить статус письма как спам-сообщения, разделив итоговую вероятность на группы.

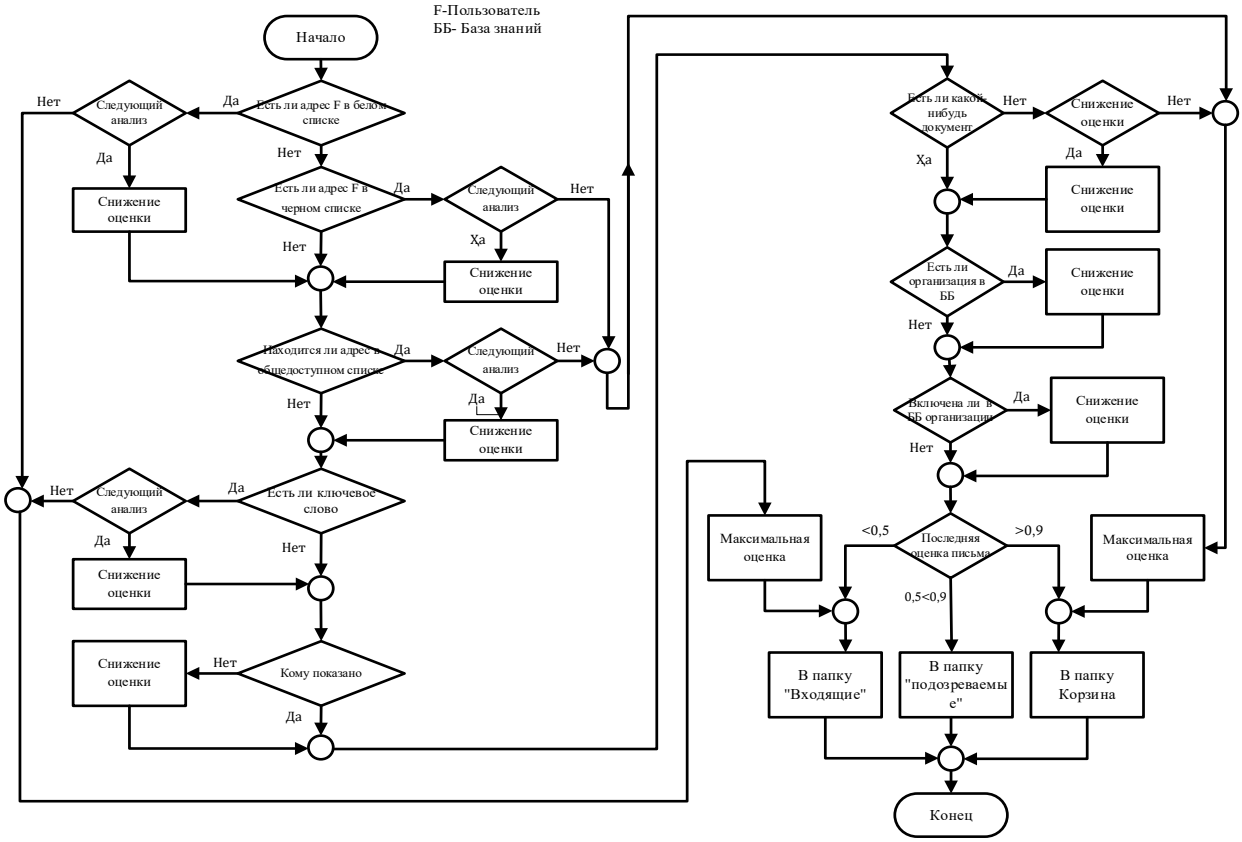


Рисунок 3. Блок-схема алгоритма фильтрации документов

В третьей главе диссертационной работы под названием **«Интеллектуальные методы и алгоритмы обнаружения спам-сообщений»** представлены схема обнаружения спам-сообщений на основе машинного обучения, модифицированного метода опорных векторов и алгоритма обнаружения спам-сообщений, а также алгоритма обнаружения спам-документов на основе логистической регрессии.

В первом параграфе предложено классифицировать сообщения электронной почты путем добавления статистических и нестатистических признаков к входным параметрам нейронной сети, используемой для обнаружения спам-сообщений на основе существующих методов. В соответствии с этой классификацией были сформированы интеллектуальная система фильтрации сообщений электронной почты и схема обнаружения спам-сообщений на основе машинного обучения (рис. 4).

Непосредственное построение эффективной нейросетевой модели (разновидности нейронной сети) фильтрации электронной почты осуществляется с использованием технологии распознавания знаний в базе данных, которое включает следующие этапы:

1. Получение предварительных данных электронной почты, включая образцы спамовых и не спамовых писем;

2. Формирование обучающей выборки для предобработки исходных данных и обучения нейронной сети;

3. Разработка структуры нейронной сети: определение количества входов, выходов, слоев сети и нейронов в каждом слое;

4. Обучение нейронной сети созданию модели фильтрации спам-сообщений;

5. Тестирование и оценка нейросетевой модели фильтрации спама.

