

**ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ ва
ЎЗБЕКИСТОН МИЛЛИЙ УНИВЕРСИТЕТИ ҲУЗУРИДАГИ ФАН
ДОКТОРИ ИЛМИЙ ДАРАЖАСИНИ БЕРУВЧИ 16.07.2013.Т/ФМ.29.01
РАҚАМЛИ ИЛМИЙ КЕНГАШ**

**ЎЗБЕКИСТОН РЕСПУБЛИКАСИ АЛОҚА, АХБОРОТЛАШТИРИШ ВА
ТЕЛЕКОММУНИКАЦИЯ ТЕХНОЛОГИЯЛАРИ ДАВЛАТ ҚЎМИТАСИ**

ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ

АХАТОВ АКМАЛ РУСТАМОВИЧ

**ЭЛЕКТРОН ҲУЖЖАТ АЙЛАНИШ ТИЗИМЛАРИДА
МАЪЛУМОТЛАР ИШОНЧЛИЛИГИНИ PR-ОРТИҚЧАЛИК АСОСИДА
ТАЪМИНЛАШ УСУЛИ ВА ДАСТУРИЙ-АЛГОРИТМИК
МАЖМУАЛАРИ**

**05.13.01 – Тизимли таҳлил, бошқарув ва ахборотни қайта ишлаш
(техника фанлари)**

ДОКТОРЛИК ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ

Тошкент – 2014

Докторлик диссертацияси автореферати мундарижаси
Оглавление автореферата докторской диссертации
Content of the abstract of doctoral dissertation

Ахатов Акмал Рустамович

Электрон хужжат айланиш тизимларида маълумотлар
ишончилигини PR-ортиқчалик асосида таъминлаш усул ва
дастурий-алгоритмик мажмуалари.....3

Ахатов Акмал Рустамович

Методы и программно-алгоритмические комплексы
обеспечения достоверности информации в системах электронного
документооборота на основе PR-избыточности.....29

Akhatov Akmal

Methods, software and algorithmic complexes of providing
information authenticity in electronic document
interchange systems on the basis of PR-redundancy.....55

Эълон қилинган ишлар рўйхати

Список опубликованных работ
List of published works.....80

**ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ ва
ЎЗБЕКИСТОН МИЛЛИЙ УНИВЕРСИТЕТИ ҲУЗУРИДАГИ ФАН
ДОКТОРИ ИЛМИЙ ДАРАЖАСИНИ БЕРУВЧИ 16.07.2013.Т/ФМ.29.01
РАҚАМЛИ ИЛМИЙ КЕНГАШ**

**ЎЗБЕКИСТОН РЕСПУБЛИКАСИ АЛОҚА, АХБОРОТЛАШТИРИШ ВА
ТЕЛЕКОММУНИКАЦИЯ ТЕХНОЛОГИЯЛАРИ ДАВЛАТ ҚЎМИТАСИ**

ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ

АХАТОВ АКМАЛ РУСТАМОВИЧ

**ЭЛЕКТРОН ҲУЖЖАТ АЙЛАНИШ ТИЗИМЛАРИДА
МАЪЛУМОТЛАР ИШОНЧЛИЛИГИНИ PR-ОРТИҚЧАЛИК АСОСИДА
ТАЪМИНЛАШ УСУЛИ ВА ДАСТУРИЙ-АЛГОРИТМИК
МАЖМУАЛАРИ**

**05.13.01 – Тизимли таҳлил, бошқарув ва ахборотни қайта ишлаш
(техника фанлари)**

ДОКТОРЛИК ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ

Тошкент – 2014

Докторлик диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида 20.02.2014/В2013.1.Т1 рақам билан рўйхатга олинган.

Докторлик диссертацияси Самарқанд давлат университетида бажарилган.

Докторлик диссертациясининг тўлиқ матни Тошкент ахборот технологиялари университети ва Ўзбекистон Миллий университети ҳузуридаги Фан доктори илмий даражасини берувчи 16.07.2013.Т/ФМ.29.01 рақамли илмий кенгаш www.tuit.uz манзилидаги веб-саҳифасига жойлаштирилган.

Диссертация автореферати уч тилда (ўзбек, рус, инглиз) www.tuit.uz манзилида веб-саҳифа ҳамда www.ziyonet.uz манзили «ZIYONET» ахборот-таълим порталига жойлаштирилган.

**Илмий
маслаҳатчи:**

Жуманов Исраил Ибрагимович
техника фанлари доктори, профессор

**Расмий
оппонентлар:**

Набиев Озод Маликович
техника фанлари доктори, профессор

Нишанов Акрам Хасанович
техника фанлари доктори, профессор

Зайнидинов Хақимжон Насридинович
техника фанлари доктори, профессор

**Етакчи
ташкilot:**

Тошкент давлат техника университети

Диссертация ҳимояси Тошкент ахборот технологиялари университети ва Ўзбекистон Миллий университети ҳузуридаги 16.07.2013.Т/ФМ.29.01 рақамли Илмий кенгашнинг «28» июн 2014 йил соат 10⁰⁰ даги мажлисида бўлиб ўтади (Манзил: 100202, Тошкент, Амир Темур кўч., 108. Тел.: (99871) 238-64-43; факс: (99871) 238-65-52; e-mail: tuit@tuit.uz).

Докторлик диссертацияси билан Тошкент ахборот технологиялари университетининг Ахборот-ресурс марказида танишиш мумкин (01 рақами билан рўйхатга олинган). Манзил: 100202, Тошкент, Амир Темур кўч., 108. Тел.: (99871) 238-64-43.

Диссертация автореферати 2014 йил « » майда тарқатилди.
(2014 йил « » майдаги 1 рақамли реестр баённомаси).

Х.К.Арипов

Фан доктори илмий даражасини берувчи
илмий кенгаш раиси т.ф.д., профессор

М.С.Якубов

Фан доктори илмий даражасини берувчи
илмий кенгаш илмий котиби т.ф.д., профессор

Т.Ф.Бекмуратов

Фан доктори илмий даражасини берувчи
илмий кенгаш ҳузуридаги илмий семинар раиси
т.ф.д., академик

ДОКТОРЛИК ДИССЕРТАЦИЯСИ АННОТАЦИЯСИ

Диссертация мавзусининг долзарблиги ва зарурияти. Ахборотлаштириш жаҳон тажрибасида электрон ҳужжат айланиш тизимлари (ЭХТ), кейинги йилларда нафақат бошқарув жараёнларини автоматлаштирувчи тизимлари, балки ягона ахборот майдонини яратиш учун тўлақонли платформаси сифатида ҳам қаралмоқда ва жорий этилмоқда, бу эса, сўзсиз, улардан фойдаланиш соҳаларини кенгайтиришга ва илмий ҳамжамиятда унга бўлган қизиқишнинг ортишига сабаб бўлди. Ахборот-коммуникация технологиялари соҳасининг етакчи олимлари тадқиқотларида ЭХТ тезкорлиги ва унумдорлигини оширадиган самарали технологияларини яратиш учун иш юритиш ҳужжатларини шакллантиришда ахборотни ишончли йиғиш, узатиш, таҳлил, кодлаш масалаларининг ечимига бўлган эҳтиёж алоҳида ажратиб кўрсатилмоқда.

