

ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМий ДАРАЖАЛАР БЕРУВЧИ
DSc.13/30.12.2019.Т.07.01 РАҚАМЛИ ИЛМий КЕНГАШ

САМАРҚАНД ДАВЛАТ УНИВЕРСИТЕТИ

КАРШИЕВ ХУСАН БЕРКИНБАЕВИЧ

ЭЛЕКТРОН ҲУЖЖАТ ЭЛЕМЕНТЛАРИ МАНТИҚИЙ ВА СЕМАНТИК
БОҒЛАНИШИНИ АЖРАТАДИГАН МЕХАНИЗМЛАР БЎЙИЧА
АХБОРОТ ИШОНЧЛИГИНИ ОШИРИШ

05.01.02- Тизимли таҳлил, бошқарув ва ахборотни қайта ишлаш

ТЕХНИКА ФАҢЛАР БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ

Тошкент–2021

**Техника фанлари бўйича фалсафа доктори (PhD) диссертацияси
автореферати мундарижаси**

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

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

Каршиев Хусан Беркинбаевич

Электрон ҳужжат элементлари мантикий ва семантик боғланишини
ажратадиган механизмлар бўйича ахборот ишончилигини ошириш.....3

Каршиев Хусан Беркинбаевич

Повышение достоверности информации с использованием механизмов
извлечения логических и семантических связей элементов электронных
документов19

Karshiev Khusan Berkinbayevich

Increasing the reliability of information using mechanisms for extracting logical
and semantic links between elements of electronic documents.....35

Эълон қилинган ишлар рўйхати

Список опубликованных работ

List of published works 38

ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМий ДАРАЖАЛАР БЕРУВЧИ
DSc.13/30.12.2019.T.07.01 РАҚАМЛИ ИЛМий КЕНГАШ

САМАРҚАНД ДАВЛАТ УНИВЕРСИТЕТИ

КАРШИЕВ ХУСАН БЕРКИНБАЕВИЧ

ЭЛЕКТРОН ҲУЖЖАТ ЭЛЕМЕНТЛАРИ МАНТИҚИЙ ВА СЕМАНТИК
БОҒЛАНИШИНИ АЖРАТАДИГАН МЕХАНИЗМЛАР БЎЙИЧА
АХБОРОТ ИШОНЧЛИГИНИ ОШИРИШ

05.01.02- Тизимли таҳлил, бошқарув ва ахборотни қайта ишлаш

ТЕХНИКА ФАҢЛАР БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ

Тошкент–2021

Техника фанлари бўйича фалсафа доктори (PhD) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида В2019.3.PhD/T1294 рақам билан рўйхатга олинган.

Диссертация Самарқанд давлат университетида бажарилган.

Диссертация автореферати уч тилда (ўзбек, рус, инглиз (резюме)) Илмий кенгаш веб-саҳифасида (www.tuit.uz) ва «ZiyoNet» Ахборот таълим порталида (www.ziyounet.uz) жойлаштирилган.

Илмий раҳбар:

Жуманов Исроил Ибрагимович
техника фанлари доктори, профессор

Расмий оппонентлар:

Кабулов Анвар Василевич
техника фанлари доктори, профессор
Сулюкова Лариса Фаритовна
техника фанлари доктори

Етакчи ташкилот:

Тошкент давлат техника университети

Диссертация ҳимояси Тошкент ахборот технологиялари университети ҳузуридаги DSc.13/30.12.2019.T.07.01 рақамли Илмий кенгашнинг 2021 йил «29» ОКТАБР соат 16:00 даги мажлисида бўлиб ўтади. (Манзил: 100202, Тошкент шаҳри, Амир Темур кўчаси, 108-уй. Тел.: (99871) 238-64-43, факс: (99871) 238-65-52, e-mail: tuit@tuit.uz).

Диссертация билан Тошкент ахборот технологиялари университети Ахборот-ресурс марказида танишиш мумкин (225 рақам билан рўйхатга олинган). (Манзил: 100202, Тошкент шаҳри, Амир Темур кўчаси, 108-уй. Тел.: (99871) 238-65-44).

Диссертация автореферати 2021 йил «18» ОКТАБР да тарқатилди.
(2021 йил «11» ОКТАБР даги 31 рақамли реестр баённомаси.)



Р.Х.Хамдамов

Илмий даражалар берувчи илмий кенгаш раиси,
техника фанлари доктори, профессор

Ф.М.Нуралиев

Илмий даражалар берувчи илмий кенгаш илмий
котиби, техника фанлари доктори, доцент

А.В. Кабулов

Илмий даражалар берувчи илмий кенгаш қошидаги
илмий семинар раиси, техника фанлари доктори,
профессор

КИРИШ (фалсафа доктори (PhD) диссертацияси аннотацияси)

Диссертация мавзусининг долзарблиги ва зарурати. Жаҳонда, ишлаб чиқариш самарадорлиги, технологик агрегатлар қувватини оширишга ҳамда ишлаб чиқариш мажмуаларини замонавий технологиялар асосида яратишга ва оптимал бошқаришга алоҳида эътибор берилмоқда. Ривожланаётган фан соҳасида энг муҳим ва долзарб масалалардан энергетик ҳамда турли хил ресурсларнинг кам харажатида сифатли маҳсулотлар ишлаб чиқаришни юқори даражада ошириш имконини берувчи рақамли технологияларни ҳар хил иқтисодий объектлар бошқарувида қўлланиши муҳим масалалардан бири саналади. Шу билан бир қаторда, автоматлаштириш тизимлар яратилишини тадқиқ қилиш ҳамда такомиллаштириш, электронлаштириш, инновацион ғоя ва технологиялар, электрон ҳужжат алмашуви тизимлари (ЭҲАТ), замонавий математик усуллар, ахборотга тезкор ишлов берувчи технологияларни иқтисодиётнинг турли соҳаларида қўллаш ва ривожлантиришга жиддий эътибор берилмоқда.

Дунёда Microsoft, Lucent, Lernout&Hauspie, Unisys хорижий фирмалари, NLU – мулоқат технологияси, электрон почта, ўқитиш воситалари, табиий тил орфографиясини назорат ва таҳрир қилиш, машина таржимаси, нутқ, матн синтези ва таҳририни бажарувчи дастурий воситалар, ташкилий-фармойишли ҳужжатга ишлов берувчи ҳамда маълумот базаси (МБ) ва билимлар базасини (ББ) шакллантирувчи ишланмалар маълум. Замонавий ЭҲАТ калит бўғинларидан ҳисобланган энг оммабоп тизимлар Google, Bing, Yahoo, рус тили - Yandex, Rambler, корпоратив тизимлардан Sphinx, Lucene, Харіан қўлланилиши мутахассисларга маълум. Бу соҳада, АҚШ, Россия Федерацияси, Англия, Франция, Германия, Испания, Хитой, Ҳиндистон, Жанубий Корея, Япония, Украина, Беларусия, Латвия ва бошқа давлатларда тадқиқотлар жадал бажарилмоқда.

Ўзбекистон Республикасида бошқарув тизимларини ривожлантириш, ҳамда ягона информацион муҳитни пайдо қилиш мақсадида қабул қилинган чора-тадбирлар мажмуаси доирасида, жиддий эътиборни замонавий ахборот - коммуникация технологиялари, математик усул, моделларга асосланган тизимларни яратишга йўналтирилган. 2017-2021 йилларда Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегиясида, жумладан, «... юқори технологияли қайта ишлаш тармоқларини, энг аввало, маҳаллий хомашё ресурсларини чуқур қайта ишлаш асосида юқори қўшимча қийматли тайёр маҳсулот ишлаб чиқаришни жадал ривожлантириш ...»¹ тараққиётни юксалтириш вазифалари белгилаб олинган. Шу қаторда, ахборотлар ишончлигини оширишнинг долзарб муаммоси ечимига йўналтирилган ёндашув, усул, моделлар, ҳамда қарор қабул қилишни оптималлаштириш учун ҳужжатлар ҳимояланиши, релевантлиги, тўлалигига жиддий эътибор қаратилган. Ахборот ишончлиги пасайиши, энг аввало, ҳужжатни тайёрлаш ва киритиш, кейинча, узатиш ва хотиралаш, ҳамда қайта

¹ Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги ПФ-4947-сон «Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида» ги Фармони

ишлаш, фойдаланувчилар звеноларида содир бўлади. Ахборот ишончсизлиги инсон-оператор хатолари, сканерловчи ва матнни таниш воситалари хатоликлари, алоқа каналида помехалар таъсири, электрон ускуналар ишончсизлиги туфайли ифодаланади. Ахборот ишончлигини ошириш муаммоси ечимини ташкил қилувчи илмий-услубий асослар, маълумотлар танқислиги, параметрли ноаниқлик, жараёнлар ностационарлиги мавжуд шартларда хатоларни аниқлаш ва тузатишни амалга ошириш дастурий воситаларини кам харажатларда қўллаш имкониятини таъминлайди.

Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги ПФ-4947-сон «Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси», 2018 йил 19 феврал ПФ-5349-сон «Ахборот технологиялари ва коммуникациялари соҳасини янада такомиллаштириш чора-тадбирлари» Фармонлар, 2017 йил 29 августдаги ПҚ-3245-сон «Ахборот-коммуникация технологиялари соҳасида лойиҳа бошқаруви тизимини янада яхшилашга доир чора-тадбирлар тўғрисида»ги Қарори ҳамда мазкур фаолиятга тегишли бошқа меъёрий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишга ушбу диссертация тадқиқоти муайян даражада хизмат қилади.

Тадқиқотнинг республика фан ва технологиялари ривожланишининг устувор йўналишларига мослиги. Мазкур тадқиқот республика фан ва технологиялар ривожланишининг IV. «Ахборотлаштириш ва ахборот-коммуникация технологияларини ривожлантириш» устувор йўналиши доирасида бажарилган.

Муаммонинг ўрганилганлик даражаси. Сўнгги 10-15 йиллар илмий адабиётлари таҳлили ахборотни ишончли узатиш назарияси ривожига Ch.W.Bachman, W.W.Peterson, D.Ritchi, E.F.Codd, C.J.Date, Alan Turing, E.N.Gilbert, R.C.Bose, Danish Jamil, Peter Pin-Shan Chen, Чжо Зо Е, В.Б.Кудрявцев, Д.А.Полянский, Э.Э.Гасанов, А.И.Репин, А.Н.Савельев, бошқа мамлакат олимларининг катта ҳиссаси борлигини кўрсатади.

Ўзбекистонда ахборотга оптимал ишлов бериш назарияси ривожига Д.А.Абдуллаев, Ф.Б.Абуталиев, Т.Д.Раджабов, Т.Ф.Бекмуратов, М.М.Камилов, Н.Р.Юсупбеков, Х.З.Игамбердиев, И.И.Жуманов, Р.Х.Хамдамов, М.А.Рахматуллаев, Ш.Х.Фазилов, А.А. Саидов, А.Р. Ахатов ва бошқа олимлар салмоқли ҳисса қўшган.

Ахборот ишончлигини оширишнинг замонавий назарияси сунъий ахборот ортиқчалигидан фойдаланишга асосланган кодлаштирувчи ва аппаратли усуллардан иборат ва ушбу йўналишдаги тадқиқотлар кўлами кўп сонли монографик ишлар, диссертациялар ва илмий мақолаларда акс эттирилган. Ахборот-коммуникация тармоқларида сунъий ахборот ортиқчалигидан фойдаланувчи кўп сонли маълумотлар узатиш протоколлари (МУП) мавжуд. Уларда катта ҳажмдаги ресурслар фақат МУП сарлавҳасини химоя қилиш ва узатишда сарфланади, аксарият хабарлар узатиш сеанслари давомида ўзгаришсиз қолади. Айниқса таъкидлаш керакки, инсон – оператор, сканер ва таниш воситалари, фойдаланувчи орфографик хатолари OSI модели Тадбиқий ва Намойиш босқичларида намоён бўлади ва улар умумий

хатолар ҳажмида катта қисми ташкил этади, бу воқелик эса, замонавий МУПда инобатга олинмайди. Шу туфайли, ахборот ишончилигини ошириш имконини яратадиган ҳужжат элементлари орасидаги статистик, корреляция, мантиқий - семантик, технологик боғланишлар сабабли пайдо бўладиган ахборот ортиқчалигидан фойдаланишга асосланган модел, алгоритм, умуман, механизмлар етарли даражада ўрганилмаган.

Диссертация тадқиқотининг диссертация бажарилган олий таълим муассасасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги. Диссертация иши Самарқанд давлат университети ва Тошкент ахборот технологиялари университети Самарқанд филиалининг А5-039-«Маълумотларни интеллектуал таҳлил қилишнинг параллел алгоритмларини куриш дастурий воситаларини ишлаб чиқиш» (2015-2017) мавзусидаги лойиҳа доирасида бажарилган.

Тадқиқотнинг мақсади турли табиатдаги ахборот ортиқчалигидан фойдаланишга асосланган ахборот ишончилигини оширувчи модел, алгоритм ва дастурий мажмуани ишлаб чиқишдан иборат.