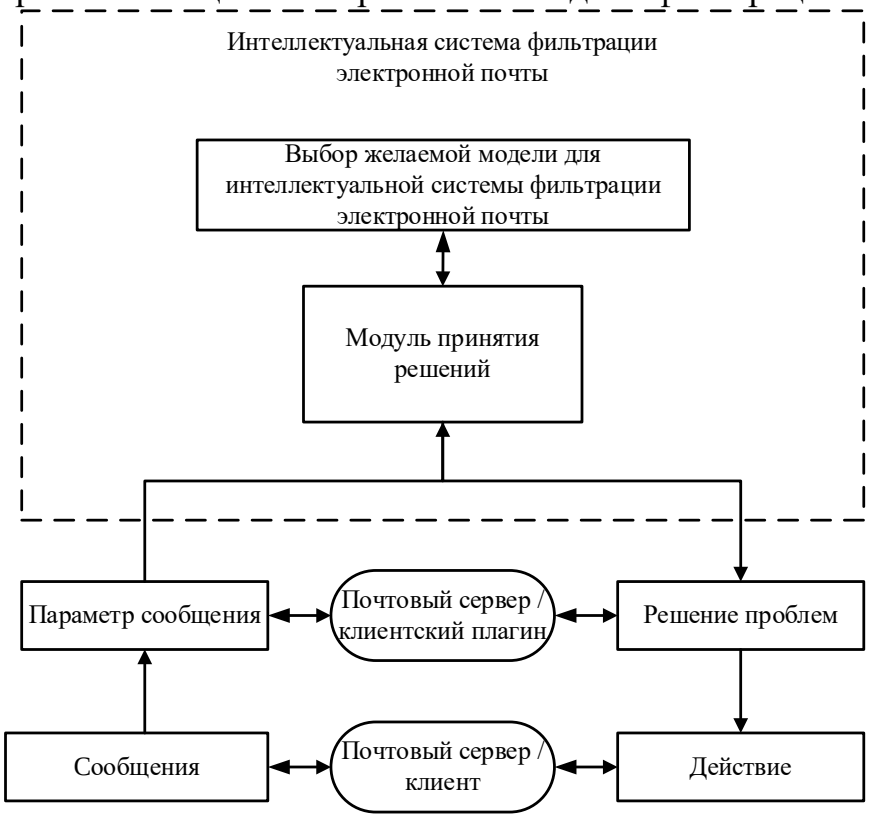


Рисунок 4. Интеллектуальная система фильтрации электронной почты

Во втором параграфе описан модифицированный метод опорных векторов выявления спам-сообщений с обучающей выборкой, то есть с учетом выборок, в какой категории (для каждой категории) находятся и индуктивным обучением на основе заданного описания документов, что позволяет с помощью разработанного метода опорных векторов и алгоритма разделить выборки на две категории(рис. 5).

Любая информация *x* электронного письма (spam или ham) представлена *n*-мерной системой символов (вектором):

$$x=(x^1,x^2, \dots , x^n) \quad (4)$$

Два класса сообщений электронной почты (spam или ham) представлены в виде набора объектов (сообщений), представленных следующей *n*-мерной системой обозначений (5):

$$x_{pi} = (x_{pi}^1, x_{pi}^2, \dots , x_{pi}^n), \quad p = 1, 2, \dots , m; \; i = 1, 2, \dots , k_p \quad (5)$$

здесь x_{pi}^j - числовое значение *j*-го символа, принадлежащего *i*-му объекту в *p*-м классе, *m*-количество заданных классов, k_p-количество объектов в *p*-м классе.

$$x_i = \begin{Bmatrix} \text{spam}, & \text{если} \quad \sum_{j=1}^{6} x_i^j \geq b_q \text{ или } (x_i^7 > 0.06 \text{ или } x_i^8 > 0.03\ ) \\ \text{ham}, & \text{в других случаях} \end{Bmatrix} \quad (6)$$

В качестве классификатора используется гиперплоскость, которая оптимально разделяет два класса в пространстве R^n методом опорных векторов.

$$\omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n + \omega_0 = 0 \quad (7)$$

Суть метода заключается в том, чтобы найти уравнение (7). Выражение преобразования функции F объекта x в метку класса Y можно представить в виде:

$$F(x) = sign(\omega^T x - b). \quad (8)$$

здесь

$$\omega = (\omega_1, \omega_2, \ldots, \omega_n), \quad b = -\omega_0 \quad (9)$$

После настройки весовых коэффициентов (неизвестных параметров) алгоритма обучения ω и b все объекты, попадающие в одну сторону построенной гиперплоскости, определяются как первый класс, а объекты, попавшие по другую сторону, определяются как второй класс.