Ўзбекистон Республикаси ҳукумати томонидан амалга оширилаётган ҳудудий автоматлаштирилган бошқарув тизимларини ривожлантириш ҳамда ягона ахборот майдонини яратиш бўйича чора-тадбирлар ахборот тизимлари, ЭХТ, маълумотлар базаларини (МБ) замонавий ахборот-коммуникация технологиялари асосида кенг жорий этишга қаратилган. Шу сабабли, электрон ҳужжат айланиш тизимлари тузилмасида минимал даражадаги моддий ҳаражатлар ва вақт сарфи билан хатоларни самарали аниқлаш ва тузатиш имконини берадиган, ахборот узатиш сифатини ошириш учун фойдаланадиган ахборот ресурсларига тафаккурли ишлов бериш усулларини ишлаб чиқиш халқ хўжалигида жиддий аҳамият касб этиб, ҳозирга қадар тўлиқ ечимини топмаган назарий ва амалий муаммолар қаторига киради.

Ахборот ресурслари ва маълумотларни узатиш оқимларига қўйиладиган талаблар ЭХТ самарадорлиги ва ишлаш сифатинининг муҳим омили сифатида маълумотлар барқарорлиги, яхлитлиги, бутунлиги ва ишончилигини таъминлашда ўз ифодасини топган. Бундай омиллар орасида маълумотлар ишончилиги мезони, узатиладиган маълумотларда техник воситаларнинг турли узилиш ҳолатлари ва носозликлари, алоқа каналларига бўладиган ташқи таъсирлар, оператор хатолари ва сканерловчи ҳамда танувчи тизимлар ноаниқликлари сабабли хатоликлар пайдо бўлиши туфайли жиддий эътиборга молик.

Шунингдек, корхона ва ташкилотларда автоматлаштирилган бошқарув ва электрон ҳужжат айланиш шароитларига хос, маълумотларга ишлов беришнинг устувор технологияси сифатида ахборотларни узатиш ва уларга ишлов беришнинг ишончилигини назорат қилувчи самарали тизимларини қуриш бошқача илмий қизиқиш уйғотади.

Мавжуд усуллар маълумотларни узатиш ишончилигини юқори даражада таъминлашига қарамасдан, ўзига хос ечилмаган масалалар мавжуд. Асосийлари сифатида қуйидагиларни ажратиб кўрсатиш мумкин: электрон ҳужжат айланиш технологияларини ишлаб чиқишда маълумотларни узатиш пакетларида катта ҳажмдаги моддий манбалар сарлавҳаларни узатишга сарфланади, шу билан бирга сарлавҳа маълумотларнинг кўп қисми кадрни

узатиш оқими мобайнида пакетдан-пакетгача ўзгармас бўлади (бу жараёнда пайдо бўладиган информацион ортиқчалик ва узатиш ишончилигини таъминловчи механизмлар, асосан, хос хабарларни ва пакетларни қайта узатишга асосланади, бу эса, ўз навбатида, хатоларни аниқлаш ва тузатиш учун кўшимча вақт сарфланишига ва моддий ҳаражатларнинг ортишига олиб келади); маълумотлар ишончилигини ошириш учун кодлаш ва аппаратурани усуллар, асосан, пакетнинг бошқарув майдонида содир бўладиган бир ва икки каррали хатоларни тузатади, аммо маълумотларни узатишда ахборот майдонлар ҳам хатоликларга учрайди ва улар кўп каррали (уч-, турт-, к-каррали) матнли маълумотлардаги хатолар кўринишида бўлади.

Шу туфайли, ахборотларни узатиш ишончилигини ошириш масаласи ечимини икки соҳада кўриш мақсадга мувофиқдир.

Биринчи турдаги масалалар ечими маълумотларни киритувчи оператор, сканерлаш ва бошқа ускуналар хатоликларини ҳисобга олади. Шунини алоҳида таъкидлаш керак-ки, бу турдаги хатолар нотўғри қабул қилинган матн маълумотларининг катта ҳажмини ташкил этади ва OSI (Open System Interconnection reference model) моделининг Амалий ва Тақдимот босқичларида содир бўлади.

Ахборот ишончилиги назоратига бағишланган иккинчи турдаги масалалар OSI моделининг Транспорт, Тармоқ, Физик ва Канал босқичларида пайдо бўладиган хатоликлар эҳтимолларини ҳисобга олади.

Диссертацияни бажарилишига бўлган зарурият шунинг билан ифодаланадики, электрон хужжат айланиши IP-технологиясининг кенг доирада жорий этилиши хужжатларни тайёрланиш ва расмийлаштириш жараёнида хатоликларни аниқлаш ва бартараф этишга жиддий эътибор қаратишни талаб этмоқда.

Мазкур тадқиқот «Ахборотлаштириш», «Электрон рақамли имзо», «Электрон хужжат айланиш», «Электрон тижорат», «Электрон тўловлар» тўғрисидаги Ўзбекистон Республикасининг қонунлари ҳамда Ўзбекистон Республикаси Вазирлар Маҳкамасининг 2011 йил 4 майдаги «Вазирлар Маҳкамасининг ижро этувчи аппаратида, давлат ва хўжалик бошқаруви, маҳаллий давлат ҳокимияти органларида ягона ҳимояланган электрон почтани ва электрон хужжат айланиш тизимини жорий этиш ҳамда улардан фойдаланиш чора-тадбирлари тўғрисида»ги 126-сонли қарорини ижросини таъминлашга хизмат қилади.

Шу жиҳатдан юқорида санаб ўтилган масалалар ечими маълумотларни узатувчи пакетлар таркибида маълумотлар ишончилигини информацион ортиқчаликдан фойдаланишга асосланган назорат қилувчи усул ва алгоритмларни, шаклланиб бўлган узатиш муҳитида ҳам ишлайдиган ва мавжуд камчиликларни бартараф этишга қаратилган махсус тадқиқотни бажариш ва ишланмаларни яратишни талаб қилади. Мазкур ҳолат ишлов бериладиган объект хусусиятига кўра аниқланадиган янги турдаги PR-ортиқчалик асосида ахборот ишончилигини таъминловчи махсус алгоритмлар синфини ажратиб олиш заруриятини асослайди.

Тадқиқотнинг Ўзбекистон Республикаси фан ва технологиялар тараққиётининг устувор йўналишларига мослиги. Диссертация фан ва технологиялар ривожланишининг устувор йўналишига мос равишда бажарилган: ИДТ-17 – «Ахборот телекоммуникация технологияларини кенг ривожлантириш ва жорий этишни таъминлайдиган замонавий ахборот тизимлари, бошқарув ва ўқитишнинг интеллектуал воситалари, илмий-техникавий маълумотлар базаси ва дастурий маҳсулотларини ишлаб чиқиш»; ИТД-5 – «Жамиятни ахборотлаштириш даражасини оширишга йўналтирилган илмий ҳажмдор ахборот технологияларни, телекоммуникацион тармоқларни, аппарат-дастурий воситаларни интеллектуал бошқариш, ўқитиш усуллари ва тизимларини ишлаб чиқиш».