Тадқиқотнинг вазифалари:

замонавий бошқарув тизимларида электрон ҳужжатларга ишлов бериш, ахборот ишончилигини ошириш чораларини такомиллаштиришга йўналтирилган илмий - услубий асослар, модел ва алгоритмларни яратиш;

ахборот ишончилигини таъминлаш модел, алгоритмларини ҳужжат элементлари эҳтимолий-статистик боғланишига таянган ёндошув бўйича рухсат этилган тўплам қисмига мувофиқлиги, кластерга мансублиги, эталон – жамламага ҳосиллигини амалга оширувчи механизмларни ишлаб чиқиш;

ЭҲ концептлари муносабати, элементлари мантиқий, семантик боғланишларини аниқловчи, матн статистик хусусиятларидан фойдаланувчи услубият, фрактал, фрейм тавсифлари бўйича ахборот ишончилигини оширувчи механизмларга асосланган алгоритмларни ишлаб чиқиш;

ахборот ишончилигини оширишни баҳоловчи усулларни оптималлаштириш, алгоритмлар самарадорлигини мураккаблик, харажат нархи қиймати коэффицентлари бўйича таҳлил қилиш;

ҳужжат элементларининг статистик, мантиқий, семантик, технологик боғланишларига таянган алгоритмлар, функционал модуллар, дастурий мажмуаларни ишлаб чиқиш, синовдан ўтказиш ва ва жорий қилиш.

Тадқиқотнинг объекти информацион жараёнлар, иш юритиш ҳужжатлари, ЭҲАТ, корпоратив тармоқлар.

Тадқиқот предмети турли табиатли ахборот ортиқчалигидан фойдаланишга асосланган ахборот ишончилигини ошириш усул, модел, алгоритмлари ва дастурий таъминоти.

Тадқиқотнинг усуллари. Тизимли таҳлил, маълумотларни узатиш, маълумот қидириш, таниш, кластерлаш, компьютер лингвистикаси, алгоритмлаш ва дастурлаш усуллари ва назариялари.

Тадқиқотнинг илмий янгилиги қуйидагилардан иборат:

ҳужжат элементлари статистик, мантиқий, семантик, таркибий-технологик боғланиши ва концептлар муносабати сабабли, ҳосил бўлувчи

ахборот ортиқчалигидан фойдаланувчи ахборот ишончилигини ошириш механизмларини яратиш илмий услубиятлари ишлаб чиқилган;

ахборот хатоларини аниқлаш ва тузатиш учун кетма-кетликда ишончли ва ишончсиз элементларга, сегментларга ажратувчи, эквивалентлик коэффициентини мувофиқлаштирувчи, назорат қилиш бўсағаси ва жамламаси ўлчамини оптималлаштирувчи алгоритмлар ишлаб чиқилган;

ахборот хатоларини назорат ва тахрир қилиш масалаларининг умумий ва хусусий ечимлари, n-грамм шартли эҳтимоллар баҳосини берувчи услубият олинган ҳамда кўп босқичли морфолигик ва таркиблаштирилган n-грамм таҳлилга асосланган механизмлар Фреймворк Сфинкс-4 технологияси муҳитида синтезлаштирилган;

ахборот ишончилигини ошириш алгоритмларининг изим самарадорлигига қўшган ҳиссаси хос мақсад функция ҳамда экстремал назорат чегаралари бўйича оптималлаштирилган.

Тадқиқотнинг амалий натижалари қуйидагилардан иборат:

Ахборот ишончилигини ошириш дастурий мажмуаси (ДМ), корхона ва муассаса ЭҲАТи таркибида идентификациялаш, назорат, ўзгарувчиларни мувофиқлаш, маълумотлар фойдали хусусиятларини ажратиш функцияларини амалга оширувчи алгоритмлар ҳамда оригинал функционал модуллар яратилган;

ДМ самарадорлиги реал шароитларда синовдан ўтказилган, натижада ахборот ишончилиги икки-уч тартибга оширилиши, назорат қилиш мураккаблик ва харажат нарх-қиймат коэффициентлари 7-8 мартобага камайиши асосланган.

Тадқиқот натижаларининг ишончилиги. Ишлаб чиқилган дастурий таъминот тўплами ва унинг модулларининг тўғри қўлланилиши, реал шароитларда ахборот ишончилигини ошириш муаммоларини ечишда олинган назарий ва амалий натижаларнинг мувофиқлиги, шунингдек, қабул қилинган мезонларга мос амалга оширилган алгоритмларнинг самарадорлигини қиёсий таҳлил қилиш билан исботланади.

Тадқиқот натижаларининг илмий ва амалий аҳамияти. Ишнинг назарий аҳамияти, ҳужжатларда концептлар муносабатлари ва элементлар орасидаги статистик, корреляция, мантиқий, семантик, таркибий-технологик боғланишлари сабабли пайдо бўладиган ахборот ортиқчаликдан фойдаланиш бўйича ахборот ишончилигини оширишга чамаланган ёндашув ва механизмларни қўллаш асосида изоҳланади.

Ишнинг амалий аҳамияти ахборот ишончилигини оширишга йўналтирилган модел, алгоритмлар, типик қидириш, таниш, кластерлаш, синтезлаш воситаларидан фойдаланиш асосида турли корхоналар ЭҲАТида қўлланилишига чамаланган дастурий мажмуани яратишдан иборат. Ахборот ишончилигини оширувчи дастурий мажмуа самарадорлиги ҳужжат элементлари орасидаги статистик, мантиқий, семантик, технологик боғланишлардан фойдаланувчи, матнни парсинг-кодловчи, “Сфинкс-4” муҳитида “CUDA” параллел ҳисоблаш дастурий пакети бўйича барча амалларни бажарувчи механизмлар яратилиши бўйича изоҳланади.

Тадқиқот натижаларининг жорий қилиниши. Диссертация тадқиқоти бўйича олинган натижалар модел, алгоритм, дастурий мажмуа (ДМ), функционал модуллар ва дастурий воситалар асосида:

электрон ҳужжат алмашиш тизимига кирувчи: имловий хатоларини аниқлаш ва тузатиш, керакли сўз луғатларини кидириш, таниш, синфлаштириш, идентификациялаш, ахборот ишончилигини табиий ортиқчалик асосида назорат қилиш функциялари Самарқанд вилоят соғлиқни сақлаш бошқармаси девонхона фаолиятида жорий қилинган (Ахборот технологиялари ва коммуникацияларини ривожлантириш вазирлигининг 2020 йил 23 декабрдаги 33-8/7849-сон маълумотномаси). Натижада ДМ таркибида ахборот ишончилигини ошириш, информатив элементларни танлаш ва шакллантириш, ўзгарувчилар қийматини мослаштириш, ҳамда оптималлаштириш дастурий модулларининг қўлланиши, ташкилот бир ойлик ҳужжати ва ахборотлари ҳажми бўйича ҳисобланганда сарфланадиган моддий харажат кўламини ўртача 20% қисқартириш имконини берган;

кидирув, таниш, кластерлаш дастурий модуллари ДМ таркибига синтезлаштирилган ҳамда «ЎЗБЕКТЕЛЕКОМ» АЖ Самарқанд филиалида жорий этилган (Ахборот технологиялари ва коммуникацияларини ривожлантириш вазирлигининг 2020 йил 23 декабрдаги 33-8/7849-сон маълумотномаси). Жорий қилинган ДМ, анъанавий визуал технологияга нисбатан ҳужжат ахборотини назорат қилишда аниқланган умумий вақтни 10% гача, хатоларни аниқлаш ва тузатиш ўртача вақтини 4,39 мин дан 3,3 мин гача, бир ҳужжатни қайта ишлаш вақтини 6 мин дан 1,5 мин гача камайтириши; ахборот ишончилигини уч тартибгача ошириши; ахборотга ишлов беришда мураккаблик ва харажат нархи қиймати коэффицентларини 7-8 маротобага камайтиради.

Тадқиқот натижаларининг эълон қилинганлиги. Диссертация мавзуси бўйича 26 та илмий мақола чоп этилди, шулардан 2 таси Scopus маълумотлар базасида индексланган журналларда, 8 таси республика журналларида; 3 та ЭҲМ дастурларини рўйхатдан ўтказганлик тўғрисидаги гувоҳномалар олинди.

Тадқиқот натижаларининг эълон қилинганлиги апробацияси. Тадқиқот натижалари 3 халқаро ва 4 республика анжуманларида муҳокамадан ўтказилган.

Диссертациянинг тузилиши ва ҳажми. Диссертация кириш, тўртта боб, хулоса, фойдаланилган адабиётлар рўйхати ва иловалардан иборат. Диссертация ҳажми 119 бетни ташкил этади.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Кириш қисмида мавзунинг долзарблиги ва зарурияти асосланган, тадқиқот мақсади, вазифалари, объекти ва предметлари тавсифланган, Республика фан ва технологиялари тараққиётининг устувор йўналишларига мослиги кўрсатилган, илмий янгилиги ва амалий натижалари ҳамда уларнинг назарий, амалий моҳияти талқин қилинган, тадқиқот натижаларини жорий қилиш ҳолати, нашр этилган ишлар бўйича маълумотлар келтирилган.

Диссертациянинг **“Ҳужжат элементлари статистик боғланиши бўйича ахборот ишончилигини ошириш илмий – услубий асослари”** номли биринчи бобида турли табиатли ахборот ортиқчалиқдан фойдаланиш концептуал принципларига асосланган ахборот ишончилигини оширишнинг мавжуд методологияси назарий жиҳатдан таҳлил қилинган.

Ҳужжат концепти муносабати ва элементлари статистик, корреляция, мантиқий, семантик ва таркибий-технологик боғлиқликларидан фойдаланишга асосий ёндашувлар ва ортиқчалиқни аниқловчи моделлар алоҳида кўрсатилган.

Статистик, мантиқий-семантик, таркибли-технологик боғланишларни ажратувчи механизмларга асосланган ёндашув, принцип, ЭҲАТда ахборот ишончилигини ошириш услублари ишлаб чиқилган, бу эса биргаликда, ахборотга оптимал ишлов бериш илмий - услубий асосларини ташкил этади.

Ахборот ортиқчалигидан ишончликни оширишда фойдаланувчи механизмлар содда воситаларни тақдим этади, мавжуд усуллардаги камчиликларни бартараф қилади ҳамда ишончликни юқори қийматга оширади, маълумотга ишлов бериш мураккаблиги ва харажат нарх қийматини эса сезиларли даражада камайтиради.

Ахборот ишончилигини оширишга чамаланган усул, модел, алгоритмлар туркуми ва амалий иловаларнинг онтологияси таклиф этилган.

Биринчи гуруҳ усуллар, ҳужжат концептлари муносабати ва элементларининг статистик, корреляция, мантиқий, семантик, таркибий - технологик боғланишидан фойдаланувчи механизмларни қўллашга асосланиб ахборот ишончилигини оширишга йўналтирилган. Иккинчи гуруҳ усуллар, ҳужжат матнларидан хусусий билим, текстура хоссаси, калитли сўз, шаблон – эталон, фрактал, статистик таснифларни ажратувчи механизмларга асосланиб ахборот ишончилигини оширишга йўналтирилган.

Табиий тилларда берилган матн орфографиясини графаматик, лингвистик, морфологик таҳлил моделлари асосида назорат қилиш усуллари тадқиқ қилинган. Тенгсиз эҳтимоллар тақсимоти, статистик ва корреляция боғланишлари туфайли ҳосил бўлувчи шартли энтропия баҳолари ҳамда матн n – граммли графаматик таҳлилинини қўллаш саволлари тадқиқ қилинган. Монограмм, диграмм, учграмм ва k – граммли энтропиялар бўйича ахборот ортиқчалиги аниқланган.

Матн k-граммли шартли энтропияси қуйидагича аниқланган

$$\tilde{H}(x_{ij}^{(1)}, x_{ij}^{(k)}) = - \sum_{\xi_{ij}^{(1)}, \xi_{ij}^{(k)}} \tilde{P}(\xi_{ij}^{(1)}, \xi_{ij}^{(k)}) \log \tilde{P}(\xi_{ij}^{(1)}, \xi_{ij}^{(k)}), \quad \xi_{ij}^{(1)}, \xi_{ij}^{(k)} \in x_{ij}^{(1)}, x_{ij}^{(r)}$$

k-граммлар ўлчови; $\tilde{P}(\xi_{ij}^{(1)}, \xi_{ij}^{(k)})$ – k – граммли эҳтимол.

Хатоликлар ўртача эҳтимолини инобатга олганда талаб этилган ортиқчалик қуйидаги кўринишда аниқланади $R(S_i) = 1 - \frac{1}{(1-P)} \log \frac{(2^m - 1)}{nP}$.

Хужжат элементлари ахборот ортиқчалигининг қуйидаги: 0,5 – статистик боғланишлар; 0,3 – мантиқий боғланишлар; 0,6 – семантик боғланишлар; 0,8 – технологик боғланишлар мавжуд қийматларида ахборот ишончилигини талаб этилган даражада ошириш мумкинлиги исботланган. ЭХ элементлари статистик, корреляция, мантиқий, семантик ва таркибий - технологик боғланишларидан фойдаланувчи механизмларга асосланган алгоритмлар ўз самарадорлиги ва шаффоф қўлланувчи хатоларни аниқлаш ва тузатиши билан фарқланишлиги асосланган.

Асосий информация бўғинларида содир этиладиган хатолар статистикаси тадқиқ қилинган. Хатоликлар ўртача эҳтимоли қийматлари алоқа каналлари помехалари туфайли $P_1 = 5 \cdot 10^{-4}$, оператор хатолари сабабли $P_2 = 1,9 \cdot 10^{-3}$, имловий хатолар бўйича $P_3 = 10^{-4}$, сканер ва таниш воситалари хатоликлари сабабли $P_4 \approx P_5 \approx 10^{-2} - 10^{-3}$ бўлишлиги тажрибалар бўйича аниқланган.