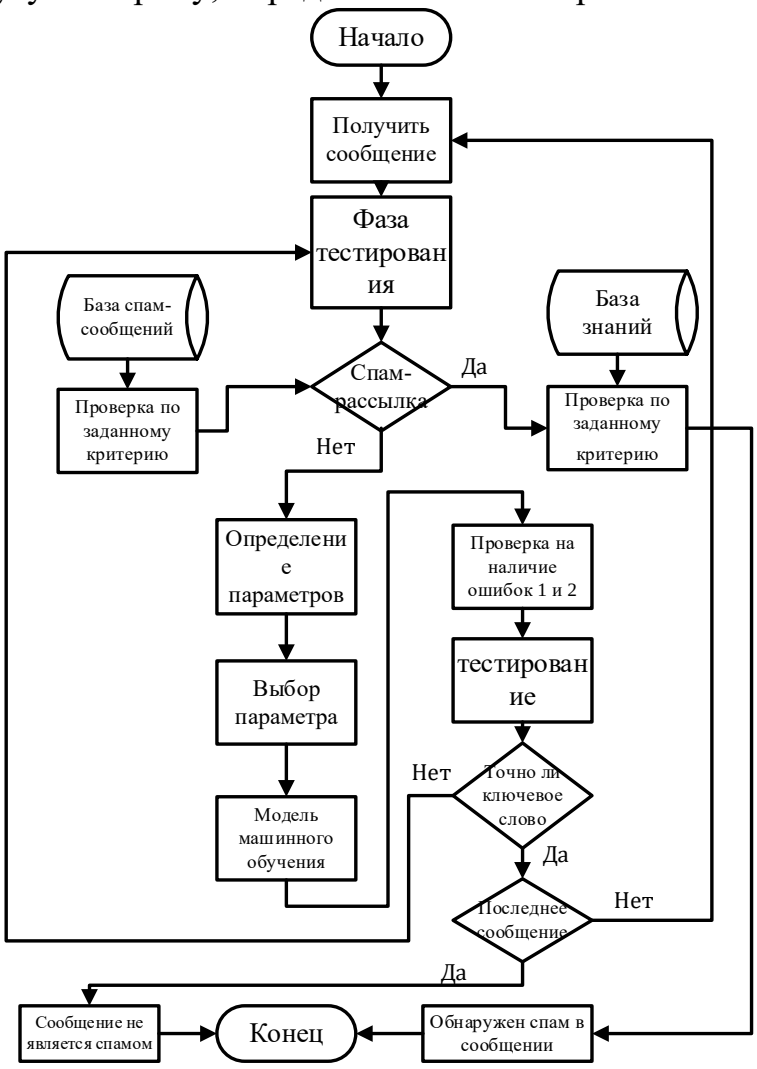


Рисунок 5. Блок-схема модифицированного алгоритма опорных векторов для обнаружения спам-сообщений

В третьем параграфе используя метод логистической регрессии с применением метода адаптивного случайного поиска Растригина Л.А., удалось повысить надежность решения задачи классификации и быстродействие системы обнаружения спам-сообщений за счет нахождения наиболее вероятного спама среди признаков сообщения.

Модель логистической регрессии (ЛР):

$$f_{w,b}(x) = \frac{1}{1+e^{(wx-b)}} \tag{10}$$

для оптимизации, то есть критерий максимизации записывается в следующем виде:

$$q(w,b) = \sum_{i=1}^{N}[y_i \ln f_{wb}(x_i) + (1-y_i)\ln(1-f_{wb}(x_i))] \to \underset{w,b}{max} => (w*, b*) \tag{11}$$

здесь

$$y_i = \begin{cases} 1, & i = 1,2,\dots,N_1 \\ 0, & i = N_1+1,\dots,N \end{cases} \tag{12}$$

Используя условия (10) и (12), критерий оптимизации (11) можно записать в следующем виде:

$$q(w) = \sum_{i=1}^{N} \ln\left(1 + e^{\sum_{j=0}^{n} w_j x_i^j}\right) - \sum_{i=N_1+1}^{N} \sum_{j=0}^{n} w_j x_i^j \tag{13}$$

здесь

$$\sum_{j=0}^{n} w_j\, x_i^j = wx_i + b,\ i = 1,2,\dots.,N;\ x_i^0 = 1,\ w_0 = \text{b}.$$

С учетом сущности модели логистической регрессии задачу оптимизации можно представить в следующем виде:

$$q(w) \to \underset{w}{max} => w^*, \tag{14}$$

$$f_w(x) = \frac{1}{1+e^{-\sum_{j=0}^{n} w_j x_i^j}} \geq \frac{1}{2},\ i = 1,2,\dots,N_1 \tag{15}$$

$$f_w(x) = \frac{1}{1+e^{-\sum_{j=0}^{n} w_j x_i^j}} < 1/2, \quad i = N_1+1,\dots,N \tag{16}$$

здесь $w = (w_0, w_1, w_2, \dots., w_n)$.

Полученную оптимизационную задачу (14)-(16) можно решить с помощью адаптивного метода случайного поиска, предложенного основоположником стохастических методов оптимизации Растригиным Л.А.

Адаптивный метод случайного поиска. Рассмотрим следующую задачу оптимизации в общем виде

$$q(\omega) \to max => \omega^* \quad , \quad \omega \in D \tag{17}$$

где, $q(\omega)$ — нелинейная функция многих переменных,

$\omega = (\omega_1, \omega_2, \dots., \omega_n)$,

$\omega^* = (\omega_1^*, \omega_2^*, \dots., \omega_n^*)$– решение задачи (17).