Диссертация мавзуси бўйича халқаро илмий тадқиқотлар шарҳи. Ахборотларни узатиш ва уларга ишлов бериш ишончилигини автоматлашган назорат қилиш усул ва воситаларини табиий тил хусусиятларини ҳисобга олган ҳолда ишлаб чиқиш бўйича тадқиқот ишлари АҚШ, Буюк Британия, Россия, Хитой, Германия, Австрия, Швеция, Норвегия, Финляндия, Туркия, Исроил, Канада, Корея, Лотин Америкаси (Бразилия, Аргентина, Чили) ва бошқа мамлакатлар олимлари томонидан олиб борилиб, маълум даражадаги натижаларга эришилмоқда.

Gartner, International Data Corporation, Wobot, Forrester, CNews Analytics, ЮНИПРАВЭКС, Прогноз, Итерфейс ИТ, Интеллект Телеком ва бошқа халқаро эксперт компанияларининг нашрларида эълон қилинишича, ЭХТ технологияларининг ривожланиши ҳужжатларга ишлов бериш, WEB орқали кўп тармоқли бошқарув ва офис жараёнларини автоматлаштириш усуллари интеграциялаш йўналишида тараққий этмоқда. Шу қаторда, уларга илгари боғлиқ бўлмаган, инсон омили таъсирини пасайтирувчи ускунавий мажмуа сифатидаги, электрон ҳужжатларнинг маълумотларини киритиш, узатиш ва ишлов беришда пайдо бўладиган хатоларини автоматлашган аниқлаш ва тузатиш, табиий тил имлосини назорат ва таҳрирлаш технологияларини бирлаштириш долзарблиги ва эҳтиёжи таъкидланиб келмоқда.

Шу каби, тармоқ хизматларидан фойдаланишда эҳтиёжлар аниқланиши билан, маълумотларни узатиш технологиясига асосланган янги йўналиш ва иловаларни яратишга бағишланган тадқиқотлар фаоллашмоқда. Илмий-техник адабиётларни таҳлилий ўрганиш шуни кўрсатади-ки, электрон ҳужжат айланишининг, электрон рақамли имзо ёки маълумотлар ҳаққонийлиги ва яхлитлигини кафолатлайдиган технологияларини қўллаш билан бирга шубҳасиз афзаллиги – электрон ҳужжатларни шакллантириш, киритиш, узатиш ва қайта ишлашда пайдо бўладиган хатоларни инсон иштирокини камайтирувчи маълумотларни автоматлашган назорат қилиш усуллари ҳисобига аниқлаш ва тузатиш имкониятидир.

Муаммонинг ўрганилганлик даражаси. Матнли маълумотларни узатиш ишончилигини ошириш муаммосининг тадқиқот этиш услубиятини тараққий этган ҳамда, ахборотнинг сунъий ортиқчалигини қўллашга асосланган кодлаш назарияси ва аппаратли услублар ташкил қилади, бу соҳа,

ўз навбатида, кўп сонли монография, диссертация тадқиқотлари ва илмий мақолаларда ўз аксини топган. Таъкидлаш жоизки, ЭХТларда матнли маълумотларни узатиш ва қайта ишлашнинг асосий хусусиятларидан бири бу уларда ишлатилаётган ҳужжат ва хабарлар таркибида сунъий, табиий, статистик инфор­мацион ва таркибий-технологик ор­тиқчаликнинг мавжудлигидир.

Тадқиқотнинг натижалари маълумотларни узатишнинг мавжуд услубларини қулай тўлдирувчи ва қўйилган турли табиатли ор­тиқчаликдан фойдаланиш муаммосининг ечимида янги йўналишнинг илмий-услубий асослари бўладиган, ЭХТларнинг ахборот ресурсларини узатиш ва қайта ишлаш ишончилигини таъминловчи усул, алгоритмик ва дастурий мажмуалар кенг кўламини касб этиши лозим.

Диссертация тадқиқотининг илмий-тадқиқот ишлари режалари билан боғлиқлиги қуйидаги лойиҳаларда ўз аксини топган:

давлат илмий-техника дастури (ДИТД)-4 – «Ахборотлаштириш, бошқарувни автоматлаштириш тизимлари ва технологик жараёнларни мувофиқлаштириш дастурий воситалари ва услубларини ишлаб чиқиш» (2000-2002 й.й.);

ДИТД-20 – «Информацион ва телекоммуникация технологияларини кенг кўламда ривожлантиришни таъминловчи замонавий инфор­мацион тизим, бошқарув воситалари, маълумотлар базаси ва дастурий маҳсулотларни ишлаб чиқиш» (2003-2005 й.й.).

Тадқиқотнинг мақсади PR-ор­тиқчаликдан фойдаланишга асосланиб, маълумотларни узатишда ахборот ишончилигини назорат қилувчи конструктив усул, модел, алгоритм ва тизимларини яратиш, уларни электрон ҳужжат айланиш технологияларини ишлаб чиқишда дастурий-алгоритмик мажмуаларни жорий этишдан иборат.

Мақсадга эришиш учун қуйидаги **тадқиқот вазифалари** қўйилган:

OSI моделининг амалий ва тармоқ босқичларида турли хил PR-ор­тиқчаликдан фойдаланиш бўйича маълумотлар ишончилигини назорат қилиш усул, дастурий-алгоритмик мажмуалар ва тизимларни яратиш ҳамда ривожлантиришнинг услубиятини танлаш;

сунъий, табиий ва семантик PR-ор­тиқчаликдан фойдаланиб йиғиндиларни ҳосил қилиш, матнларни арифметик ва статистик кодлаш механизмлари бўйича маълумотлар ишончилиги назорати усул, модел ва алгоритмларини ишлаб чиқиш;

матн элементларини парсингли кодлаш, n-граммли тавсифлаш, статистик таниш ва хеш-кодлаш механизмларини тадқиқ қилиш ва уларни маълумотлар ишончилигини назорат қилувчи усул, модел ва алгоритмларини ишлаб чиқишда қўллаш;

кўп босқичли морфологик таҳлил қилиш ва хатоларни n-граммли тавсифлаш асосида матнлар ишончилигини назорат қилувчи усул, модел ва алгоритмларни ишлаб чиқиш; имлони назорат ва таҳрир қилувчи тизимнинг асосий компонентларини моделлаштириш; назорат сифати баҳосини тавсифлаш;

таркиблашган мантиқий мезонлар, маълумотлар ва билимлар базалари, эксперт тизимлари асосида таркибий-технологик PR-ортиқчаликни шакллантириш ва маълумотлар ишончилигини назорат қилувчи модел ва алгоритмларини ишлаб чиқишда жорий этиш;

ахборот ресурсларига интерфаол ишлов бериш тизимни матнлар ишончилигини назорат қилувчи алгоритмлар синтези асосида ишлаб чиқиш, тажрибали синаш ва самарадорлигини баҳолаш, телекоммуникация тизимлари ва ЭҲТда натижаларни жорий этиш самарадорлигини ҳисоблаш.

Тадқиқот объекти сифатида электрон ҳужжат айланиш тизимларида маълумотлар ва ҳужжатларни тўплаш, шакллантириш, киритиш, узатиш ва уларга ишлов бериш ахборот ресурслари ҳамда технологиялари қаралган.