Хужжат элементлари статистик боғланишидан фойдаланувчи ахборот ишончилигини ошириш алгоритмлари ишлаб чиқилган. Ахборот назорати шароитининг турли ҳолатлари учун хатоларни аниқламаслик эҳтимоли баҳолари олинган. Масалаларни ечиш вариантлари синтези амалга оширилганда хатоларни аниқламаслик умумий эҳтимоли қуйидагича баҳоланади $P_H \approx 3,9 \cdot (1 - P_{\alpha_i})^2 \cdot 10^{-5} + 3,2 \cdot P_{\alpha_i} \cdot (1 - P_{\alpha_i}) \cdot 10^{-2}$. Бу ҳолда, алгоритмлар ахборот ишончилигини икки тартибгача оширади.

Хужжат концепти ва элементлари жамламасини қамровчи МБни шакллантириш ва фойдаланиш; статистик параметрлар, динамик, хусусий таснифлар ва тақсимот қонуниятлари маълумотлари; информацияни турли назорат ва таҳлил қилиш, статистик, мантиқий-лингвистик боғланишлар ва хусусиятлар; хужжат концепти элементлари қийматининг умумий интервалини сегмент, гуруҳ, синфларга ажратувчи қоидалардан фойдаланишга асосланган ахборот ишончилигини оптималлаштириш масалалари ечилган. Сегментлар чегараларини белгилаш ва мослаштириш қоидалари, хужжат элементи қийматининг рухсатга эга тўплам қисмига тегишлилиги функция ва интервалини аниқлаш қоидалари таклиф этилган тҳамда модал мисоллар таснифлари жамламаси аниқланган. Хужжат хоссалари, типик таснифлари, характерли қирралари, хусусиятларини умумлаштирувчи механизмга эга алгоритмлар юқори қаррали хатоларни аниқлаш ва тузатиш имконини яратади.

Диссертациянинг “Хужжат элементларида мантикий ва семантик боғланишлардан фойдаланиб ахборот ишончилигини ошириш усул ва механизмлари” номли иккинчи бобида ЭХ элементлари статистик, ностационарлик хоссалари, динамик ва хусусий таснифларини ажратиш ва қўллаш бўйича ахборот ишончилигини ошириш масалаларининг янги ечимлари берилган ва назарий жиҳатдан асосланган. Хужжат элементлари ўзаро боғланишлари, жамлама – эталон ва ўзгарувчиларни мословчи механизмлар бўйича ахборот ишончилигини ошириш модел ва алгоритмлари таклиф қилинган. Ахборотни назорат қилиш алгоритмлари, сегментнинг қўйи ва юқори чегаралари, элементлар эҳтимоли содир этишлиги тақсимот функцияси ўзгарувчилари, ишончли ва ишончсиз қийматларга элементларни кетма – кетликда ажратувчи “бўсаға” ўзгаришлари бўйича мослашади. Ахборот ишончилиги, кирувчи ва эталон хужжат сегментлари элементларининг эквивалентлик коэффициенти бўйича оптималлаштирилган. Бу механизмлар эса, сегмент элементлари сонини қисқартириш амалиётини ҳам бажаради. Кирувчи хужжат сегменти концепти k - сегментга тегишли i - элементнинг модал мисол жамлама– эталони q - сегментга тегишли j - элементга ўзаро эквивалентлиги аниқланган.

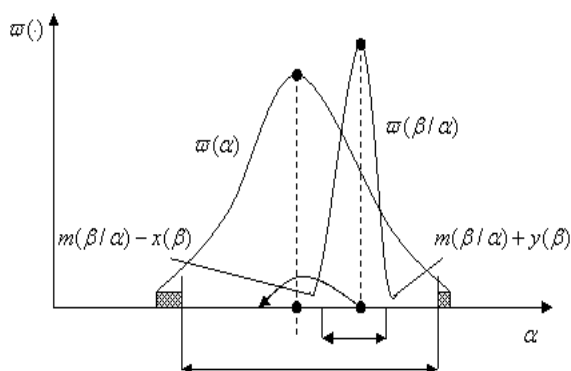
Ахборот ишончилигини оширишда N - иккилик ўлчам, n - разрядли кодлаштиришга асосланган ёндашувлар таклиф қилинган. Бунда бир, икки, уч ва q - каррали хатоларни аниқлаш учун код оптимал метрикаси, тўпلام D_1 , $d_p \in D_1$ аниқланади ва алгоритмлар параметрлари d_p “бўсаға” қийматига мослаштирилади. Аниқланган “бўсаға” хужжат элементларини ишончли ва ишончсиз қийматларга ажратишда ҳам қўлланилади. Масала ечимда, қўйи – $x(\alpha)$ ва юқори – $y(\alpha)$ чегараларда $\{\alpha_p^{1,k}\}$ – рухсат берилган тўпلام қисми ва $\{\alpha_3^{1,k}\}$ – рухсат берилмаган тўпلام қисмига элементлар тегишлилиги эҳтимоллари аниқланади ва ўз навбатида чегаралар оптималлаштирилади.

Агар элемент α тўпلام қисми $\{\alpha_p^{1,k}\}$ га тегишли бўлса, бу ҳолда ахборот ишончли ҳисобланади. Агар элемент α тўпلام қисми $\{\alpha_3^{1,k}\}$ га тегишли бўлса, бу ҳолда ахборот ишончсиз ҳисобланади. Алгоритм икки турдаги хатолар аниқланмаслигига йўл қўяди.

Қабул қилинган тақсимот функцияси $N=5$ тартиб, z аниқликдаги аппроксимация ифодасига эга

$$F(z) = \Phi(z) - \frac{1}{\sqrt{m}} \left(\frac{a_3}{3!} \right) \Phi^{(3)}(z) + \frac{1}{m} \left[\frac{1}{4!} \varepsilon_4 \Phi^{(4)}(z) + \frac{10}{6!} a_3^2 \Phi^{(6)}(z) \right],$$

бунда $\Phi(z)$ - нормал қонуниятдаги эҳтимоллар тақсимот функцияси; μ_1 – “ноль” қийматга эга ўрта катталиқ; μ_2 – “бир” қийматга эга дисперция; $\Phi^{(3)}(z)$ – функция ҳосиласи, $a_3 = \frac{\mu_1}{\sqrt{\mu_2^3}}$ - асимметрияси; $\varepsilon_4 = \frac{\mu_1}{\mu_2} - 3$ – экцессаси.



1-расм. Ахборот назорати

1- расмда α и β элементлар эҳтимолнинг $w(\alpha)$ тақсимот функцияси чизиқлари тасвирланган элементлар назорати шартли функцияси $w(\beta/\alpha)$ бўйича амалга оширилади. Хатолар импликацияси матрицаларига эга жамғарма бўйича ахборот ишончилигини назорат қилиш механизмлари ишлаб чиқилган. Ахборотни назорат қилиш принципи хужжатга хос бўлган хоссалардан

чекланган ўзгаришларни мантиқий ва семантик боғланишлар ҳамда хужжат концепти муносабатлари маънолари бўйича аниқлайди.

Тадқиқотда, бир, икки ва уч қаррали хатоларнинг энг кўп эҳтимолга эгаллиги аниқланган. Хатоларни аниқламаслик эҳтимоли, агар $10^{-4} - 10^{-5}$ қийматларда бўлса, бу ҳолда 10^8 ҳажмда ўнли белги сонига эга 100 хужжатга ишлов бериш вақт харажати қиймати, ахборот ишончилигини назорат қилувчи анъанавий визуал технология бўйича 2,4 соатга тенг бўлади. Ишлаб чиқилган алгоритм эса ахборотга ишлов бериш мураккаблигини 2,4 соатдан 1,1 соатгача қисқатириш имконини яратади. Ахборот ишончилигини таъминлаш бўйича ютуқ коэффициенти қиймати хатоларни таҳрирловчи кодлар технологиясига асосланганда 3-5 мартоба ошади. Хатоларни аниқлаш коэффициенти 88% дан 98,7 % гача ошади, ахборотга ишлов бериш харажат нархи қиймати коэффициенти 7- 8 мартобагача камаяди. Ахборот ишончилигини ошириш алгоритмида, агар барча вариантларни танлаш механизми бўйича қидириш алгоритмини умумлаштирилганда хужжат релевантлиги ўртача 40% га, куйдириш механизмига асосланган қидирувда ўртача 55% га, статистик моделлаштириш механизмига эга қидирувда 70 % га ошиш имконияти яратилиши аниқланган. Хатоларни аниқлаш ва тузатиш ўртача вақти 4,39 мин дан 3,3 мин гача; бир дона хужжатга ишлов бериш вақти эса 6 мин дан 1,5 мин гача камаяди, фойдаланувчининг тизим фаолиятдан қониқиши 98% гача ошади.

Диссертациянинг “**Хужжатлар ахборотлари ишончилигини оптималлаштириш усуллари самарадорлиги баҳоси**” номли учинчи бобида алгоритмлар бўйича ахборотга ишлов бериш самарадорлиги кўпмезонли – ишончилилик, мураккаблик, харажат нарх-қиймати мезонлари бўйича таҳлил услубияти, ҳамда хужжат элементлари статистик (алгоритм1), мантиқий (алгоритм2), семантик (алгоритм3), таркибий-технологик (алгоритм4) боғланишларидан фойдаланувчи механизмларини қўллашга йўналтирилган ёндашувлар таклиф этилган. Ишда яратилган қидирув, таниш, синфлаштириш, ишончликни бинар метрика бўйича назорат қилиш механизмлари имкониятлари умумлаштирилган, ҳамда бажарилган таҳлил натижалари умумлашган алгоритмнинг анъанавий ахборотга пакетли ишлов бериш технологияси самарадорлиги билан қиёсланган. ЭҲ ахборотига ишлов

бериш учун таклиф этилган алгоритмлар ишончликни икки дан уч тартибгача оширади; мураккаблик ва харажат нарх қиймати коэффицентларини 7-8 маротобага камайтириши асосланган. Барча вариантларни тўла танлаш бўйича қидирув, куйдириш механизми бўйича эвристик қидирув, кесик Марков занжири бўйича моделлаштирилган статистик қидирув механизмлари синтези асосида умумлашган алгоритм оптималлаштирилган.

Ахборот ишончилиги баҳосини олиш учун элемент α_{ij} нинг эталон функция $F^*(\beta_{ij})$ β_{ij} элементига хослигини берувчи $F(\alpha_{ij})$ функция киритилган. Ушбу таҳлика функцияси бажарилиши натижалари икки ҳолатдаги эҳтимол функциялари: $\mu(A_1)$ - биринчи турдаги хатолик содир бўлиши, $\mu(A_2)$ - иккинчи турдаги хатолик содир бўлиши, бунда A_1 - элемент α_{ij} ишончилиги бўйича гипотеза ва A_2 - элемента β_{ij} ишончилиги бўйича гипотеза. Таҳлика функцияси $\mathfrak{R} = \mu(A_1) \cdot \omega_{A_1} + \mu(A_2) \cdot \omega_{A_2}$ кўринишда берилган, бунда $\omega_{A_1}, \omega_{A_2}$ - $\mu(A_1)$ ва $\mu(A_2)$ гипотезаларининг миқдор

коэффициентлари. Функция $\mu(A_1) \cdot \omega_{A_1} = \sum_{i=1}^s \varphi_i \cdot w_i$ берилади, бунда w_i - хужжат элементи α_{ij} моҳияти миқдорий қиймати $s = 1, 2, \dots, 16$. Таҳлика

функцияси баҳоси, биринчи турдаги хатолик эҳтимоли $P_1 = \frac{z_1 + z_2}{D} p_{A_1}(\alpha_{ij})$,

D - α_{ij} элементи қиймати умумий интервали; $p_{A_1}(\alpha_{ij})$ - $\alpha_{ij} \rightarrow \beta_{ij}$ хатолар содир бўлиши априор эҳтимоли, ҳамда иккинчи турдаги хатолик эҳтимоли

бўйича $P_2 = (1 - p_{A_1}(\alpha_{ij})) \left[1 - \int_{z_1}^{z_2} \mu(y) dy \right]$ ифода олинган. Биринчи ва иккинчи

турдаги хатоликлар аниқланмаслиги умумий эҳтимолини камайтириш учун $\alpha_{ij} \rightarrow \beta_{ij}$ z_1 ва z_2 - интервалида хослигини текшириш экстремал чегаралари аниқланган. Оптималлаштириш масаласи хатолар аниқламаслик умумий эҳтимоли математик ифодасида хусусий ҳосилаларни олиш, айрим ўгиришларни амалга ошириш ҳамда оптимал бўлган z_1^{onm} ва z_2^{onm} чегараларни аниқлашдан иборат. Гаусс функцияси учун қуйидаги экстремал назорат чегараларининг қийматлари $z_1^{onm} = z_2^{onm} = \pm 2\sigma[1 - \mu(A_1)]$ олинган, бу ерда σ - тақсимот функциясининг ўрта квадратик хатолиги. Кўлланилган механизмлар асосида эришилган хужжатга ишлов бериш аниқлиги қуйидаги мезон бўйича баҳоланади

$$\Omega(a, b, \lambda) = \sqrt{\sum_{i=1}^N (\zeta(A_i) - \bar{\zeta})^2},$$

бу ерда $\zeta(A_i), \bar{\zeta}$ - кирувчи ва эталон хужжат элементларининг A_i нуктада мослиги.