D — область определения задачи оптимизации, которая обычно может быть задана в виде равенств и неравенств.

В методе адаптивного случайного поиска используется следующая рекуррентная формула:

$$\omega^{k+1} = \omega^k + \Delta\omega^{k+1}, \tag{18}$$

$$\Delta\omega^{k+1} = \begin{cases} a^{k+1}\Delta\omega^{k}, & \text{если } q(\omega^{k}) > q(\omega^{k-1}) \\ a^{k+1}\cdot\xi^{k+1}, & \text{если } q(\omega^{k}) \le q(\omega^{k-1}) \end{cases} \quad (19)$$

где, a^{k+1} – параметр, характеризующий длину (k+1) шагов.

В четвертой главе диссертации под названием **«Оценка эффективности процесса обнаружения спам-сообщений и результатов его реализации»** представлена оценка эффективности методов выявления спам-сообщений из сообщений электронной почты, функциональная структура и принцип работы программного средства обнаружения спам-сообщений, а также результаты экспериментальных расчетов, полученные при её реализации.

В первом параграфе проведена оценка эффективности модифицированного метода опорных векторов при обнаружении спам-сообщений по трем основным показателям (таблица 1). Сюда относится спам по электронной почте (ЭПС), спам в социальных сетях (ССС), спам на форумах (ФС), спам, распространяемый через комментарии на сайтах (СШТС), спам в виде каталогов и бюллетеней (КБС), СМС-спам (СС).

Таблица 1

Результаты тестирования модифицированного метода опорных векторов обнаружения спам-сообщений

Типы спам-сообщений	Общее количество спам-сообщений	% обнаруженных спам-сообщений	% необнаруженных спам-сообщений	% ложных срабатываний системы
ЭПС	1364	97,3	2.1	0,6
ССС	1223	95,4	3,9	0,7
ФС	1393	96,3	3.2	0,5
СШТС	1676	95,1	4.1	0,8
КБС	1128	93,3	6.1	0,6
СС	1410	92,1	7.1	0,8

Таблица 2

Результаты оценки эффективности модифицированного метода опорных векторов для обнаружения спам-сообщений

Системы обнаружения спама	Общее количество спам-сообщений	% обнаруженных спам-сообщений	% необнаруженных спам-сообщений	% ложных срабатываний системы
Программа обнаружения спама	1665	96,2	3.4	0,4
AntiSpam Sniper Pro	1665	94,8	3.1	2.1
Malwarebytes Anti-Exploit Free	1665	93,4	4.4	2.2
Spybot Search & Destroy	1665	92,7	4,8	2,5
Spamhaus	1665	97,1	2.1	0,8

Полученные результаты сравнения эффективности систем, основанных на модифицированном методе опорных векторов обнаружения спама, с другими системами обнаружения спама представлены в таблице 2. В данном случае были протестированы 3 основных индикатора обнаружения спам-сообщений, упомянутых выше.

Таблица 3

Результаты тестирования на компьютерах с различной производительностью процессора и графического процессора для фильтрации электронной почты

Требования к устройству / Фильтрация электронных почтовых сообщений (ЭПХФ)	Intel Core i5 4,4 ГГц, 16 ГБ ОЗУ	Intel Core i7 7900K, 64 ГБ ОЗУ	Intel Core i9 7900K, 64 ГБ ОЗУ, графический процессор GTX 1080 TI
ПОС (ЭПХФ)	0,7 с	39 мс	11 мс
MailWasher Free (ЭПХФ)	1,5 с	65 мс	26 мс
SpamBytes (ЭПХФ)	0,9 с	43 мс	13 мс
Spamihilator (ЭПХФ)	0,8 с	41 мс	12 мс
Spamfence (ЭПХФ)	1,2 с	49 мс	15 мс

Во втором параграфе представлена функциональная структура программного средства обнаружения спама (рис. 6).

Как известно ответственным за информационную безопасность сотрудникам, программные средства, позволяющие фильтровать электронные письма и выявлять из них спам-сообщения, сегодня широко используются практически во всех организациях. Программный инструмент, созданный на основе разработанных алгоритмов, не ограничивается только обнаружением спам-сообщений, оно также позволяет выполнить ряд таких задач, как:

- обнаружение спам сообщений;
- сохранение спам-сообщений в карантине;
- удаление спам-сообщений;
- отправка сообщений;
- получение сообщений и т. д.

В третьем параграфе представлены результаты внедрения программного средства защиты системы электронной почты организации от спам-сообщений.

Основной целью программного обеспечения ПОС является защита электронной почты организации на основе комплексной защиты от вирусов, хакерских атак и спама, а также анализа входящей почты с использованием сложного подхода.

Программное средство фильтрации адресов электронной почты организаций и предприятий и выявление спам-сообщений внедрено в ГУП «Центр кибербезопасности», в обществе с ограниченной ответственностью «Единый интегратор по созданию и сопровождению государственных информационных систем UZINFOCOM», в обществе с ограниченной ответственностью «ASR KABEL», полученные результаты приведены ниже.