Тадқиқот предмети – турли табиатдаги информацион PR-ортиқчалигидан фойдаланиш механизмига асосланган матнли маълумотларни назорат қилувчи усул, модел, алгоритмлари ва тизими.

Тадқиқот усуллари. Тадқиқот жараёнида тизимли таҳлил, ахборот узатиш назарияси, тармоқ технологиялари, эҳтимоллар назарияси, қиёфаларни таниш, компьютер лингвистикаси, алгоритмлаш ва моделлаштириш ҳамда объектга йўналтирилган дастурлаш усул ва алгоритмлари қўлланилган.

Диссертация тадқиқотининг илмий янгилиги қуйидагилардан иборат: электрон ҳужжат айланиш тизимларида маълумотлар ишончилигини таъминлаш усул, модел ва алгоритмларни яратишнинг концепция, услубияти ва дастурий-алгоритмик асослари ишлаб чиқилган, PR-ортиқчаликка эга бўлган объектлар синфлари ЭҲТда ахборот ресурслари аниқлиги, яхлитлиги, тезкорлиги, ихчамлиги, мурожаатини таъминлаш учун ажратиб қаралган;

сунъий ортиқчаликдан фойдаланувчи чизиқли, модулли ва текислик бўйича йиғиндиларни ҳосил қилиш ҳамда кодланган тўпламларга тегишлиликни аниқлаш механизмларига асосланган маълумотлар ишончилигини назорат қилувчи усул ва алгоритмлар таклиф қилинган;

матн элементларини статистик, арифметик, парсингли кодлаш, таркиб-лаштирувчи n-граммалар бўйича тавсифлаш, статистик таниш ва ҳешлаш алгоритмлари асосида табиий ортиқчаликдан фойдаланувчи маълумотларни назорат қилиш усул ва дастурий мажмуалари ишлаб чиқилган;

кўп босқичли морфологик таҳлил ва n-граммалар грамматикасининг тавсифи моделлари асосида ўзбек тилида бериладиган матнлардаги орфографик хатоларни назорат ва таҳрир қилувчи усул ва алгоритмлари таклиф қилинган;

ички ўрнатиладиган эксперт тизимлардаги таркиблашган мантиқий мезонлар, маълумотлар ва билимлар базалари асосида таркибий-технологик PR-ортиқчаликдан фойдаланувчи ахборот ишончилигини назорат қилиш модел ва алгоритмлари ишлаб чиқилган;

электрон ҳужжат айланиш технологияларини ишлаб чиқишда хатоларни назорат ва таҳрир қилувчи интерфаол тизимлар таркибида матнли маълумотлар ишончилигини назорат қилиш алгоритмларини синтези усуллари ишлаб чиқилган.

Тадқиқотнинг амалий натижалари қуйидагилардан иборат:

сунъий, табиий ва статистик, таркибий-технологик, дастурий-вақтли PR-ортиқчаликлардан фойдаланувчи дастурий усулларни синтезлаш асосида маълумотлар ишончилигини ошириш тизимлари ишлаб чиқилган ва уларнинг мажмуи OSI моделининг тадбиқий ва тармоқ даражаларида қўлланаётган кодли ва аппаратурали усуллар имкониятларини кенгайтиради ҳамда матнли маълумотлар ишончилигини таъминлайди;

ишлаб чиқилган усуллар самарадорлиги хато аниқланмаслиги эҳтимоли мезони асосида намоёниши бўйича, улар 92% хатоларни аниқлайди ва бир, икки қаррали ва қўшма транспозицияли хатоларни тузатишга қодир; бундан ташқари, назорат мураккаблиги ва жорий этиш ҳаражатларини хатолар содир бўлишининг ўртача эҳтимоли шароитида 2-3 маротаба пасайтиради, ахборот ишончилигини эса, уч тартибгача оширади;

морфологик ва n-граммли таркибланишган моделлар ҳамда луғат, статистик ва хеш-кодлаш усулларига асосланган матнли ахборотларни узатишда орфографик хатоларни назорат ва таҳрир қилувчи тизимлар ишлаб чиқилган, уларни параллел ҳисоблаш NVIDIA дастурий-аппарат муҳитида сонли таҳлил стандарт кутубхоналари ҳамда CPU ва GPU орасида мақбул маълумотлар алмашувини берувчи воситаларни амалда қўллаш таклиф этилган;

фреймворк Сфинкс-4 архитектураси турли тил моделларига асосланган дастурий модуллари учун амалга оширилган ва тизимнинг яратилган вариантида морфологик таҳлилга асосланган имлони назорат қилувчи тизимга нисбатан аниқланмаган хатолар сонини, жорий этиш ҳаражати ҳамда назорат мураккаблигини икки маротаба камайтиришга эришилди.

Олинган натижаларнинг ишончилиги таклиф этилган моделлар учун ўтказилган математик тадқиқот, олинган формула ва амалий натижалар, умумқабул қилинган мезонлар асосида ўтказилган қиёсий таҳлил билан асосланди. Тадқиқот натижаларини мантиқий асосланишини баҳолаш учун дастурий мажмуалар оммабоп адабий ва техник адабиётларнинг кенг контексти бўйича тест синовидан ўтказилди, дастурий мажмуалар самарадорлигининг қиёсий таҳлили назорат мураккаблиги, жорий этиш ҳаражатлари ва хатолар аниқланмаслиги эҳтимоли мезонлари бўйича ахборот ишончилигини назорат қилувчи кўп синфли алгоритмлари учун бажарилган.

Тадқиқот натижаларининг назарий ва амалий аҳамияти. Тадқиқотда олинган натижаларнинг назарий моҳияти шундан иборатки, турли табиатли – сунъий, табиий, статистик, таркибий-технологик, дастурий-вақтли PR-ортиқчаликдан фойдаланиш тамойилларига асосланиб ишлаб чиқилган усул, модел ва алгоритмлар OSI моделининг тадбиқий ва тармоқ босқичларида мавжуд кодли ва аппаратурали усулларни муваффақиятли тўлдирувчи ва ахборот ишончилигини таъминловчи тадбиқий усуллар синтези йўли билан ЭҲТда ахборот ишончилигини оширувчи тизимларни яратиш концепцияси ва дастурий-алгоритмик мажмуалари асосларини ташкил этади.

Диссертация ишининг амалий аҳамияти матнлар грамматикасини кўп босқичли тавсифлаш ва n-граммли таркибланишган моделлари, ўзбек тили имлосини назорат ва таҳрир қилиш жараёнларини дастурий жорий этиш учун PR-ортиқчаликни қўллаш билан аниқланади ва бу жараёнлар n-граммли кўрсаткичларни самарали ҳисоблаш ва улардан фойдаланиш, статистик ва парсингли таниш ва кодлаш механизмлари асосида маълумотлар ишончилигини назорат қилиш қоидаларини шакллантириш, OSI тармоқ моделининг тадбиқий ва тақдимот босқичларида хатоларни аниқлаш ҳамда тузатиш имкониятини яратади. Маълумотлар ишончилигини назорат қилиш дастурий мажмуаларини ЭХТ тузилмасида жорий этиш ҳужжатларни шакллантиришда вақт сарфи ва моддий харажатларни камайтиради, ахборот алмашув жараёнларини инсон омилига боғлиқ бўлган хатоларни аниқлаш ва тузатиш ҳисобига жадаллаштиради.