Оптималлаштириш қўйидаги шартларда тадқиқ қилинади

$$D = [\Omega(a, b, \lambda) - \Omega(a^*, b^*, \lambda^*)] \rightarrow \min_{(a, b, \lambda)},$$

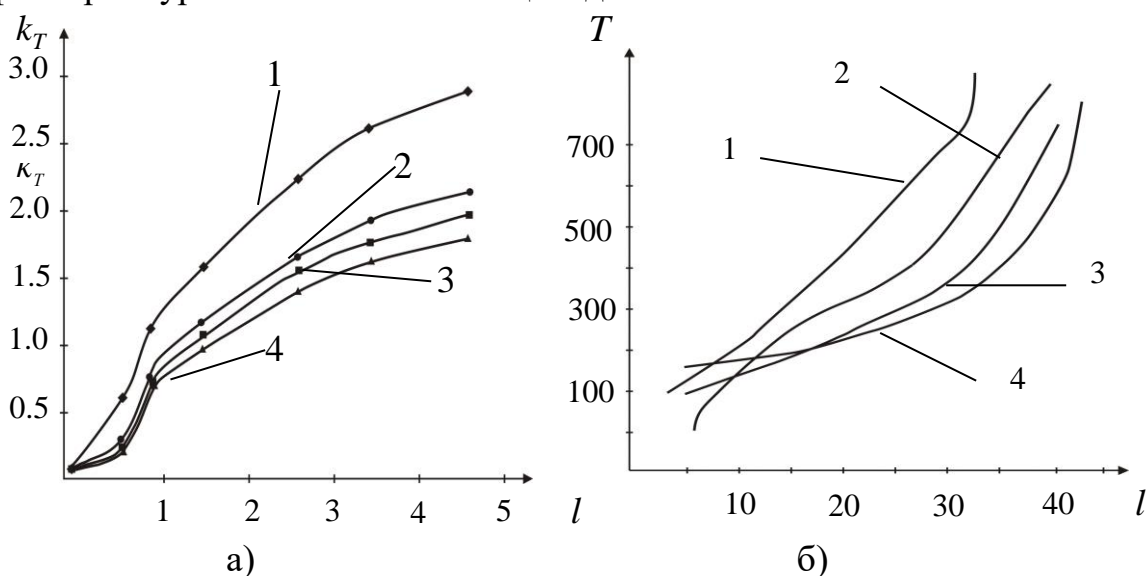
бу ерда $D_a = a_{\min} \leq a \leq a_{\max}$; $D_b = b_{\min} \leq b \leq b_{\max}$, $D_\lambda = \lambda_{\min} \leq \lambda \leq \lambda_{\max}$ - a, b, λ параметрлар қийматлари соҳалари; a - ҳужжат тасвири растрининг баландлиги; b - ҳужжат тасвири растрининг кенглиги; λ - ахборот ишончилигини оширувчи алгоритм самарадорлиги коэффиценти (ҳужжат тасвирларининг ҳар бир узатиш сеанси учун).

Ҳужжатга ишлов бериш аниқлиги коэффиценти қўйидагича

$$k_T = \sum_{i=1}^q \sum_{j=1}^l \lambda_i d_j^2, \quad d_i \in D.$$

2-расм а) ахборотга ишлов бериш аниқлиги k_T коэффиценти 1-4 графикларини ахборот ҳажми l турли қийматларида тузилган, a, b параметрлар қиймати ўнли кодлаш тизимида ўлчанади, ахборот ҳажми 10^6 ўнли белгида тенг қабул қилинган. График 1 - ҳужжат элементларининг статистик; график 2 - мантиқий; график 3 - технологик; график 4 - семантик боғланишларига таянган алгоритмлар самарадорлигини акслантиради.

2 - расм б) ахборотга ишлов бериш мураккаблиги T коэффиценти графиклари кўринишини намоиш қилади.



2-расм. Ахборотга ишлов бериш самарадорлиги

Ахборотга ишлов бериш мураккаблиги, $N=100$ ҳужжат сонидан тўрт ядроли процессорда анъанавий технология бўйича 5.21×10^{-4} сек тенг бўлади. Параллел ҳисобланиш умумий алгоритмга синтезлашганда эса ишлаб чиқилган механизм ахборотга ишлов бериш мураккаблиги коэффицентини 0.17×10^{-4} сек га тушириш имконига эгаллиги аниқланган. Шу қаторда, технологик ортиқчалигидан фойдаланишга асосланган алгоритм ишончилилик ва релевантлик коэффицентлари қийматларини сезиларли даражада оширади.

Диссертациянинг “Электрон ҳужжат алмашуви тизимларида ахборот ишончилигини оширувчи дастурий мажмуа” номли тўртинчи бобида ишлаб чиқилган ҳамда турли механизмлар асосида ҳужжат маълумотларини назорат қилиш ва қайта ишловчи функционал модулларни қамраган ахборот ишончилигини оширувчи дастурий мажмуа(ДМ) жорий этилган. ЭХАТ куйидаги функционал модулларни: мословчи воситалар – интегратори, умумлаштирувчи, маълумотлар импорт ва экспорти; матн семантик анализатори; маълумотларни юклаш интерфейси, калитли сўзлар ва луғатлар библиотекаси, маълумотларни ўгирувчи – конвертор; семантик гипер тармоқ бўйича қидирувчи, дастурлар библиотекаси синтезлаш схемаси жорийлаштирилган.

Жорийлаштирилган ДМ устунлиги куйидагилардан иборат: «CUDA» параллел ҳисобланишнинг Си тилида яратилган кенгайтирилган иловаси; мултипроцессорда 16 Кб ўлчамида оқимлар орасида тақсимланадиган хотирага эришиш имконини яратиш; тизим ва видео хотира орасида маълумот узатилишини таъминлаш; gather, scatter хотираларини чизиқли адреслаш, ҳамда эркин адрес бўйича олинган ёзувни таъминлаш.

МБда жойлаштирилган жадваллар гуруҳи, SQL форматда конвертлашувчи, ББ муҳитида ДМ ни синтезлаш схемаси таклиф этилган.

Жорийлаштирилган ахборотга ишлов бериш модуллари куйидаги функцияларини бажаради: маълумотларни *.xls форматга ўгириш ва юклаш; ҳужжат элементлари боғланишларини қидириш, мослаштириш, маълумотлар майдонни белгилаш ва мувофиқлаш; сўз, луғатларни, дастур библиотекаларини шакллантириш; тематик йўналишларни қидириш, танлаш, улар таркибини мувофиқлаш; ЭХАТ менюсини, ЭХ га ишлов бериш жадвалини яратиш; идентификация қилиш, ЭХ матнларини генерация, таҳлил ва синтезлаш, ЭХ ахбороти хатоларини назорат ва таҳрир қилиш; ББни шакллантириш вазибаларини амалга оширувчи семантик гипертармоқ ишлаб чиқилган. ЭХАТда ахборот ишончилигини анъанавий технология бўйича оширувчи ДМ бир сеанс даврида керакли ҳужжатга берилган 2 сўровни 6 мин да бажарса, жорийлаштирилган технология эса шу операцияларни бажариши учун умумий вақтни 3-4 мин қисқартириши асосланган. Бундан ташқари, 100 ҳужжатдан иборат коллекцияга ишлов бериш вақти 40% гача, сўровга жавоб тайёрлаш вақти 53% га, ахборот таркибида мавжуд бўлган хатоларни тузатиш учун қайта сўров вақти 50% гача қисқариши кузатилди.

Тўла матнли ҳужжат ишончилигини оширувчи алгоритмларнинг кўп босқичли морфологик ва таркиблаштирилган n – грамм таҳлил қилувчи моделлари механизмлари имкониятларини умумлаштирувчи технология таклиф этилган ва ўзбек тили матнлари орфографиясини назорат қилиш «Сфинкс-4» АДП муҳитидаги технология сифатида жорий этилган. «Сфинкс-4» АДПга кирувчи воситалар куйдириш, чеклашлар, стахостик моделлаштириш, n – грамм таҳлил асосида яратилган эвристик қидириш алгоритмлари матнларни таҳлил, нутқга ўгириш ва генерация қилиш ҳамда

нутқни матнга ўгириш ва турли мақсадга чамаланган ахборотни қайта ишлашда қўшимча имкониятларни яратади.

Жорий қилинган ахборот ишончилигини оширувчи механизмлар бир ЭХ ахбороти назоратига кетадиган умумий вақтни 9-10 % гача, ахборотга ишлов бериш мураккаблик ва нарх қиймати коэффициентлари қийматларини 8 маротобага қисқартириши аниқланган.

Тадқиқотлар қуйидаги мезонлар: P_H – хатолар аниқланмаслиги эҳтимоли; P_H/P – ахборот ишончилиги бўйича ютуқ коэффициенти; T_M , C_M – ахборотга ишлов бериш мураккаблиги ва нарх қиймати коэффициентлари бўйича бажарилган.

I – жадвалда концептлар муносабати ва элементлар боғланишига асосланган ЭХ ахбороти ишончилигини оширувчи турли механизмлар дастурий модуллари бўйича оптималлаштириш натижалари келтирилган.

Жадвал 1.

Хужжат элементлари информация ортиқчалигидан фойдаланиш бўйича ахборот ишончилигини ошириш механизмлари самарадорлиги

№	Механизмлар	Самарадорлик мезонлари қиймати			
		P_H	P_H/P	T_M	C_M
1	Статистик боғланиш	$12 \cdot 10^{-5}$	$0,35 \cdot 10^{-1}$	6-7	3-4
2	Мантиқий боғланиш	$2,34 \cdot 10^{-5}$	$0,62 \cdot 10^{-1}$	8-9	3-4
3	Корреляция боғланиш	10^{-5}	$0,3 \cdot 10^{-1}$	10-12	5-7
4	Семантик боғланиш	$1,5 \cdot 10^{-6}$	$0,44 \cdot 10^{-2}$	10-13	8-9
5	Технологик боғланиш	$1,3 \cdot 10^{-6}$	$0,3 \cdot 10^{-2}$	13-15	10-12

Хато пайдо бўлиш эҳтимоли ўртача қиймати $10^{-4} - 10^{-5}$ тенг бўлганда, 100 хужжатдаги ахборот ишончилигини оширишга кетадиган умумий вақт 1,1 соат қийматигача қисқариши, ахборот ишончилиги коэффициенти қиймати 3-5 маротобага оширилиши, ахборотга ишлов бериш мураккаблиги ва нарх қиймати коэффициентлари қиймати 7-8 маротобага камайиши аниқланган.

Ишлаб чиқилган ДМ самарадорлигини ҳисоблаш услубияти Самарқанд вилояти турли корхоналари фаолиятида жорий қилинган. ДМ дастурий таъминоти аниқ предмет соҳаларда қўлланишга мослаштирилган. Дастурий воситалар, модуллар ва дастурий мажмуа имкониятлари тавсифи Ўзбекистон Республикаси интеллектуал мулк Агентлиги томонидан берилган гувоҳномалар аннотацияларида келтирилган. Самарқанд вилояти соғлиқни сақлаш бошқармаси муассаларида, ДМни қўллаш самарадорлиги бир йиллик хужжатларга ишлов бериш маълумотлари ҳажми ҳисобида 612,6 млн. с бўлишлиги аниқланган.

“Электрон хужжат элементлари мантикий ва семантик боғланишини ажратадиган механизмлар бўйича ахборот ишончилигини ошириш” мавзуси бўйича диссертация тадқиқотлари натижалари қуйидагилардан иборат

АСОСИЙ ХУЛОСА

1. ЭҲАТ хужжат концептлари орасидаги муносабат, ҳамда элементларининг статистик, корреляция, мантикий-семантик, таркибий - технологик боғланишлари бўйича ифодаланувчи сунъий, статистик, мантикий-семантик, таркибий-технологик ахборот ортиқчалигидан фойдаланувчи механизмлар асосида ахборот ишончилигини оширишга йўналтирилган конструктив ёндашув, принцип, модел ва алгоритмлар ишлаб чиқилган.

2. Ахборот ишончилигини ошириш масалалари ечимига таклиф қилинган моделлар назарий жиҳатдан асосланган ва параметрлари статистик, динамик, хусусий таснифларни ажратувчи механизмларни қўллаш асосида оптималлаштирилган, ҳамда ахборот ишончилигини назорат қилиш чегараларини мақбуллаштиришга чамаланган алгоритмлар ишлаб чиқилган.

3. Қидириш, сегментация, таниш, синфлаштириш, бинар метрикада мақбуллаштириш механизмларига эга умумлашган алгоритмлар маълумотга пакетли ишлов беришга нисбатан, ахборот ишончилигини 2 дан 3 тартибгача оширади, ҳамда ахборотга ишлов бериш мураккаблик ва харажат нарх қиймати коэффицентларини 7-8 маротобага камайтиради.

4. Барча вариантни кетма-кет танлаш бўйича қидириш механизмига эга алгоритмлар хужжат релевантчилигини ўртача 40 % га, “қуйдириш” га асосланган қидириш 55% га, стохастик моделлаштириш бўйича қидириш 70% гача оширади. Хатоларни аниқлаш ва тузатишга кетадиган ўртача вақт 4.39 мин дан 3.3 мин гача, бир хужжатга кетадиган вақт 6 мин дан 1,5 мин гача қисқаради, фойдаланувчи қониқиши 98% гача оширилиши аниқланган.