Программное средство, разработанное для фильтрации сообщений электронной почты, выявление спама и его блокировки в ГУП «Центр кибербезопасности», отфильтровало 4673 сообщения электронной почты и определило в них 1327 из 1364 спам-сообщений с точностью 97,3% .

В процессе экспериментальной апробации с целью фильтрации сообщений электронной почты и выявления спам-сообщений в ООО «UZINFOCOM» программный инструмент для выявления спам-сообщений в 3624 сообщениях электронной почты эффективно отфильтровал 1465 сообщений электронной почты и из них позволил идентифицировать 1431 спам-сообщение с точностью 97,6%.

В процессе тестирования в целях фильтрации сообщений электронной почты, а также выявления и блокировки сообщений спама в ООО «ASR KABEL» программным средством для выявления сообщений спама в сообщениях электронной почты эффективно отфильтровано 2526 сообщений электронной почты: идентифицировано 1248 сообщений электронной почты из 1280 спам-сообщений с точностью 97.5% и, что позволило успешно заблокировать их путем выявления и протоколирования ошибок.

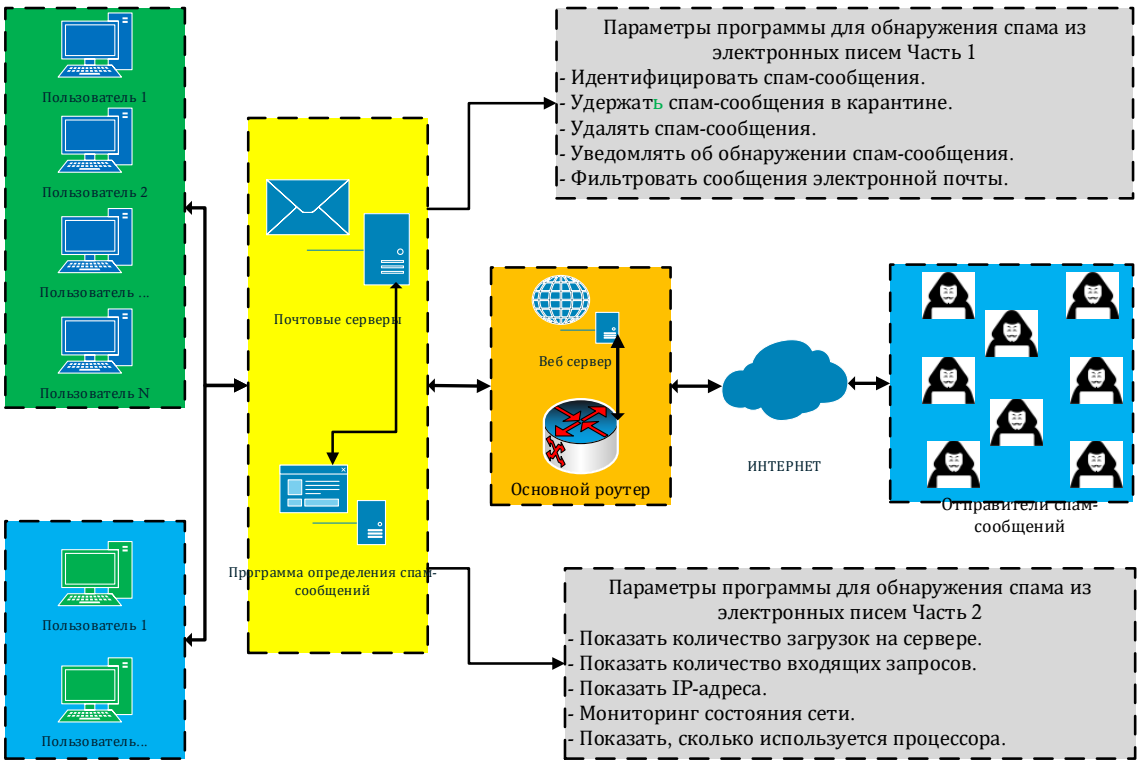


Рисунок 6. Функциональная структура программного средства для обнаружения спам-сообщений

По результатам, полученным от использования разработанного программного обеспечения на предприятиях и в организациях, можно сказать, что в результате невнимательности одного пользователя, повреждения информационной системы организации и, как следствие, негативного влияния на работу почтового сервера, а также потеря, удаление или модификация необходимой информации из базы данных организации.

Если спам-сообщения, кажущиеся обычными на первый взгляд, не будут вовремя выявлены и устранены, то по их последствиям будет открыт путь к другим системам организации. Потому что, нет ни одной организации, не имеющей адреса электронной почты. Электронная почта является одной из единственных технологий, существующих во всех современных организациях и позволяющих обмениваться информацией по всему миру. Поэтому целесообразно уделить особое внимание его безопасности, чтобы эффективно использовать эти технологии.