Тадқиқот натижаларининг жорий қилиниши. PR-ортиқчаликка эга бўлган объектлар синфлари ҳисобига ЭХТда ахборот ресурслари аниқлиги, яхлитлиги, тезкорлиги ва ихчамлиги юзасидан ишлаб чиқилган модел, алгоритм, дастурий мажмуалар Ўзбекистон Республикаси Алоқа, ахборотлаштириш ва телекоммуникация технологиялари давлат қўмитасининг «ЎзТелеком» компанияси Самарқанд филиалида амалиётга жорий этилган («СамТелеком» ШК, 23.05.2010 й.). Маълумотларни адаптив узатиш, қайта ишлаш ва таҳлил қилиш тизимлари бўйича ишланмалар «Lazzat Meva» қўшма корхонаси фаолиятига жорий этилиб, йиллик иқтисодий самара 62,2 млн. сўмни ташкил этган («Lazzat Meva» ҚК, 14.05.2010 йилдаги далолатнома).

Ишнинг апробацияси. Тадқиқот натижалари 30 та илмий-амалий анжуманлар, шу қаторда 14 та халқаро симпозиум, конгресс ва семинарларда, хусусан «Инновация» (Тошкент, 2004-2009), «Алоқанинг техника ва технологияси» (Алмати, 2002), «Ал-Хоразмий 2009 – амалий математика ва ахборот технологияларининг долзарб муаммолари» (Тошкент, 2009), «ICI-2006», «ICI-2007» (Тошкент, 2006-2007), «ICEIC 2008» (Тошкент, 2008), Телекоммуникация бўйича олий таълим муассасаларининг X Халқаро илмий-услубий анжумани (Россия, Ростов-на-Дону, 2008), «WCIS-2008», «WCIS-2010», «WCIS-2012» (Тошкент, 2008-2012), Турк математиклар III конгресси (Козоғистон, Алмати, 2009) ҳамда «Технологияларнинг замонавий ҳолати ва ривожланиш йўллари» (Тошкент, 2008), «Математика, механика ва ахборот технологияларининг ҳозирги муаммолари» (Тошкент, 2008), «Математика ва информатиканинг долзарб муаммолари» (Самарқанд, 2007-2010) республика илмий анжуманлари ва Республика инновацион ғоя, технология ва лойиҳалар Ярмаркаларида (Ташкент, 2008-2011) апробациядан ўтказилган.

Натижаларнинг эълон қилинганлиги. Диссертация мавзуси бўйича 87 та илмий иш, жумладан, 9 та илмий мақола халқаро журналларда чоп этилган.

Диссертациянинг тузилиши ва ҳажми. Диссертация кириш, олти боб, хулоса, фойдаланилган адабиётлар рўйхати, 11 та илова ва 197 саҳифа матн, 47 та расм ва 12 та жадваллардан иборат.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Кириш қисмида диссертация тадқиқотининг долзарблиги ва зарурияти асосланган, тадқиқот мақсади ва вазифалари ҳамда объект ва предметлари тавсифланган, Ўзбекистон Республикаси фан ва технологияси тараққиётининг устувор йўналишларига мослиги кўрсатилган, тадқиқот илмий янгилиги ва амалий натижалари баён қилинган, олинган натижаларнинг назарий ва амалий аҳамияти очиб берилган, тадқиқот натижаларини жорий қилиш рўйхати, нашр этилган ишлар ва диссертация тузилиши бўйича маълумотлар келтирилган.

Диссертациянинг **биринчи бобида** мавжуд маълумотлар ишончлилигини оширишнинг кодли ва аппаратурали усуллари назарий-информацион таҳлили амалга оширилган; ахборот ишончилигини оширувчи тизимларни қуриш концепциясини касб этадиган, тадқиқотда PR-ортиқчалик деб қабул қилинган турли табиатли – сунъий, табиий ва статистик, таркибий-технологик ва дастурий-вақтли ортиқчаликдан фойдаланиш асосида маълумотларни узатиш ишончилигини назорат қилиш усул, модел ва алгоритмларини ишлаб чиқишнинг вазифалари аниқланган.

Ишда рақамли телекоммуникация соҳаси иловаларининг кўпида кенг қўлланиладиган Рид-Соломон, БЧХ, FEC каби услубларида амалга ошириладиган кодли ортиқчалик тамойилини тасвирлайдиган хатоларни назорат ва таҳрир қилиш схемаси берилган. Хатоларни код ортиқчилиги бўйича аниқловчи ва тузатувчи алгоритмларнинг тамойил ва механизмларининг тизимли таҳлили, шуни кўрсатди-ки, улар бир, икки, уч каррала хатоларни аниқлаш билан чекланади, ҳамда сунъий равишда назорат разрядини код таркибига киритишга асосланади. Бу эса, хатоларни назоратловчи кодли ва аппаратли усуллар имкониятини оширишга йўналтирилган тадқиқотни асослайди.

Матнли маълумотларни узатиш ва ишлов бериш тизимида ахборот ишончилигини ошириш учун турли табиатли – сунъий, таркибий-технологик, дастурий-вақтли, статистик (табиий) ортиқчаликдан фойдаланиш бўйича маълумотларни узатувчи пакетларда ишончилиликни назорат қилувчи усул ва алгоритмларни ишлаб чиқишга чамаланган тадқиқот ва ишланмаларни бажариш гипотезаси таклиф қилинган. Матнли маълумотлар ишончилигини назорат қилиш усул, алгоритмлари ва тизимларини синтезлашга қаратилган асосий ёндошувлар ишлаб чиқилган. Бажарилган тадқиқотларга кирган масалалар кетма-кетлиги матнли маълумотлар ишончилигини назорат қилувчи автоматлаштирилган тизимларнинг (ММИНТ) таркибий-функционал компонентларни қуриш схемаси сифатида тасвирланган.

Қўйилган масалаларни ечиш ва ишлаб чиқиладиган усул ва алгоритмларни самарадорлигини PR-ортиқчилиги бўйича оширишни баҳолашда хатоларни аниқланмаслик эҳтимоли талаб қилинган чегараларгача камайтирилиши мезон сифатида қабул қилинган.

Кодли S_0 сўзнинг қабули эҳтимолини баҳолаш формуласи қуйидагича

$$P(S_0/S_1') = \sum_{i=1}^{2^m-1} (1 - P(S_0))^{-1} P(S_i/S_i'),$$

бунда $P(S_0)$ – дастлабки кодли сўзда хатонинг аниқланиш эҳтимоли,

$P(S_i/S_1')$ – S_i кодли сўзни қабул қилинган S_1' -га нисбатан бўлган шартли эҳтимоли,

m – сўзга кирган кодли разрядлар сони.

Кодли сўзни нотўғри қабул қилиш эҳтимолини баҳолаш қуйидаги ифода бўйича амалга оширилади:

$$P_H(S) = \sum_{i=0}^{2^m-1} P(S_i) \sum_{\substack{j=0 \\ j \neq i}}^{2^m-1} P(S_j) = \sum_{j=0}^{2^m-1} P(\Lambda_j),$$

бунда Λ_j - нотўғри қабул қилинган кодли сўзлар элементларининг тўплами.