5. Хужжатлар коллекцияси МБ си ва ахборотни назорат қилиш қоидалари ББ сини “ клиент – сервер ” принципида фойдаланувчи МББТ MS SQL муҳитида яратиш таклиф этилган. С# интеграллашган муҳитида хужжат қидирувига йўналтирилган, ҳамда редактор, отладчик, дастур тили, жойлаштирилган график амаллар, Microsoft стандарти модуллари дастур мажмуасини қўллаш синовдан ўтказилган. ЭҲ ахбороти ишончилигини ошириш учун С++ платформаси ва СГС муҳитида тақсимотли қидириш амалларини бажариш имкониятларига эга ҳисобланиш схемаси MySQL асосида қўлланилади.

6. Ахборот ишончилигини оширувчи ДМни корхоналарда йил мобайнида тўпланган коллекция маълумотлари ҳажми бўйича ҳисобланган иқтисодий самарадорлик 612,6 млн. с ни ташкил қилади. ДМ дастурий модуллари “Ўзбектелеком”, Самарқанд соғлиқни сақлаш бошқармаси корхона ва муассасаларига жорий этилган.

**НАУЧНЫЙ СОВЕТ DSc.13/30.12.2019.Т.07.01 ПО ПРИСУЖДЕНИЮ
УЧЕНЫХ СТЕПЕНЕЙ ПРИ ТАШКЕНТСКОМ УНИВЕРСИТЕТЕ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

САМАРКАНДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

КАРШИЕВ ХУСАН БЕРКИНБАЕВИЧ

**«ПОВЫШЕНИЕ ДОСТОВЕРНОСТИ ИНФОРМАЦИИ С
ИСПОЛЬЗОВАНИЕМ МЕХАНИЗМОВ ИЗВЛЕЧЕНИЯ ЛОГИЧЕСКИХ
И СЕМАНТИЧЕСКИХ СВЯЗЕЙ ЭЛЕМЕНТОВ ЭЛЕКТРОННЫХ
ДОКУМЕНТОВ»**

05.01.02 - Системный анализ, управление и обработка информации

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

Ташкент – 2021

Тема диссертации доктора философии (PhD) зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за B2019.3.PhD/T1294

Диссертация выполнена в Самаркандском государственном университете.

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.ziynet.uz).

Научный руководитель: **Жуманов Исраил Ибрагимович**
доктор технических наук, профессор

Официальные оппоненты: **Кабулов Анвар Васильевич**
доктор технических наук, профессор
Сулюкова Лариса Фаритовна
доктор технических наук

Ведущая организация: **Ташкентский государственный технический университет**

Защита диссертации состоится «29» октября 2021 г. в 16:00 часов на заседании Научного совета DSc.13/30.12.2019. T.07.01 при Ташкентском университете информационных технологий. (Адрес: 100202, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-43; факс: (99871) 238-65-52; e-mail: tuit@tuit.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского университета информационных технологий (регистрационный номер № 225). (Адрес: 100202, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-65-44).

Автореферат диссертации разослан «18» октября 2021 года.
(протокол рассылки № 31 от «11» октября 2021 г.).



Р.Х.Хамдамов

Председатель научного совета по присуждению ученых степеней, доктор технических наук, профессор

Ф.М.Нуралиев

Ученый секретарь научного совета по присуждению ученых степеней, доктор технических наук, доцент

А.В.Кабулов

Председатель научного семинара при научном совете по присуждению ученых степеней, доктор технических наук, профессор

ВВЕДЕНИЕ

Актуальность и востребованность темы диссертации. В мире особое внимание уделяется совершенствованию технологий повышения эффективности производства, мощности технологических агрегатов, производственных комплексов и оптимальному управлению предприятиями. Одной из наиболее актуальных задач, в развивающихся отраслях является создание и применение цифровых технологий управления различными экономическими объектами, позволяющих улучшить и увеличить выпуск высококачественных продукции с наименьшими энерго и другими ресурсными затратами. В качестве, важных вопросов отмечаются исследования и разработки систем автоматизации, электронизации среди них, большое внимание уделяется ускоренному введению и развитию автоматизированных систем управления в различные отрасли хозяйства, современных математических методов, информационных технологий, которые направлены в первую очередь на внедрении инновационных идей и технологий, в том числе, систем электронного документооборота (СЭД).

В мире известны такие разработки зарубежных фирм как, Microsoft, Lucent, Lernout & Hauspie, Unisys, которые позволяют вести технологии NLU - полноценный диалог с компьютером, средства чтения электронной почты, редакторы контроля орфографии естественных языков, инструментов машинного перевода, синтезаторов и анализаторов речи и текстов, обработки организационно – распорядительных документов (ОРД), а также баз данных (БД) и баз знаний (БЗ). Ключевые звенья современных СЭД используют наиболее популярные системы: Google, Bing, Yahoo, в русскоязычных системах Yandex, Rambler, а в корпоративных сетях Sphinx, Lucene, Xapian. Интенсивно проводится исследование в США, России, Англии, Франции, Германии, Испании, Китая, Индии, Южной Кореи, Японии, Украины, Беларуси, Латвия и в других.

В рамках, принятого в Республике Узбекистан комплекса мер по развитию систем управления и созданию единого информационного пространства особое внимание уделено разработке методов, моделей, систем, базирующихся на современных информационно-коммуникационных технологиях. В Стратегии действий по дальнейшему развитию на 2017-2021 годы определены задачи, направленные на опережающее развитие «...высокотехнологических обрабатывающих отраслей, прежде всего по производству готовой продукции с высокой добавленной стоимостью на базе глубокой переработки местных сырьевых ресурсов...»², в которой большое внимание уделено разработкам новых подходов, методов, предпринятую мер по повышению достоверности информации, полноты, точности, релевантности электронных документов (ЭД). Снижение достоверности информации наиболее вероятно происходит при подготовке и вводе,

² Указ Президента Республики Узбекистан «О Стратегии действий по дальнейшему развитию Республики Узбекистан» ПФ-4947 от 7 февраля 2017 года.

передачи, хранение и обработке документов, обмене между пользователями, содержания которых обусловлены ошибками человека – оператора, погрешностями устройства сканирования и распознавания, воздействием помех в каналы связи, неудовлетворительным обслуживанием системы. Научные и методические основы проблемы повышения достоверности информации должны выполняться в условиях ограниченности априорных сведений, нестационарности, неопределенности, а инструментальные средства обладать большим удобством и скоростью обработки информации при обнаружении и коррекции ошибок ЭД.

Данное диссертационное исследование направлено в определенной степени на выполнение задач, предусмотренных Указом Президента Республики Узбекистан от 7 февраля 2017 года №УП-4947 «О стратегии действий по дальнейшему развитию Республики Узбекистан», от 29 августа 2017 года №ПП-3245 «О мерах по дальнейшему совершенствованию системы управления проектами в сфере информационно-коммуникационных технологий», от 19 февраля 2018 года №УП-5349 «О мерах по дальнейшему совершенствованию информационно-коммуникационных отраслей» и другими нормативно-правовыми документами, принятыми в данной сфере.

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий Республики Узбекистан IV. «Развитие информатизации и информационно-коммуникационных технологий».

Степень изученности проблемы. По результатам анализа литературы за последнее 10-15 лет можно отметить, что большой вклад в теорию и практику достоверной передачи информации внесли зарубежные ученые Ch.W.Bachman, W.W.Peterson, D.Ritchi, E.F.Codd, C.J.Date, Alan Turing, E.N.Gilbert, R.C.Bose, Danish Jamil, Peter Pin-Shan Chen, Чжо Зо Е, В.Б.Кудрявцев, Д.А.Полянский, Э.Э.Гасанов, А.И.Репин, А.Н.Савельев и др.

В теории оптимизации обработки данных значительный вклад внесли ученые Узбекистана Д.А.Абдуллаев, Ф.Б.Абуталиев, Т.Д.Раджабов, Т.Ф.Бекмуратов, М.М.Камилов, Н.Р.Юсупбеков, Х.З.Игамбердиев, И.И.Жуманов, Р.Х.Хамдамов, М.А.Рахматуллаев, Ш.Х.Фазилов, А.А.Саидов, А.Р. Ахатов и др.

Методологию повышения достоверности информации составляют развитые методы теории кодирования и аппаратного использования искусственной информационной избыточности. Исследования в данном направлении отражены в большом количестве монографических работ, диссертаций и научных статей. В современных информационно-коммуникационных сетях используется множество протоколов передачи данных (ППД), которые опираются на использование искусственной информационной избыточности. В них значительный объем ресурса расходуется на передачу заголовков ППД, большая часть которых остается неизменной от пакета к пакету на протяжении сеансов передачи сообщения. Основные источники ошибок, обусловленных человеком – оператором,

орфографией текстов, погрешностью сканирующих и распознающих устройств, образующих наибольшую часть общего объема искаженной информации на Прикладном и Представительском уровнях модели OSI не учитываются. Следовательно, повышения достоверности информации требует развитие методов на основе разработки моделей, алгоритмов и механизмов с использованием статистической, семантической, технологической информационной избыточности, обусловленной статистическими, корреляционными, логико – семантическими связями элементов и отношениями концептов документов, которые, к сожалению, недостаточно изучены и в существующих подходах рассмотрены.

Связь темы диссертации с планами научно-исследовательских работ высшего образовательного и научно-исследовательского учреждения, где выполнена диссертация. Исследование выполнено в рамках проектов плана НИР Самаркандского государственного университета и Самаркандского филиала ТУИТ, в частности по проектам А5-039 «Интеллектуальный анализ данных на основе построения программных инструментов параллельных алгоритмов» (2015-2017).

Цель исследования. Разработка моделей, алгоритмов и программного комплекса повышения достоверности информации на основе использования информационной избыточности различной природы.

Задачи исследования:

теоретико-информационный анализ современных систем управления, методов анализа и обработки ЭД, разработка научно - методических основ повышения достоверности информации;

разработка вероятно – статистических моделей и алгоритмов повышения достоверности информации с механизмами проверки соответствия элементов документа в разрешенное подмножество, кластеру и набор - эталону;

разработка методики выявления и оценки логических, семантических связей элементов и отношений концептов ЭД, алгоритмов повышения достоверности информации с механизмами использования фрактальных, фреймовых характеристик, текстурных особенностей документов;

разработка методики анализа эффективности алгоритмов контроля, по критериям повышения достоверности, трудоемкости и стоимости обработки информации;

разработка программного комплекса и его функциональных модулей, исследование эффективности алгоритмов, использующие статистические, логические, семантические, технологические связи элементов документа.

Объектом исследования являются информационные процессы, документы делопроизводства, СЭД, автоматизированные системы и сети.

Предмет исследования методы, модели, алгоритмы и программные комплексы повышения достоверности информации с использованием информационной избыточности различной природы.

Методы исследования. Теория системного анализа, передачи информации, информационного поиска, распознавания, кластеризации, компьютерной лингвистики, алгоритмизации и программирования.

Научная новизна исследования:

разработаны научно – методические основы построения механизмов повышения достоверности информации, которые основаны на использовании взаимосвязанности элементов и отношений концептов документов, обусловленных статистическими, логическими, семантическими, структурно-технологическими связями;

разработана методика оптимизации порога разделения последовательности по вероятностным критериям обнаружения и исправления ошибок на достоверные и недостоверные элементы, сегменты и регулирования коэффициентов эквивалентности;

получены общее и частные решения задач оценки условных вероятностей n -грамм, а также механизмы контроля и коррекции орфографии текстов на основе синтеза многоуровневого морфологического и n -грамм структурированного анализом с технологией Фреймворк Сфинкс-4;

оптимизировано определение долевого вклада каждого алгоритма повышения достоверности информации в общую эффективность системы по целевой функции соответствия путем определения границ проверки.

Практические результаты исследования:

разработаны алгоритмы, оригинальные функциональные модули, выполняющие функции идентификации, контроля, регулирования переменных, извлечения полезных свойств информации из документов, для использования в СЭД учреждений и организаций в виде программного комплекса (ПК) повышения достоверности информации;

эффективность ПК продемонстрирована в реальных условиях, результаты характеризуются повышением достоверности информации от двух до трех порядков, снижением коэффициентов трудоемкости и стоимости контроля информации в 7-8 раз.

Достоверность результатов исследования. Доказано корректным применением разработанного программного комплекса и его модулей, согласованностью полученных теоретических и практических результатов при решении задач повышения достоверности информации в реальных условиях, а также сравнительным анализом эффективности реализованных алгоритмов по принятым критериям.

Научная и практическая значимость результатов исследования. Теоретическая значимость работы заключается в разработке конструктивных подходов, принципов, моделей, которые обосновываются реализациями механизмов повышения достоверности информации на основе использования избыточности, обусловленной статистическими, корреляционными, логическими, семантическими, структурно-технологическими связями элементов и отношениями концептов документов.

Практическая значимость работы доказывается разработкой программного комплекса повышения достоверности информации,

основанный на синтез алгоритмов повышения достоверности информации, типичных инструментов поиска, распознавания, кластеризации и механизмов применения в СЭД предприятий и организаций.

Разработаны оригинальные функциональные модули повышения достоверности информации с механизмами использования статистических, логических, семантических, технологических связей элементов ЭД, правил парсингового кодирования, распознавания текстовой информации в среде ППП «Сфинкс-4» и параллельных вычислений «CUDA».