ЗАКЛЮЧЕНИЕ

В результате выполнения диссертационной работы на тему «Разработка алгоритмов фильтрации электронной почты и обнаружения спама» были достигнуты следующие результаты:

1. Проанализированы методы фильтрации электронной почты в сравнении с различными характеристиками, параметрами и средами. Было установлено, что метод опорных векторов обладает большой эффективностью. Сокращение количества перегрузок системы фильтрации достигнуто за счет выбора необходимых критериев, общих для всех форм спам-сообщений.

2. Разработан алгоритм пополнения базы знаний новыми знаниями в режиме реального времени, позволяющий выявлять новые виды спам-сообщений путем совместного использования существующей базы знаний в организации. В результате при резком увеличении количества знаний в базе знаний было достигнуто сокращение времени ответа на отправленные запросы на 1 мс.

3. В соответствии с многоуровневой иерархией базы знаний системы обнаружения спама разработан трехэтапный алгоритм фильтрации документов. В результате удалось оценить итоговую вероятность принадлежности электронного сообщения к спам-сообщению.

4. Разработан модифицированный метод опорных векторов и алгоритм обнаружения спам-сообщений. В результате стало возможным на основе определенного описания документов F-классификатора разделить объекты на два класса.

5. Программное обеспечение для обнаружения спама, разработанное на основе модифицированного метода опорных векторов, показало точность 97,6%.

6. С помощью разработанного программного обеспечения для обнаружения спама время, необходимое для идентификации спам-сообщений

из электронной почты, составило 11 мс. Было обнаружено, что оно на 1 мс быстрее, чем другие подобные программные инструменты.

Программный инструмент фильтрации сообщений электронной почты и обнаружения спам-сообщений, разработанный на основе предложенных методов, позволяет снизить вероятность ложного срабатывания системы защиты, используемой для защиты информации в организации, в 1,4 раза и увеличить скорость работы в 1,1 раза, а также обеспечить защиту сообщений электронной почты в режиме 24/7. Для повышения эффективности обнаружения спама рекомендуется постоянно обновлять базу данных спама.

SCIENTIFIC COUNCIL AWARDING SCIENTIFIC DEGREES DSc.13/30.12.2019.T.07.02 AT TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES

TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES

HAYDAROV ELSHOD DILSHOD UGLI

DEVELOPMENT OF ALGORITHMS FOR FILTERING EMAILS AND DETECTING SPAM MESSAGES

05.01.05 – Methods and systems of information protection. Information security.

DISSERTATION ABSTRACT OF THE DOCTOR OF PHILOSOPHY (PhD) ON TECHNICAL SCIENCES

Tashkent-2023

The theme of doctor of philosophy (PhD) on technical sciences was registered at the Supreme attestation commission at the Cabinet of Ministers of the Republic of Uzbekistan under number B2022.4.PhD/T3425.

The dissertation has been prepared at Tashkent University of Information Techologies.

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

Scientific adviser:	**Khamdamov Rustam Khamdamovich** doctor of Technical Sciences, professor
Official opponents	**Kerimov Kamil Fikratovich** doctor of technical sciences, assistant professor
	Nasrullayev Nurbek Bakhtiyorovich doctor of Physical and Mathematical sciences, assistant professor
Leading organization:	**Scientific-Engineering and Marketing researches Center «UNICON.UZ»**

The defense will take place «____» _____ 2023 at _____ the meeting of Scientific council No. DSc.13/30.12.2019.T.07.02 at Tashkent University of Information Technologies (Address: 100084, Tashkent city, Amir Temur street, 108. Tel.: (+99871) 238-64-43, e-mail: tuit@tuit.uz).

The dissertation can be reviewed at the Information Resourse Centre of the Tashkent University of Information Technologies (is registered under No.___). (Address: 100084, Tashkent city, Amir Temur street, 108. Tel.: (+99871) 238-64-70).

Abstract of dissertation sent out on «____» _____ 2023 y.
(mailing report No. ___ on «____» _____ 2023 y.).

B.Sh. Makhkamov
Chairman of the scientific council
awarding scientific degrees,
doctor of economical sciences, professor

E.Sh. Nazirova
Scientific secretary of scientific council
awarding scientific degrees,
doctor of technical sciences, professor

S.K. Ganiev
Chairman of the academic seminar under the
scientific council awarding scientific degrees,
doctor of technical sciences, professor

INTRODUCTION (abstract of PhD thesis)

The aim of the research work is to develop methods and algorithms for detecting spam messages, allowing to increase the efficiency of the e-mail protection system.

The object of the research work is a spam email message.

The scientific novelty of the research work is as follows:

uniform criteria have been formed for e-mail messages with extensive capabilities and many properties, resulting in the extraction of features that are common to spam messages of all kinds;

taking into account the frequency of use of knowledge in the existing knowledge base in the organization, the most used group of knowledge was formed, as a result, an algorithm for filling the knowledge base was developed that allows identifying new types of spam messages in real time;

taking into account the types of documents in e-mail, a multi-level knowledge base on file extension among these documents was formed, as a result, an algorithm for filtering documents in e-mail was developed, consisting of three stages, using a multi-level knowledge base of the spam detection system;

developed a modified base vector method and an algorithm for detecting spam messages based on the learning method with a training set;

summing the method of logistic regression with a flexible method of random search, an algorithm for detecting spam documents was developed by finding the probability of the largest spam message among the message.