Хатоларни аниқланмаслик эҳтимолини баҳолаш ифодаси қуйидагича:

$$P_{ocm}(S) = 1 - (1 - P_{\alpha_i})^m - \sum_{j=1}^{2^m-1} P(\Lambda_j).$$

Маълумотларни пакетли узатиш ишончилиги кўрсаткичларини, ҳамда кўп босқичли OSI модели бўғинларида маълумотлар ишончилигини баҳолаш учун ахборот узатиш жараёни эҳтимоллари шакллантирувчи схема ва ўтиш эҳтимоллари диаграммалари таклиф этилган. Маълумотларни узатиш хатоларини аниқланишини таъминловчи қўшимча ахборотни минимал миқдорини топиш услубияти ишлаб чиқилган. Маълумот узатувчи пакетларда қўлланиладиган ва S_i хабардаги хатони S'_k қабул қилинган белги маълум бўлганида аниқлашга чамаланган ихтиёрий код усули учун ахборот миқдори қуйидагича ҳисобланади:

$$J(S_i/S'_k) = m + \frac{1}{2^m} \sum_{k=1}^{2^m} \sum_{i=1}^{2^m} P(S_i/S'_k) \log(S_i/S'_k) = m + \sum_{i=1}^{2^m} P(S_i/S'_k) \log(S_i/S'_k) = \quad (1)$$

$$= m + Q' \log Q' + P_H' \log P_H(S).$$

Ифода (1) кейинги ўзгартиришларида, $Q'(S) = Q(S)[1 - P_0(S)]^{-1}$ ва $P_H'(S) = P_H(S)[1 - P_0(S)]^{-1}$ ҳисобга олиниб, ушбу кўринишга олиб келинади

$$J(S_i/S'_k) = m + \frac{Q(S)}{1 - P_0(S)} \log Q(S) - \frac{Q(S)}{1 - P_0(S)} \log[1 - P_0(S)] + \frac{P_H(S)}{1 - P_0(S)} \times$$

$$\times \log P_H(S) - \frac{P_H(S)}{1 - P_0(S)} \log[1 - P_0(S)] - \frac{P_H(S)}{1 - P_0(S)} \log(2^m - 1) = m - \frac{P_H(S)}{1 - P_0(S)} \times$$

$$\times \log(2^m - 1) - \log[1 - P_0(S)] + \frac{Q(S)}{1 - P_0(S)} \log Q(S) + \frac{P_H(S)}{1 - P_0(S)} \log P_H(S)$$

бунда $P_0(S)$ - хатолик S хабарда аниқланиш эҳтимоли;

$Q(S)$ - маълумотни тўғри қабул қилиш эҳтимоли.

Ифодадаги $J(S_i/S'_k)$ қиймати – маълумотларда хатолар аниқланмаслиги туфайли содир бўладиган йўқотишларни бартараф этиш учун керак бўладиган минимал қўшимча ахборот миқдори.

Ихтиёрий кодли усул самарадорлигини аниқлаш учун ортиқчалик ҳажми баҳосини соддалаштирилган ифодаси қуйидаги кўринишда олинган:

$$R(S_i) = 1 - \frac{1}{(1-P)} \log \frac{(2^m - 1)}{nP}.$$

Яратилган услубият бўйича турли табиатдаги информацион ортиқчалигидан фойдаланишга асосланган ахборот ишончилигини назорат қилувчи дастурий услублар самарадорлигини баҳоланиши кўрсатилди.

Диссертациянинг **иккинчи бобида** маълумотларни узатиш пакетлар таркибида сунъий ортиқчаликдан фойдаланиш тамойиллари бўйича маълумотлар ишончилигини назорат қилиш усул, модел ва алгоритмлари ишлаб чиқилган.

Мавжуд дастурий усулларга хос бўлган камчиликларни бартараф этувчи, чизикли, текислик ва модуллар бўйича йиғиндилар ҳосил қилишга асосланган маълумотлар ишончилигини назорат қилиш усул, модел ва алгоритмлари ишлаб чиқилган. Шунинг алоҳида таъкидлаш керак-ки, йиғиндиларни ҳосил қилиш бўйича маълумотлар ишончилигини назорат қилувчи алгоритмлар қабул қилинган кетма-кетликларда нафақат хато борлигини, балки унинг жойлашиш ўрни хатоликларини автоматик тузатиш учун кўрсатади. Улар қуйидаги қоидаларга асосланган: π_1 – маълумотлар ишончилиги назорати; π_2 – Q_1 хатоларни аниқлаш ва тузатиш; π_3 – Q_2 хатоларни аниқлаш ва тузатиш; π_4 – Q_3 хатоларни аниқлаш; π_5 – Q_4 хатоларни аниқлаш.

Турли хил алгоритмлар учун маълумотлар ишончилигини назорат қилиш самарадорлигини хатоларни аниқланмаслик эҳтимоли мезони бўйича баҳолаш масалаларининг умумий ва хусусий ечимини олиш услубияти ишлаб чиқилган.

Чизикли йиғинди бўйича маълумотлар ишончилигини назорат қилиш қоидаси учун хатолар аниқланмаслик эҳтимоли йиғиндиси қуйидагича.

$$P_H = P_1 + P_2 + P_3 + P_4 = \{67[mnp^2(1-p)^{m+n-2} + C_m^2 C_n^2 p^4 (1-p)^{m+n-4} + C_m^k C_n^3 p^{3k} (1-p)^{m+n-k-3}] + 5,76 \cdot 10^{-6}\}.$$

Ишда назорат мураккаблиги ва жорийлаш ҳаражати мезонлари бўйича алгоритмлар самарадорлигининг графиклари тадқиқ қилинган ва уларда ишлаб чиқилган усул мавжуд дастурий усулларга нисбатан назорат мураккаблиги ва жорийлаш ҳаражатларини хатоларни тузатишда 2-3 мартаба камайтириши мумкинлиги кўрсатилган.

Назорат қилувчи алгоритмларнинг P_H -хатолар аниқланмаслик эҳтимоли P -хатолар эҳтимолидан боғлиқлиги графиклари бўйича маълумотлардаги хатолик эҳтимоли $10^{-3} - 10^{-4}$ бўлганда, хатоларни аниқлаш имконияти тадқиқ қилинган алгоритмларда икки тартибга, назоратнинг мураккаблиги ва ҳаражати мезонлари эса хатоларни тузатишда 2-3 мартаба камаяди.

Тадқиқотда йиғиндиларни разрядлар - Π_2 , рақам - Π_3 , текислик - Π_6 бўйича ҳосил қилишга асосланган алгоритмларнинг ахборот ишончилиги кўрсаткичлари энг самарадор эканлиги кўрсатилди.

Рақамли ва сонли модуллар бўйича йиғиндиларни 8, 9, 10, 11 базисларда ва ихтиёрий q га асосланган алгоритмлар самарадорлиги таҳлил қилинди. Тадқиқотда, рақамли 8, 10 ва 11 модулли йиғиндиларга асосланган алгоритмларни қўллашда ахборотларнинг ишончилилик кўрсаткичи энг катта самарадорликга эга эканлиги кўрсатилди.