Внедрение результатов исследования. По теоретическим результатам диссертационного исследования разработаны модели, алгоритмы, программный комплекс(ПК), функциональные модули и программные инструменты повышения достоверности и обработки информации, которые включены в систему электронного документооборота. Разработанный ПК внедрен в деятельности отделов и канцелярии Самаркандского областного управления здравоохранения для выявления и исправления ошибок, поиска объекта, распознавания, классификации, идентификации, контроля достоверности информации на основе механизмов использования естественной информационной избыточности (Справка Министерства по развитию информационных технологий и коммуникаций №33-8/7849 от 23 декабря 2020 г.). В результате использования ПК повышения достоверности информации, а также других программных модулей в среднем материальные затраты за один месяц сокращаются на 20% ; ПК повышения достоверности информации, а также программные модули, синтезированные в ПК внедрены также в деятельности Самаркандского филиала АО «УЗБЕКТЕЛЕКОМ» (Справка Министерства по развитию информационных технологий и коммуникаций 33-8/7849 от 23 декабря 2020 г.). ПК при применении способствует улучшению следующих технико-экономических показателей системы: общее время, затрачиваемое на контроль достоверности информации в документе, чем по традиционной визуальной технологии сокращается до 10%; среднее время обнаружения и исправления ошибок уменьшается от 4,39 мин до 3,3 мин; среднее время обработки документа уменьшается от 6 мин. до 1 мин.; достоверность информации повышается на три порядка; коэффициенты трудоемкости и стоимости обработки информации снижаются 7-8 раз.

Апробация результатов исследования. Результаты данного исследования обсуждены на 3 международных и 4 республиканских научно-практических конференциях.

Публикация результатов исследования. По теме диссертации опубликовано 26 научных статей, из них 2 в изданиях Scopus, 8 в республиканских журналах, получено 3 свидетельства о регистрации программ для ЭВМ.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав, заключения, списка использованной литературы и приложений. Объем диссертации состоит из 119 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обосновываются актуальность и востребованность темы диссертации, формулируются цель, задачи, объект, предмет исследования, научная новизна, практические результаты исследования, обоснована достоверность полученных результатов, раскрыта теоретическая и практическая значимость результатов, приведены перечень внедрений, сведения об опубликованных работах и структуре диссертации.

В первой главе диссертации **«Научно - методические основы повышения достоверности информации на основе использования статистических связей элементов документа»** проведен теоретический анализ существующей методологии повышения достоверности информации на основе концептуальных принципов использования информационной избыточности различной природы.

Выделены основные подходы, модели определения избыточности, обусловленной статистическими, корреляционными, логическими, семантическими и структурно – технологическими связями элементов и отношениями концептов документа.

Разработаны конструктивные подходы, принципы, методы повышения достоверности информации в СЭД, основанных на применении механизмов извлечения статистических, логика – семантических, структурно – технологических связей, которые в совокупности составляют научно-методические основы оптимизации обработки информации.

Доказано, что механизмы повышения достоверности данных за счет использования избыточности информации представляют эффективные, простые инструменты, устраняют недостатки в существующих методах, а также обеспечивают высокую достоверность со существенно уменьшенной трудоемкостью и стоимостью контроля информации.

Разработана онтология группы методов, моделей, алгоритмов повышения достоверности информации и практических приложений.

Первая группа методов, направлены на повышении достоверности информации с применением механизмов извлечения статистических, корреляционных, логических, семантических, структурно – технологических связей элементов и отношений концептов документа. Вторая группа методов, направлены на повышении достоверности информации полнотекстовых документов на основе механизмов извлечения специфических знаний, текстурных особенностей, ключевых слов, шаблон – эталонов, фрактальных, статистических характеристик.

Исследованы методы контроля орфографии с применением графематических, лингвистических, морфологических моделей анализа текстов на естественных языках. Исследован вопрос использования n – грамм грамматики текстов и оценки условных энтропий, которые обусловлены неравномерностью распределения вероятностей, статистическими и корреляционными связями. Определена информационная избыточность по энтропиям монограмм, диграмм, триграмм и k – грамм.

Условные энтропии k-грамм текстов находятся, как $\tilde{H}(x_{ij}^{(1)}, x_{ij}^{(k)}) = - \sum_{\xi_{ij}^{(1)}, \xi_{ij}^{(k)}} \tilde{P}(\xi_{ij}^{(1)}, \xi_{ij}^{(k)}) \log \tilde{P}(\xi_{ij}^{(1)}, \xi_{ij}^{(k)})$, $\xi_{ij}^{(1)}, \xi_{ij}^{(k)} \in \overline{x_{ij}^{(1)}, x_{ij}^{(r)}}$ реализации k-грамм; $\tilde{P}(\xi_{ij}^{(1)}, \xi_{ij}^{(k)})$ – вероятности k – грамм.

Избыточность требуемая с учетом средней вероятности ошибок, определяется в виде $R(S_i) = 1 - \frac{1}{(1-P)} \log \frac{(2^m - 1)}{nP}$.

Доказано, что достоверность информации повышается до требуемого уровня при следующих значениях информационной избыточности: 0,5 - статистических связях; 0,3 - логических связях; 0,6 - семантических связях; 0,8 - технологических связях элементов документа. Обосновано, что алгоритмы с механизмами использования статистических, корреляционных, логических, семантических и структурно - технологических связей элементов ЭД отличаются эффективностью и простотой применения инструментов обнаружения и исправления ошибок.

Изучена статистика ошибок, происходящих в основных информационных звеньях. Установлено, что значения средней вероятности ошибок, обусловленной помехами в каналах связи $P_1 = 5 \cdot 10^{-4}$, оператором $P_2 = 1,9 \cdot 10^{-3}$, орфографическими ошибками $P_3 = 10^{-4}$, погрешностью устройства сканирования и распознавания $P_4 \approx P_5 \approx 10^{-2} - 10^{-3}$.

Разработаны алгоритмы повышения достоверности информации на основе использования статистических связей элементов документа. Получены оценки вероятностей необнаруженных ошибок для различных ситуаций контроля информации. Проведен синтез вариантов решения задач, который позволяет провести оценку общей вероятности необнаруженных ошибок в виде $P_H \approx 3,9 \cdot (1 - P_{\alpha_i})^2 \cdot 10^{-5} + 3,2 \cdot P_{\alpha_i} \cdot (1 - P_{\alpha_i}) \cdot 10^{-2}$. Алгоритмы достоверность информации повышают до двух порядков.

Решены задачи оптимизации достоверности информации на основе формирования и использования БД, включающий наборы элементов, концептов, ЭД; информации о статистических параметрах, динамических, специфических характеристиках и законах распределений; БЗ, включающий различные правила контроля, анализа, определения статистических, логических, лингвистических связей элементов и отношений концептов, особенностей документов; правил разбивки общего интервала значений элементов концепта ЭД на сегменты, группы, классы. Предложены правила установления и настройки границ сегментов, определения функции и интервала принадлежности в подмножество разрешенных значений элементов, а также набора характеристик модальных примеров. Алгоритмы с механизмами обобщения свойств, типичных характеристик, характерных черт, особенностей документов позволяют обнаруживать и исправлять ошибок высокой кратности.

Во второй главе диссертации «Методы и механизмы повышения достоверности информации с использованием логических и семантических связей элементов документов» теоритически обоснованы новые решения задач повышения достоверности информации за счет извлечения и использования статистических, нестационарных свойств, динамических и специфических характеристик элементов ЭД. Предложены модели и алгоритмы повышения достоверности информации с механизмами использования взаимосвязанности элементов документа, набора – эталона и настройки переменных. Алгоритмы контроля информации адаптируются на изменения нижней и верхней границы сегмента, переменных функции плотности распределения вероятностей элементов, порога разделения последовательности на достоверные и недостоверные элементы. Оптимизация достоверности информации проводится по коэффициенту эквивалентности сегмента элементов вводимого и эталонного документа. Механизмы, включают процедур сокращения число элементов в сегменте. Определена взаимная эквивалентность i -го элемента, k -го сегмента концепта вводимого документа, j -му элементу, q -го сегмента в набор – эталоне модального примера.

Предложены подходы, направленные на повышении достоверности кодированной n - разрядным кодом информации размера N - двоичных символов. Параметры алгоритмов при обнаружении одно, две, три и q - кратных ошибок регулируются на оптимальное кодовое расстояние метрики, которое равняется пороговой величине d_p из множества D_1 , $d_p \in D_1$. Порог используется также при разграничении достоверной и недостоверной части элементов документа. Фиксируются $\{\alpha_p^{1,k}\}$ – разрешенное подмножество, которое определяется нижней - $x(\alpha)$ и верхней - $y(\alpha)$ границами и $\{\alpha_3^{1,k}\}$ – запрещенное подмножество, определяются вероятности принадлежности элементов в эти границы, которые затем оптимизируются.

Если элемент α принадлежит подмножеству $\{\alpha_p^{1,k}\}$, то информация считается достоверной. Если элемент α принадлежит подмножеству запрещенных значений $\{\alpha_3^{1,k}\}$, то информация считается недостоверной. Алгоритмом оставляется необнаруженными ошибок двух родов.

Рассматривается функция условного распределения вероятностей элемента z с точностью до членов порядка $N = 5$, которая аппроксимируется по выражению

$$F(z) = \Phi(z) - \frac{1}{\sqrt{m}} \left(\frac{a_3}{3!} \right) \Phi^{(3)}(z) + \frac{1}{m} \left[\frac{1}{4!} \varepsilon_4 \Phi^{(4)}(z) + \frac{10}{6!} a_3^2 \Phi^{(6)}(z) \right],$$

где $\Phi(z)$ - функция нормального распределения вероятностей с нулевым средним μ_1 и единичной дисперсией μ_2 ; $\Phi^{(i)}(z)$ - ее производные,

$a_3 = \frac{\mu_1}{\sqrt{\mu_2^3}}$ - асимметрия, $\varepsilon_4 = \frac{\mu_1}{\mu_2} - 3$ - эксцесса функции.

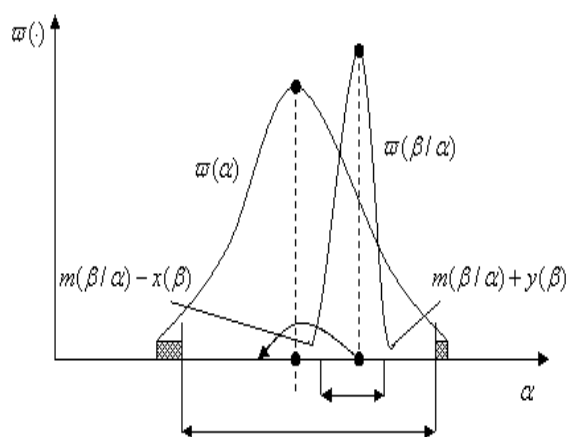


Рис. 1. Контроль информации

На рис.1 проиллюстрированы кривые функции распределения вероятностей $w(\alpha)$ элементов α и β проверочного условного распределения $w(\beta/\alpha)$.

Разработаны механизмы повышения достоверности информации с использованием базы матриц импликаций ошибок. Принцип контроля информации заключается в выявлении нарушений, заложенных в документах, логических и семантических связей

элементов, смысловых отношений концептов.

Определено, что наиболее вероятными являются одно-, дву- и трехкратные ошибки. Когда вероятность необнаруженных ошибок алгоритма находится в пределах $10^{-4} - 10^{-5}$, тогда временные затраты на обработку 100 документов с объемом информации 10^8 десят. знак по традиционной технологии визуального контроля достоверности информации равны 2,4 часа. В тоже время, разработанный алгоритм уменьшает трудоёмкость обработки информации с 2,4 час до 1,1 часа.

Значение коэффициента выигрыша в достоверности информации по сравнению технологии использования корректирующих кодов увеличивается 3-5 раза. Коэффициент обнаружения ошибок повышается с 88% до 98,7%, а значение коэффициента стоимости обработки информации уменьшается 7-8 раза. Определено, что, когда в алгоритмах повышения достоверности информации применяются механизм поиска с полным перебором всевозможных вариантов, тогда релевантность документа повышается в среднем на 40%, поиска с механизмом отжигом в среднем на 55%, поиска со стохастическим моделированием в среднем на 70%. Средняя время на обнаружение и исправления ошибок уменьшается с 4.39 мин до 3.3 мин; на обработку одного документа с 6 мин до 1,5 мин, удовлетворенность пользователя работой системы повышается до 98%.

В третьей главе диссертации **«Оценка эффективности методов оптимизации достоверности информации электронных документов»** предложены подходы, направленные на проведении многокритериального анализа эффективности алгоритмов, использующих статистические (алгоритм 1), логические (алгоритм 2), семантические (алгоритм 3), структурно – технологические связи (алгоритм 4) элементов документа по критериям достоверности, трудоемкости и стоимости обработки информации. Обобщены возможности разработанных механизмов поиска, распознавания, классификации и контроля достоверности по бинарным метрикам, результаты анализа эффективности обобщенного алгоритма получены путем сравнения с эффективностью традиционной технологии

пакетной обработки данных. Рекомендованные алгоритмы достоверность информации ЭД повышают от двух до трех порядков; коэффициенты трудоемкости и стоимости обработки информации снижаются 7-8 раз.

Оптимизирован обобщенный алгоритм, синтезирующий механизмы поиска с полным перебором всех вариантов, эвристического поиска с отжигом, со стохастическим моделированием по усеченной цепи Маркова.