Implementation of the research results. Based on scientific results of email filtering and spam detection algorithms and software tools:

a software tool developed on the basis of algorithms for filtering e-mail messages and detecting spam messages has been introduced into the practice of the State Unitary Enterprise "Cybersecurity Center" (certificate of the Ministry for the Development of Information Technologies and Communications of October 18, 2022, No. 33-8/6761). As a result of scientific research, 4673 email messages from the corporate network were filtered and 1327 spam messages were filtered out of 1364.

This made it possible to identify with an accuracy of 97.3% and an error of 2.7%.

a software tool developed on the basis of algorithms for filtering e-mail messages and detecting spam messages has been introduced into the practical activities of the limited liability company "Single integrator for the creation and maintenance of state information systems UZINFOCOM" (certificate of the Ministry for the Development of Information Technologies and Communications of October 18, 2022, No. 33-8/6761). As a result of scientific research, it was possible to effectively filter 3624 email messages in the local network of the organization and identify 1431 out of 1465 spam messages with an accuracy of 97.6% and an error of 2.4%;

a software tool developed on the basis of email filtering and spam detection algorithms has been introduced into the practical activities of ASR KABEL Limited

Liability Company (certificate of the Ministry for the Development of Information Technologies and Communications of October 18, 2022, No. 33-8/6761). The scientific study effectively filtered 2,526 corporate emails and successfully blocked 1,248 out of 1,280 spam emails with 97.5% accuracy and 2.5% error rate.

Structure and volume of the dissertation. The dissertation consists of an introduction, four chapters, a conclusion, a list of references and appendices. The volume of the dissertation is 112 pages.

E’LON QILINGAN ISHLAR RO‘YXATI
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I bo‘lim (I часть; I part)

1. R.Khamdamov, E.Haydarov. Mathematical Model and Methods for Filtering an Email Message // International Conference on Information Science and Communications Technologies: Applications, Trends and Opportunities (ICISCT 2021). Tashkent-2021. -4p. (3) Scopus (ОАК раёсатининг қарори 30.09.2021 йил №525)
2. R.Khamdamov, E.Haydarov. Detecting spam messages using the naive Bayes algorithm of basic machine learning // International Conference on “Information Science and Communications Technologies: Applications, Trends and Opportunities (ICISCT 2021). Tashkent-2021. -3p. (3) Scopus (ОАК раёсатининг қарори 30.09.2021 йил №525)
3. Bekmuratov T.F., Botirov F.B., Haydarov E.D. Electronic spam filtering based on neural networks // Chemical technology. Control and management, 2020, №3(93), p. 59-65. (05.00.00; №12)
4. Р.Хамдамов, Э.Ҳайдаров, Электрон почта хабарларини фильтрлашда зарур бўлган мезонларни танлаш // “Илмий хабарнома Физика –математика тадқиқотлари” журнали. 2021/№2(3). Андижон-2021. –Б. 86-89. (13.00.00; №12)
5. Э.Ҳайдаров, Электрон почта хабарларидаги спам хабарларни фильтрлаш иерархияси // “Муҳаммад ал-Хоразмий авлодлари” журнали. № 3(21)/2022. –Б. 187-191. (05.00.00; №10)
6. Р.Хамдамов, Э.Ҳайдаров, Электрон почта хабарларида номақбул хабарларни аниқлашда фильтрлаш алгоритмини ишлаб чиқиш // “Муҳаммад ал-Хоразмий авлодлари” журнали. № 2(20)/2022. –Б. 3-7. (05.00.00; №10)

II bo‘lim (II часть; II part)

7. Khamdamov R.H., Ibrohimov A.R., Haydarov E.D., Logistik regressiya asosida tasniflash masalalarini yechish // “RESEARCH AND EDUCATION” Multidisciplinary Scientific Journal. ISSN: 2181-3191 VOLUME 1, ISSUE 9, Impact Factor 2022:4.628, December 2022. –B. 162-171.
8. Khamdamov R.Kh., Ibrohimov A.R., Haydarov E.D., Tayanch vektorlar usuli yordamida spam hujjatlarni aniqlash algoritmi // “RESEARCH AND EDUCATION” Multidisciplinary Scientific Journal. ISSN: 2181-3191 VOLUME 1, ISSUE 9, Impact Factor 2022:4.628, December 2022. –B. 152-161.
9. R.Xamdamov, E.Haydarov, Elektron pochta xabarlarini filtrlashda zarur bo‘lgan mezonlarini tanlash // Central Asian Journal of Education and Computer Sciences (CAJECS). VOLUME 1, ISSUE 3, August 2022. –B. 17-21.