Ишлаб чиқилган алгоритмлар барча турдаги хатоларни 92% гача аниқлаши ва бир-, икки- каррали қўшма транспозицияли хатоларни ҳам тузата олишлиги исботланди.

Бир каррали хатоларни назорат қилувчи алгоритмнинг биринчи ва иккинчи турдаги хатолар эҳтимоли аниқловчи формулалар олинган:

Биринчи турдаги:

$$P_1 \approx 2P_{\alpha_i}(1 - P_{\alpha_i}) \cdot 10^{-2} + P_{\alpha_i}^2 \cdot 10^{-2} + 2P_{\alpha_i}(1 - P_{\alpha_i}) \cdot 10^{-4}.$$

Иккинчи турдаги:

$$P_2 \approx (1 - P_{\alpha_i})^2 \cdot 1.5 \cdot 10^{-5} + P_{\alpha_i}(1 - P_{\alpha_i}) \cdot 1.5 \cdot 10^{-5}.$$

Алгоритмнинг хато аниқланмаслиги эҳтимоллари йиғиндиси:

$$P_H \approx 2P_{\alpha_i}(1 - P_{\alpha_i}) \cdot 10^{-2} + (1 - P_{\alpha_i})^2 \cdot 1.5 \cdot 10^{-5}.$$

Хато аниқланмаслик эҳтимоли мезони бўйича масалани ечиш учун яратилган услубият бўйича барча турдаги хатоларни назорат ва аниқлашга чамаланган алгоритмлар самарадорлиги ҳам таҳлил қилинди.

Диссертациянинг **учинчи бобида** статистик, таркибий-технологик, дастурий-вақтли ортиқчаликдан фойдаланиш тамойиллари бўйича маълумотлар ишончилигини назорат қилиш усул, модел ва алгоритмларини ишлаб чиқилган.

Мантиқий, технологик, статистик боғланишлар ва ахборот хоссаларидан фойдаланувчи алгоритмлар маълумотлар назорати учун қўлай шароитлар яратиши, ҳамда ишончилиликни юқори даражада таъминлаши кўрсатилган.

Маълумот ва билимлар базаси (МБ ва ББ), МБни бошқариш тизими (МББТ) ва эксперт тизими асосида маълумотларни назорат қилиш дастурий-алгоритмик мажмуасини яратиш тамойиллари шакллантирилган.

Ишда, асосан ахборотларни узатиш ва ихчамлаштириш масалаларини ечишда қўлланиб келинган статистик ортиқчаликдан маълумотларни узатиш ишончилигини таъминлашда фойдаланиш механизмлари ишлаб чиқилган.

Тадқиқ қилинган матн элементлари тасвирларини назорат қилувчи алгоритмларни қўллашдаги хатолар аниқланмаслик эҳтимоли баҳосини берувчи умумлашган модел таклиф этилган:

$$P_H = P_1W_1 + P_2W_2,$$

бунда $W_1, W_2 - [0,1]$ хақиқий интервалида бериладиган бир ва иккинчи турдаги хатоликлар билан боғлиқ бўлган йўқотиш коэффициентлари.

Буни қўллашда назорат қилинадиган матн элементлари икки синфга бўлинади: ишончли ва бузилган маълумотлар.

Биринчи синфга тегишли катта бўлган N_1 сонли элементлар бўйича ўрта эҳтимол қуйидагича

$$x_1 = N_1 p_1 (1 - P_n) W_1 + N_1 (1 - p_1) P_n W_2,$$

бунда P_n – элемент бошқа синфга тегишли эканлиги эҳтимоли.

Иккинчи синфга тегишли N_2 элементлар бўйича ўрталаштириш:

$$x_2 = N_2 p_1 P_n W_1 + N_2 (1 - p_1) (1 - P_n) W_2.$$

Микдорлашган айирмаларни ҳисоблаш эса:

$$\frac{x_1}{N_1} - \frac{x_2}{N_2} = p_1 (1 - 2P_n) W_1 - (1 - p_1) (1 - 2P_n) W_2.$$

Ихтиёрий $P_n \neq 1/2$ учун $p_1 W_1$ ва $(1 - p_1) W_2$ йиғинди ва айирмалар баҳоланади. W_1 ва W_2 қийматлари қуйидаги ифодалар бўйича аниқланади:

$$W_1 = \frac{x_1 N_2 - x_2 N_1}{N_1 N_2 p_1 (1 - 2P_n)}; \quad W_2 = \frac{x_1 N_2 - x_2 N_1}{N_1 N_2 (1 - p_1) (1 - 2P_n)}.$$

Ишда Хаффман ва арифметик кодлаш, Барроуз-Вилер ўгириши, ҳамда Лемпел-Зив алгоритми, улар модификациялари бўйича маълумотни ихчамлаштириш механизмларга асосланган ахборотни назорат қилиш алгоритмлари ишлаб чиқилган.

Тадқиқ қилинган усулларда код комбинацияларини рухсат этилган тўпламининг эҳтимолли қуйи ва юқори чегараларини аниқлаш, ҳамда назорат қилинадиган маълумот элементларининг энг юқори эҳтимоллик баҳосини олиш муҳим масаладир.

Бу ҳолат тасвири сифатида қабул қилинган β га нисбатан узатилган α маълумотларнинг $\varpi(\alpha)$ ва $\varpi(\beta/\alpha)$ – нормал қонуният учун шартли эҳтимол тақсимот функцияларининг графиклари кўрсатилган.

Натижаларнинг амалий тадбиқи узатилган маълумот қийматининг назорат қилинувчидан четланиш чегараси қанча қисқа бўлса, шунча кўп биринчи турдаги хатоликларни икки ўлчамли регрессия усули билан аниқлаш мумкинлигини ёки назорат қилинувчи параметрлар боғлиқлиги қанча яқин бўлса, уларнинг қийматлари берилганларга шунча яқин бўлишини кўрсатади.

Маълумотларни назорат қилишда, шартли ҳолатнинг мақбул чегараларини аниқлаш билан боғлиқ бўлган масалаларнинг умумий ва хусусий ечимлари олинган ва бунда минимал хатолар аниқланмасликнинг эҳтимоли қуйидаги ифода бўйича ҳисобланади:

$$P_H = P_1 + P_2 = \frac{P}{B} \int_{x(\alpha)}^{y(\alpha)} W(\alpha) d\alpha + (1 - P) \left(1 - \int_{x(\alpha)}^{y(\alpha)} W(\alpha) d\alpha \right).$$

Бу ҳолда, ахборот узатиш ишончлилиги баҳосини олиш учун тақсимот функциялари $W(\alpha)$ ва $W(\beta)$ фарқлари чегараси $W(\beta/\alpha)$ шартли эҳтимол тақсимоти бўйича аниқланади.

Ишда ахборот ишончлилигини назорат қилиш масаласи маълумот узатувчи пакетдаги тасвир контейнерларида мавжуд бўладиган таркибий PR-ортиқчаликни қўллаш бўйича ҳам ечилган.

Қуйидаги белгиланишлар киритилди: $E(V, W) = V_W$ – тасвир контейнерига таркибий ортиқчаликни киритиш амали; $D(V, V_W) = W'$ (ёки $D(V_W) = W'$) – талаб қилинганда таркибий ортиқчаликдан қутилиш амали; α - масштабловчи коэффициент; C – икки сигнал ўхшашлигининг бинар белгиси: $C = 1$ агар ўзаро корреляция коэффициенти қандайдир δ чегарадан юқори; акс ҳолда $C = 0$.