Для оценки достоверности информации вводится функция соответствия $F(\alpha_{ij})$ элемента α_{ij} к элементу β_{ij} эталонной функции $F^*(\beta_{ij})$. Результат выполнения этой функции риска обуславливается вероятностными ситуациями $\mu(A_1)$ - появлением ошибок первого рода, $\mu(A_2)$ - появлением ошибок второго рода, где A_1 - гипотеза о достоверности элемента α_{ij} и A_2 - гипотеза о достоверности элемента β_{ij} .

Функция риска задаётся в виде $\mathfrak{R} = \mu(A_1) \cdot \omega_{A_1} + \mu(A_2) \cdot \omega_{A_2}$, ω_{A_1} , ω_{A_2} - соответственно, весовые коэффициенты $\mu(A_1)$ и $\mu(A_2)$. Обозначено в виде

$\mu(A_1) \cdot \omega_{A_1} = \sum_{i=1}^s \varphi_i \cdot w_i$, w_i - весовое значение важности α_{ij} элемента документа, $s = 1, 2, \dots, 16$.

Получены оценки функции риска по вероятности ошибок первого рода $P_1 = \frac{z_1 + z_2}{D} p_{A_1}(\alpha_{ij})$, D - общий интервал значений элемента α_{ij} документа; $p_{A_1}(\alpha_{ij})$ - априорная вероятность появления ошибок типа $\alpha_{ij} \rightarrow \beta_{ij}$.

Вероятность ошибки второго рода $P_2 = (1 - p_{A_1}(\alpha_{ij})) \left[1 - \int_{z_1}^{z_2} \mu(y) dy \right]$.

Для уменьшения общей вероятности необнаруженных ошибок первого и второго родов определены экстремальные значения границ z_1 и z_2 - интервала проверки соответствия $\alpha_{ij} \rightarrow \beta_{ij}$ в их пределах. Оптимизация заключается в нахождение частных производных, в проведении некоторых преобразований в выражение общей вероятности необнаруженных ошибок, а также в определении оптимальных границ z_1^{opt} и z_2^{opt} .

Для Гауссовой функции получены следующие экстремальные значения границ контроля $z_1^{opt} = z_2^{opt} = \pm 2\sigma[1 - \mu(A_1)]$, где σ - среднеквадратическое отклонение функции распределения. Достигнутая точность при обработке информации за счет использования разработанных механизмов оценивается по критерию

$$\Omega(a, b, \lambda) = \sqrt{\sum_{i=1}^N (\zeta(A_i) - \zeta(\bar{A}_i))^2},$$

где $\zeta(A_i), \zeta(\bar{A}_i)$ - соответственно, значения элемента вводимого и эталонного документов в точке A_i .

Оптимизация проводится при следующих условиях

$$D = [\Omega(a, b, \lambda) - \Omega(a^*, b^*, \lambda^*)] \rightarrow \min_{(a, b, \lambda)}$$

где $D_a = a_{\min} \leq a \leq a_{\max}$; $D_b = b_{\min} \leq b \leq b_{\max}$, $D_\lambda = \lambda_{\min} \leq \lambda \leq \lambda_{\max}$ - области допустимых значений для параметров a, b, λ ; a - высота раstra кадра документа; b - ширина раstra кадра документа; λ - коэффициент эффективности алгоритма повышения достоверности информации, определяемый в одном сеансе передачи кадра.

Коэффициент точности обработки документа задаётся в виде

$$k_T = \sum_{i=1}^q \sum_{j=1}^l \lambda_i d_j^2, \quad d_i \in D.$$

На рис 2 а) проиллюстрированы графики коэффициента точности обработки информации k_T , которые построены в зависимости от размера объема информации l , задаваемого по шкале десятичного кода и принят равным 10^6 десят. знак., при усреднении значений параметров a, b . График 1, отражает результат использования статистических связей; график 2 - логических связей; график 3- структурно - технологических связей; график 4 - семантических связей элементов документа.

На рис 2 б) проиллюстрированы графики коэффициента трудоемкости обработки информации при реализованных механизмах.

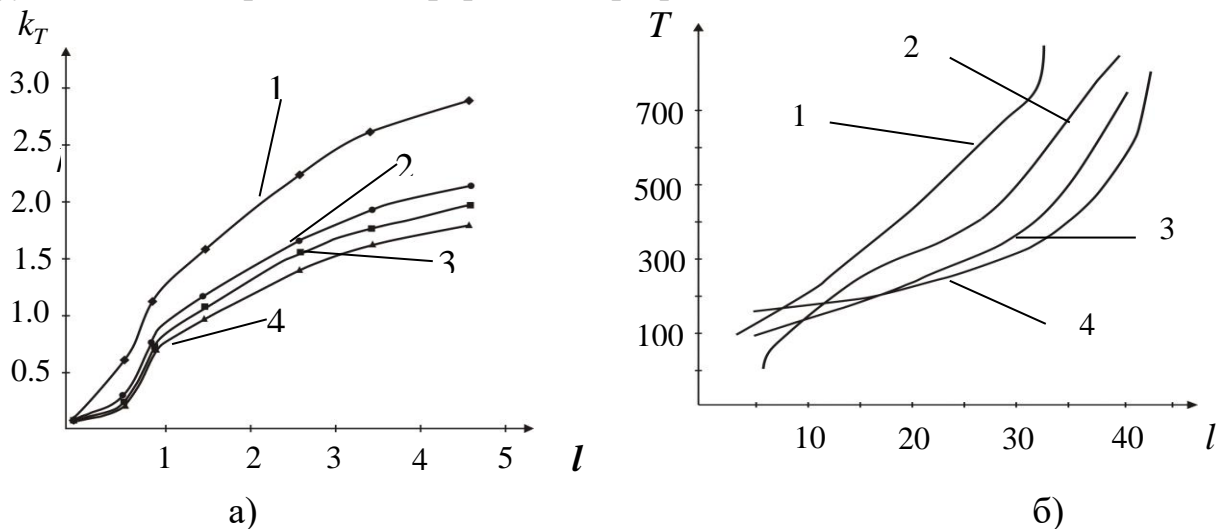


Рис. 2. Эффективность обработки информации.

Определено, что трудоемкость традиционной технологии поэлементной обработки информации при $N=100$ документах в четырех ядерном процессоре равен 5.21×10^{-4} сек.

Трудоемкость обработки документов на основе обобщенного алгоритма с параллельной обработки информации равен 0.17×10^{-4} сек.

Алгоритмы, использующие структурно- технологическую избыточность ЭД при установленных коэффициентах выигрыша в точности и релевантности информации обеспечивают существенное повышение скорости обработки документов.

В четвертой главе диссертации «**Программный комплекс повышения достоверности информации в системах электронного документооборота**» разработан и реализован программный комплекс (ПК) повышения достоверности информации, включающий функциональные модули оптимизации достоверности информации при различных механизмах контроля и обработки документов. Реализованы схемы синтеза следующих функциональных модулей СЭД: интегратор - средства настроек, объединений, импорта, экспорта данных; семантический анализатор текста; интерфейс загрузки данных, библиотек ключевых слов и словарей; конвертор - преобразователь данных; семантический гиперсеть поиска, библиотеки программ. Преимуществами реализованный ПК являются следующие: интерфейс на языке Си с расширением приложений параллельных вычислений «CUDA»; обеспечение доступа к разделяемой памяти между потоками с размером в 16 Кб на мультипроцессор; обеспечение передачи данных между системой и видеопамятью; линейная адресация памяти, gather, scatter, возможность записи по произвольным адресам.

Каждый блок может быть одно-, двух- или трехмерным по форме и может состоять из 512 потоков на текущем аппаратном обеспечении. Предложена схема синтеза ПК в среде встроенных групп таблиц БД, БЗ, которые конвертируются в формате SQL.

Разработана семантическая гиперсеть (СГС) поиска, которая выполняет следующих функций: для загрузки данных из формата *.xls с последующим преобразованием; обработки данных по реализованным модулям; поиска связей между элементами ЭД, настройки связей; фиксирования полей данных, соответствующих модулям СЭД; формировании ключевых слов, словарей, библиотек программ; выбор тематического направления; дополнения групп тем; составление меню СЭД, графика обработки ЭД; проведение идентификации, генерации, анализа и синтеза текстов в ЭД; формировании БЗ; контроль и коррекция ошибок в ЭД и вывод результата.

В одном сеансе СЭД по традиционной технологии повышения достоверности информации за 6 мин выполняет 2 запроса на нужный документ, а реализованный механизм общей времени обработки информации снижает до 3 - 4 мин.

Определено, что временные затраты на обработки коллекции из 100 документов снижаются на 40%, на подготовки ответа на запрос на 53%, выполнения запроса с исправлением искаженной информации на 50%.

Предложен обобщенный механизм, повышения достоверности полнотекстовых документов с использованием многоуровневой морфологической и структурированной n – грамм моделей анализа, которые реализованы в среде ППП «Сфинкс-4» для обнаружения и коррекции орфографии текстов ухбекского языка. Инструменты, входящие в оболочку ППП «Сфинкс-4», в частности эвристические алгоритмы поиска с отжигом, с запретами, на основе стохастического моделирования, n - грамм анализа представляют дополнительных возможностей при анализе, генерации текста в речь, а также речи в текст, обработки информации различных назначений.

Установлено, что реализованный механизм позволяет сократить общее время контроля информации в одном ЭД на 9-10 %; значения коэффициентов трудоёмкости и стоимости обработки информации уменьшаются до 8 раза.

Исследование проведено по критериям P_H – вероятности необнаруженных ошибок, P_H/P – коэффициента выигрыша в достоверности информации; T_M , C_M – коэффициентов трудоёмкости и стоимости обработки информации. В табл.1. приведены результаты оптимизации достоверности информации ЭД по программным модулям, использующих различные механизмы извлечения связей элементов и отношений концептов документа.

Таблица 1.

Эффективность механизмов повышения достоверности информации за счет использования информационной избыточности элементов документа

№	Механизмы	Значения критериев эффективности			
		P_H	P_H/P	T_M	C_M
1	Статистических связей	$12 \cdot 10^{-5}$	$0,35 \cdot 10^{-1}$	6-7	3-4
2	Логических связей	$2,34 \cdot 10^{-5}$	$0,62 \cdot 10^{-1}$	8-9	3-4
3	Корреляционных связей	10^{-5}	$0,3 \cdot 10^{-1}$	10-12	5-7
4	Семантических связей	$1,5 \cdot 10^{-6}$	$0,44 \cdot 10^{-2}$	10-13	8-9
5	Технологических связей	$1,3 \cdot 10^{-6}$	$0,3 \cdot 10^{-2}$	13-15	10-12

Установлено, что при средней вероятности ошибок $10^{-4} - 10^{-5}$ значение временных затрат на повышение достоверности информации коллекции из 100 документов сокращается до 1,1 часа, значение коэффициента выигрыша в достоверности информации увеличивается в 3-5 раза, коэффициенты трудоёмкости и стоимости обработки информации уменьшаются 7-8 раза.

Разработана методика расчета экономической эффективности внедрения ПК в деятельности различных предприятий Самаркандской области. Адаптированное программное обеспечение ПК реализовано в конкретных предметных областях. Описание возможностей программных средств, модулей и программного комплекса изложено в соответствующих аннотациях свидетельств Агенства интеллектуальной собственности РУз.

ЗАКЛЮЧЕНИЕ

Результаты диссертационного исследования по теме: «**Повышение достоверности информации с использованием механизмов извлечения логических и семантических связей элементов электронных документов**» сводятся к следующим основным выводам:

1. Разработаны конструктивные подходы, принципы, модели и алгоритмы повышения достоверности информации, основанных на концепции использования статистической, семантической, технологической информационно избыточности, обусловленных статистическим, корреляционным, логическим, семантическим, технологическим связями элементов и отношениями концептов документа.

2. Теоретически обоснованы предложенные решения задач повышения достоверности информации, оптимизации переменных за счет использования механизмов извлечения статистических, динамических, специфических характеристик ЭД, а также разработаны алгоритмы с оптимизированными границами контроля информации.

3. Разработаны обобщенные алгоритмы с механизмами поиска, сегментации, распознавания, классификации, оптимизации по бинарной метрики, эффективность которых сравнены с эффективностью традиционной визуальной технологии контроля достоверности информации, которые повышают его значения на три порядка, снижают коэффициентов трудоемкости и стоимости обработки информации почти 7-8 раз.

4. Установлено, что обобщенные алгоритмы контроля достоверности информации релевантность документов повышают при следующих механизмах поиска: с полным перебором всех вариантов в среднем на 40%; с отжигом на 55%; со стохастическим моделированием на 70%. Среднее время на обнаружение и исправление ошибок уменьшается с 4.39 мин до 3.3 мин, а на обработки информации документа с 6 мин до 1,5 мин, при этом удовлетворенность пользователя повышается до 98%.

5. Для организации БД документов и БЗ правил контроля информации применена СУБД MS SQL по принципу «Клиент-Сервер». Документный поиск осуществляется программным модулем в среде C#, которая включает редактор, отладчик, язык программирования, встроенные графические процедуры в стандарте Microsoft. Вычислительная схема повышения достоверности информации ЭД реализована в среде СГС на платформе C++ с возможностью распределённого поиска, которая производит встроенную поддержку MySQL.

6. Расчетный экономический эффект внедрения ПК в СЭД предприятия при обработке документов и информации одного года составляет 612,6 млн. сум. Программы ПАК внедрены в деятельности управления здравоохранения Самаркандской области и «Узбектелеком» Самаркандского филиала.