10. E.Haydarov. Filling the Knowledge Base of the Filtering System algorithm Development // 2nd International Conference on Pervasive Computing and Social Networking (ICPCSN 2022). Salem, India March-2022. -7p. (3) Scopus

11. E.Qahramonov, E.Haydarov, Elektron pochta tizimini filtrlashda xabarlarning xususiyatlari tahlili // “Science and Pedagogy in the modern world: problems and solutions” Conference of England-2023. VOLUME 1, ISSUE 2. –B. 63-68. (https://doi.org/10.5281/zenodo.7639026)

12. E.Haydarov, Elektron pochta tizimlari va ularda axborot xavfsizligi muammolari // “Havo hujumidan mudofaa tizimidagi mutaxassislarni tayyorlashda axborot-kommunikatsiya texnologiyalaridan foydalanish” mavzusidagi xalqaro onlayn ilmiy-amaliy konferensiyasi maqolalar to’plami. I to‘plam, Toshkent-2021. –B. 444-447.

13. E.Haydarov, Spam xabarlarni aniqlashning mashinali o‘qitish usuli // «Professional armiya boshqaruvini rivojlanishida innovatsiyalar va raqamlashtirishning o‘rni» IV Xalqaro ilmiy-amaliy konferensiyasi ma’ruzalar to‘plami. Toshkent-2022. –B. 83-85.

14. Э.Ҳайдаров, Электрон почта ва ундаги спам хабарлар муаммоси // International Conference “ICT In education: challenges and solutions” Tashkent-2021 – Б . 110-113.

15. Э.Ҳайдаров, Э.Рофиева, Elektron pochta orqali spam xabarlarni tarqatish bosqichlari // “Иқтисодиёт тармоқларининг инновацион ривожланишида ахборот-коммуникация технологияларининг аҳамияти” Республика илмий-техник анжуманининг маърузалар тўплами, 2-қисм. Тошкент-2021. –Б. 302-305.

16. Р.Хамдамов, Э.Ҳайдаров, Электрон почта хабарларини филтрлаш усуллари // «Ахборот коммуникация технологиялари ва дастурий таъминот яратишда инновацион ғоялар» Республика илмий-техник конференцияси маърузалар тўплами. II-том. Самарқанд-2021. –Б. 195-198.

17. Э.Ҳайдаров Спам омиллари ва уларни антиспам усуллари билан боғлиқлиги // “Zamonaviy axborot, kommunikatsiya texnologiyalari va AT-ta’lim tatbiqi muammolari” mavzusidagi respublika ilmiy-amaliy anjumani ma’ruzalar to‘plami, I tom. Samarqand-2021. –B. 234-236.

18. R.Xamdamov, E.Haydarov, Электрон почта хабарларини фильтрлаш усуллари // «Ахборот коммуникация технологиялари ва дастурий таъминот яратишда инновацион ғоялар» Республика илмий-техник конференцияси маърузалар тўплами. II-том. Самарқанд-2021. –Б. 195-198.

19. E.Haydarov, Sh.Sayfullayev, Neyron tarmoq algoritmlari asosida spam xabarlarni filtrlash // “Iqtisodiyot tarmoqlarining innovatsion rivojlanishida axborot kommunikatsiya texnologiyalarining ahamiyati” Respublika ilmiy-texnik anjumani ma’ruzalar to‘plami, 1-qism. Toshkent-2022. -B. 313-315.

20. Khamdamov R.Kh., Khaydarov E.D., Pre-processing of primary spam classification data from email messages // Современное состояние и перспективы развития цифровых технологий и искусственного интеллекта Сборник докладов республиканской научно-технической конференции Самарканд, 26-27 октября 2022 г., pp. 238-242.

21. R.X.Xamdamov, E.D.Haydarov, Электрон почта хабарларидаги спам хабарларни аниқлаш дастури// Dasturga guvohnoma № DGU 13397. Toshkent 04.12.2021.

22. R.X.Xamdamov, E.D.Haydarov, Электрон почта хабарларидаги спам хабарларни блоклаш дастури // Dasturga guvohnoma № DGU 13398. Toshkent 04.12.2021.

23. R.X.Xamdamov, E.D.Haydarov, Спам хабарлар статистикасини шакллантириш дастури // Dasturga guvohnoma № DGU 13399. Toshkent 04.12.2021.

24. T.F.Bekmuratov, A.A.Ganiyev, F.B.Botirov, E.D.Haydarov, SPAM SCANNER // Dasturga guvohnoma № DGU 08872. Toshkent 28.08.2020.

Avtoreferat «Muhammad al-Xorazmiy avlodlari» ilmiy jurnali tahririyatida tahrirdan o‘tkazildi va o‘zbek, rus va ingliz tillaridagi matnlarini mosligi tekshirildi.

Bosmaxona litsenziyasi:

9338

Bichimi: 84x60 $^{1}/_{16}$. «Times New Roman» garniturasi.
Raqamli bosma usulda bosildi.
Shartli bosma tabog‘i: 2,5. Adadi 100 dona. Buyurtma № 28/23.

Guvohnoma № 851684.
«Tipograff» MCHJ bosmaxonasida chop etilgan.
Bosmaxona manzili: 100011, Toshkent sh., Beruniy ko‘chasi, 83-uy.