**SCIENTIFIC COUNCIL AWARDING SCIENTIFIC DEGREES
DSc.13/30.12.2019.T.07.01 AT TASHKENT UNIVERSITY OF
INFORMATION TECHNOLOGIES**

SAMARKAND STATE UNIVERSITY

KARSHIEV KHUSAN BERKINBAEVICH

**INCREASING THE RELIABILITY OF INFORMATION USING
MECHANISMS FOR EXTRACTING LOGICAL AND SEMANTIC LINKS
BETWEEN ELEMENTS OF ELECTRONIC DOCUMENTS**

05.01.02 – System analysis, management and information processing

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

Tashkent–2021

The theme of dissertation 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 B2019. 3. PhD/T1294

The dissertation has been prepared at the Samarkand Staty University.

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

Scientific adviser: **Jumanov Isroil Ibragimovich**
Doctor of Technical Sciences, Professor

Official opponents: **Kabulov Anvar Vasilovich**
Doctor of Technical Sciences, Professor

Slyukova Larisa Faritovna
Doctor of Technical Sciences


Leading organization: **Tashkent State Technical University**


The defense will take place "29" October 2021 at 16:00 the meeting of Scientific Council № DSc.13/30.12.2019.T.07.01 at Tashkent University of Information Technologies (Address: 100202, Tashkent city, Amir Temur street, 108. Tel.: (99871) 238-64-43, fax: (99871) 238-65-52, e-mail: tuit@tuit.uz).


The dissertation can be reviewed at the Information Resourse Centre of the Tashkent University of Infornation Technologies (is registered under № 225). (Address: 100202, Tashkent city, Amir Temur street. 108. Tel.: (99871) 238-64-43, Fax: (99871) 238-65-52).

Abstract of the dissertation sent out on "18" October 2021 y.
(mailing report № 31 on "11" October 2021 y.).




R. Kh. Khamdamov
Chairman of the scientific council
awarding scientific degrees,
Doctor of Technical Sciences, Professor


F. M. Nuraliev
Scientific secretary of scientific council
awarding scientific degrees,
Doctor of Technical Sciences, Docent


A. V. Kabulov
Chairman of the academic seminar under the
Scientific council awarding scientific degrees,
Doctor of Technical Sciences, Professor

INTRODUCTION (abstract of PhD dissertation)

The aim of the research work. Development of models, algorithms and a software package for increasing the reliability of information based on the use of information redundancy of various nature.

The object of the research work. Information processes, office documents, EDMS, corporate networks.

The scientific novelty of the research work is as follows:

scientific and methodological foundations for constructing mechanisms for increasing the reliability of information have been developed, which are based on the use of the interconnectedness of elements and relations of document concepts, due to statistical, logical, semantic, structural and technological links;

a method for optimizing the threshold for dividing the sequence according to probabilistic criteria for detecting and correcting errors into reliable and unreliable elements, segments and regulation of equivalence coefficients has been developed;

general and particular solutions to the problems of assessing the conditional probabilities of n-grams, as well as mechanisms for controlling and correcting the spelling of texts based on the synthesis of multi-level morphological and n-grams of structured analyzes with the Sphinx-4 Framework technology were obtained;

the determination of the share contribution of each algorithm for increasing the reliability of information to the overall efficiency of the system is optimized for the target function of compliance by determining the boundaries of verification.

Implementation of the research results. Based on the theoretical results of the dissertation research, models, algorithms, a software package (PC), functional modules and software tools for increasing the reliability and processing of information have been developed, which are included in the electronic document management system. The developed PC has been implemented in the activities of the departments and the office of the Samarkand Regional Health Administration for detecting and correcting errors, searching for an object, recognizing, classifying, identifying, controlling the reliability of information based on the mechanisms of using natural information redundancy (Reference of the Ministry for the Development of Information Technologies and Communications No. 33-8 / 7849 dated December 23, 2020). As a result of using a PC to increase the reliability of information, as well as other software modules, on average, material costs per month are reduced by 20%; PCs for increasing the reliability of information, as well as software modules synthesized in PCs are also introduced in the activities of the Samarkand branch of UZBEKTELECOM JSC (Reference of the Ministry for the Development of Information Technologies and Communications 33-8 / 7849 dated December 23, 2020). When used, the PC contributes to the improvement of the following technical and economic indicators of the system: the total time spent on the control of the accuracy of information in the document, which is reduced by up to 10% according to the traditional visual technology; the average time of error detection and correction decreases from 4.39 minutes to 3.3 minutes; the average document processing time is reduced from 6

minutes. up to 1 min .; the reliability of information increases by three orders of magnitude; the coefficients of labor intensity and the cost of information processing are reduced 7-8 times.

Structure and volume of the dissertation. The structure of the dissertation consists of an introduction, four chapters, a conclusion, a list of used literature and appendix. The volume of the thesis consists of 119 pages.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I бўлим (I часть; I part)

1. Ахатов А.Р., Каршиев Х.Б. Контроль достоверности информации электронных документов на основе логических критериев и семантической избыточности // «Илмий тадқиқотлар ахборотномаси» илмий-назарий, услубий журнал. – Самарқанд: СамДУ. - №3, 2014. – с.55-62.

2. Jumanov I.I., Karshiev Kh. B. Expanding the possibilities of instruments to improve the information reliability of electronic documents of industrial management systems // «Chemical technology. Control and management» (WCIS-2018), October. – Tashkent, Uzbekistan, 2018. – p.146-150. (05.00.00; № 12).

3. Ахатов А.Р., Холмонов С.М., Каршиев Х.Б. Узбекский журнал «Проблемы информатики и энергетики», Изд-во «Фан» АН РУз. - Ташкент, 2018. - №5. - с. 28-31. (05.00.00; № 5).

4. Жуманов И.И., Каршиев Х. Б. Методика оценки и анализа эффективности алгоритмов повышения достоверности информации систем электронного документооборота // «Илмий тадқиқотлар ахборотномаси» илмий-назарий, услубий журнал. – Самарқанд: СамДУ. - №1 (113), 2019. – с. 64-69.

5. Жуманов И.И., Каршиев Х. Б. Анализ эффективности обобщения алгоритмов повышения достоверности информации на основе использования информационной избыточности электронных документов // «Илмий тадқиқотлар ахборотномаси» илмий-назарий, услубий журнал. – Самарқанд: СамДУ. - № 3(115), 2019. – с. 34 -39.

6. Жуманов И.И., Каршиев Х. Б. Оптимизация достоверности информации в системах электронного документооборота с использованием оценок критериальных характеристик и особенностей документов // «Илмий тадқиқотлар ахборотномаси» илмий-назарий, услубий журнал. – Самарқанд: СамДУ. - №5, 2019. – с. 108 – 116.

7. Жуманов И.И., Каршиев Х. Б. Методы повышения достоверности информации с механизмами использования специфических характеристик и текстурных особенностей электронных документов // Узбекский журнал «Проблемы информатики и энергетики», Изд-во «Фан» АН РУз. - Ташкент, 2019. - №2. - с. 44-51. (05.00.00; № 5).

8. Жуманов И.И., Каршиев Х. Б. Повышение достоверности информации на основе набора –эталонов элементов ключевого концепта документа // Узбекский журнал «Проблемы информатики и энергетики», Изд-во «Фан» АН РУз. - Ташкент, 2019. - №3. - с. 21-30. (05.00.00; № 5).

9. Жуманов И.И., Каршиев Х. Б. Повышение достоверности информации на основе логической и семантической избыточности, связей элементов и отношений ключевых концептов документа // Журнал «Вестник ТУИТ»,

Ташкентский университет информационных технологий - Ташкент, 2019. - №1(49). - с.2-16. (05.00.00; № 31).

10.Жуманов И.И., Каршиев Х. Б. Повышение достоверности информации систем мониторинга организационно – распорядительной деятельности на основе проверки эквивалентности элементов документов // Журнал «Вестник ТУИТ», Ташкентский университет информационных технологий - Ташкент, 2019. - №2(50). - с.2-12. (05.00.00; № 31).

11.Жуманов И.И., Каршиев Х. Б. Повышение достоверности информации электронных документов на основе базы знаний с настройкой матрицы импликаций // «Химическая технология. Контроль и управление» - ТГТУ, Ташкент, 2019- № 3. - с.49 -57. (05.00.00; № 12).

12.Жуманов И.И., Каршиев Х. Б. Анализ эффективности программных инструментов оптимизации достоверности информации электронных документов автоматизированных систем управления // «Химическая технология. Контроль и управление» - ТГТУ, Ташкент, - №4. - с. 57 -66. (05.00.00; № 12).

13.Жуманов И.И., Каршиев Х. Б. Оптимизация достоверности информации на основе базы электронных документов и особенностей правил контроля базы знаний // Журнал Проблемы вычислительной и прикладной математики, Научно-инновационный центр информационно-коммуникационных технологий, Ташкент, 2019.- № 3.с.57-74.(05.00.00; №23).

14.Jumanov, I.I., Karshiev, Kh.B. Increasing information validity using natural redundancy and n-grams of natural language// International journal of advanced research in science, engineering and technology volume 6, issue 9, September 2019 pp. 10937- 10945(05.00.00; № 8).

15. Jumanov, I.I., Karshiev, K.B., Tishlikov, S.A. Examination of the efficiency of algorithms for increasing the reliability of information on criteria of harness and the cost of processing electronic documents// International Journal of Recent Technology and Engineering 2019 №8(2 Special Issue 11), pp. 4133-4139

16.Jumanov, I.I., Karshiev, Kh.B. Mechanisms for optimization of detection and correction of text errors based on combining multilevel morphological analysis with n-gram models // Journal of Physics: Conference Series, Volume 1546, IV International Scientific and Technical Conference «Mechanical Science and Technology Update» (MSTU-2020) 17-19 March, 2020, Omsk, Russian Federation

II бўлим (II часть; II part)

17.Каршиев Х.Б. Исследование возможностей алгоритмов повышения достоверности и обработки информации на основе использования семантической избыточности // Иқтисодий-технологияларнинг тармоқларини инновацион ривожланишида ахборот-коммуникация технологияларининг аҳамияти Республика илмий-техник анжумани. – Ташкент: ТАТУ, 14-15 март 2019 – 269-271б.

18. Каршиев Х.Б. Повышение эффективности систем обработки информации электронных документов по критериям трудоемкости и стоимости // «Ўқув жараёнида ахборот - коммуникация технологияларидан фойдаланишнинг долзарб муаммолари» Республика илмий-амалий конференцияси материаллари тўплами. – Гулистон: Университет, 2019. 222-224 б.

19. Каршиев Х.Б. Повышение достоверности информации на основе морфологии растровых изображений электронных документов // «Современное состояние и перспективы применения информационных технологий в управлении»: доклады республиканскую научно - техническую конференции, Самарканд 5-6 сентября 2019 г. / Научно – информационный центр информационно – коммуникационных технологий. – Ташкент. 2019.с – 299-305

20. Жуманов И.И., Каршиев Х.Б. Оптимизация размещения и хранения документов в базах данных при контроле достоверности информации//»V Международная научно-практическая конференция «Наука и образование в современном мире: вызовы XXI века»10-12 декабря 2019 г. Нур-Султан, Казахстан. –с. 299-303

21. Жуманов И.И., Каршиев Х.Б. Повышение достоверности информации на основе использования ее специфических характеристик в системах электронного документооборота // Иқтисодиётнинг тармоқларини инновацион ривожланишида ахборот-коммуникация технологияларининг аҳамияти. Республика илмий-техник анжумани. – Ташкент: ТАТУ, 5 - 6 март 2020. 475-477 б

22. Жуманов И.И., Каршиев Х.Б. Оптимизация достоверности информации на основе правил контроля базы электронных документов // «V Международная научно-практическая конференция «Наука и образование в современном мире: вызовы XXI века» 10-12 декабря 2019 г. Нур-Султан, Казахстан. с - 303 – 308

23. Жуманов И.И., Каршиев Х.Б. Повышение достоверности информации в системе автоматизированного управления предприятием на основе информационной избыточности элементов документов//» АКТУАЛЬНЫЕ ПРОБЛЕМЫ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ», Международной научной конференции, Национальный университет Узбекистана имени Мирзо Улугбека 14-15 ноября Ташкент. 2019. – с. 246

24. Жуманов И.И., Каршиев Х.Б. Джуманов О.И. Программный комплекс повышения достоверности информации на основе показателя релевантности систем электронного документооборота «Девонхона» ВУЗов // Свидетельство об официальной регистрации программы для ЭВМ, Агентство по интеллектуальной собственности РУз, № DGU 06821, 10.04.2019

25. Жуманов И.И., Каршиев Х.Б. Джуманов О.И. Программный комплекс повышения достоверности информации в системах документооборота “девонхона” вузов с использованием логических,

структурно – технологических и семантических связей элементов и отношений концептов документа // Свидетельство об официальной регистрации программы для ЭВМ, Агентство по интеллектуальной собственности РУз, № DGU 06820, 10.04.2019

26. Жуманов И.И., Каршиев Х.Б., Джуманов О.И. Программный комплекс повышения достоверности информации в системах документооборота «Девонхона» ВУЗов на основе формирования и использования баз данных и знаний // Свидетельство об официальной регистрации программы для ЭВМ, Агентство по интеллектуальной собственности РУз, № DGU 06819, 10.04.2019