Таркибий PR-ортиқчалик бўйича ахборот ишончилигини назорат қилиш алгоритмлари ишининг қуйидаги вариантлари асосланган.

Биринчи вариантда назорат қилиш алгоритми - $A_1(E, D, C)$ кескин қайтариловчи жараён хоссасига асосланади. Бунда, ихтиёрий V_W учун E^{-1} аксланиш мавжуд дейилади ва $E^{-1}(V_W) = (V_F, W_F)$ ва $E(V_F, W_F) = V_W$ шартлар бажарилиши керак, яъни E^{-1} ҳисобланадиган, W_F эса таркибий ортиқчаликга тегишли, чин тасвирни ёлғонидан ажратиш имконини берадиган ва $C(D(V_W, V_F), W_F, \delta) = 1$ кўринишдаги функция сифатида тасаввур қилинади.

Иккинчи вариантда алгоритм $A_2(E, D, C)$ иши сустр қайтариловчи жараён хоссасига асосланади. Бунда, $E(V_F, W_F) = V_W$ шarti талаб қилинади. Аммо, $E(V, W) = V_W$ шarti турли хил V_W каби ахборот бузилишлари туфайли бажарилмаслиги мумкин. Қуйилган талаб жараён қайтарилиши учун сустр ҳисобланади ва шу туфайли $E(V_F, W_F) = V_W$, кўринишдаги янги талабга алмаштирилади, бунда $C(V_{W'}, V_W, \delta) = 1$.

Алгоритм $A_2(E, D, C)$ ишида жараён сустр қайтарилишининг бошқа шarti ихтиёрий V_W учун E^{-1} аксланиш мавжудлиги, бунда $E^{-1}(V_W) = (V_F, W_F)$ ва $E(V_F, W_F) = V_W$, бўлади. Яъни E^{-1} ҳисобланиши мумкин бўлган, ва $C(V_{W'}, V_F, \delta) = 1$, $C(V_{W'}, V_W, \delta) = 1$, $C(D(V_{W'}, V_F), W_F, \delta) = 1$ шартлар бажарилади.

Таркибий-технологик PR-ортиқчалик, таркиблашган эксперт тизими, билимлар базаси, мантиқий назорат амаллари ва маълумот таркибларидан фойдаланиш асосидаги маълумотларни назорат қилиш алгоритмлари ишлаб чиқилган. Таснифлар, мезонлар ва қабул қилинган ЭХТ таркибида жорий этиладиган қарор қоидалари жойларини танлаш тамойиллари таклиф этилган. Шакллантирилган тамойил ва мезонлар асосида ЭХТда маълумотлар ишончилигини назорат қилиш тизимининг алгоритм ва дастурий модуллари қурилган.

Маълумотларни PR-ортиқчалик бўйича автоматик равишда назорат қилиш ва тиклашнинг амаллари ва воситаларини жорийлаштириш учун дастурий ортиқчаликдан ҳам фойдаланишга йўналтирилган комбинациялашган ёндошувлар таклиф этилди. Бундан ташқари, ишда ЭХТ шароитида ахборот ишончилигини назорат қилишнинг ташкилий-тезкор услублари синфлаштирилган.

Ҳар хил таъсирларга турғун бўлган ва реал шароитларда ҳам сифатли натижага эришишга қодир бўлган, ўн учта вазифавий ва ўн саккизта тадбиқий дастурий жорий этилувчи модуллардан иборат ахборот ишончилигини назорат қилувчи тизим объектга-йўналтирилган дастурлаш технологиясига асосан яратилган.

Ахборот ишончилигини назорат қилиш учун диссертациянинг назарий қисмларида ишлаб чиқилган усул, модел ва алгоритмлар самарадорлигининг қиёсий таҳлили келтирилган. Олинган натижа R^*_H тажрибавий қийматларининг таҳлиliga яқинлиги χ^2 – Пирсон мезони бўйича, 0,9 кафолатли эҳтимолда текширилди ва назарий натижаларнинг тажрибавий қийматига яхши мос келиши исбот қилинди.

Диссертациянинг **тўртинчи бобида** матнни n -граммли таркибий модели, матн элементларини статистик таниш ва таснифларини таниш моделлари бўйича матнли маълумотларни узатиш ишончилигини назорат қилувчи усул ва алгоритмлари ишлаб чиқилган.

Хатоларни назорат ва таҳрир қилувчи тизимларни яратишда асос бўлувчи, n -граммли статистик тақсимот қонуниятларини аниқлаш учун жараён моделлаштирилди ва матннинг бузилган элементларини аниқлаш, кластерлаш, қидириш, таркиблаштириш, ҳамда частотали таснифларни олувчи алгоритмлар ишлаб чиқилди.

Талқин қилинган излаш, таниш ва кластерлаш тамойил ва моделлари бўйича яратилган назорат қилиш тизими катта ҳажмдаги хотирани талаб қилиниши аниқланди. Бу камчиликларни бартараф этиш учун ва матн имлосини назорат қилиш масаласини ечишда GPU модели ва CUDA технологиясига асосланган параллел ҳисоблаш тизимларининг механизмидан фойдаланиш тавсия қилинди.

Талаб қилинган минимал хотира баҳоси қуйидагича ифодаланади

$$f(x) = -\frac{nx}{\ln(1 - P^{1/x})},$$

бунда n – код узунлиги,

x – хеш-функциялар сонини белгиловчи ўзгарувчидир.

Ишда луғатли, статистик ва хеш-кодлаш механизмлари; параллел ҳисоблаш усул, алгоритм ва амаллари имкониятларини қамровчи ахборот ишончилигини назорат қилиш тизимини яратишнинг янги тамойил, модел ва алгоритмлари таклиф қилинган ва асосланган.

Ахборот ишончилигини назорат қилувчи дастурий-техник мажмуани (ДТМ) яратишда NVIDIA видеочипида параллел ҳисоблашни қўлайлаштирувчи дастурий-аппаратли ечим қўлланилган. Параллел ҳисоблашлар асосида қурилган тизимнинг асосий хусусияти шундан иборатки, унда мантиқий оқимлар сони бажарувчи физик воситалар сонидан кўп бўлганида ечим берадиган алгоритмлар амалга оширилган.

Парсингли моделлаштириш асосида маълумотлар ишончилигини назорат қилиш алгоритмлари самарадорлигини баҳолашда сўзларни қатордан ажратиш ва таниш эҳтимолидан фойдаланиш таклиф этилган

$$W \doteq \arg \max_W P(W | A)P(A),$$

бунда W – матн элементлари $W = w_1, w_2, \dots, w_n$ тўплами кўринишидаги тадқиқ қилинувчи сўз;

$P(W / A)$ – A қаторда сўзнинг W кўринишда пайдо бўлиш шартли эҳтимоли;

$P(A)$ – матнда $A = \{a_1, a_2, \dots, a_n\}$ қатор пайдо бўлиши априор эҳтимоли.



1-расм. Матни парсингли таниш модулларининг ўзаро боғланиши

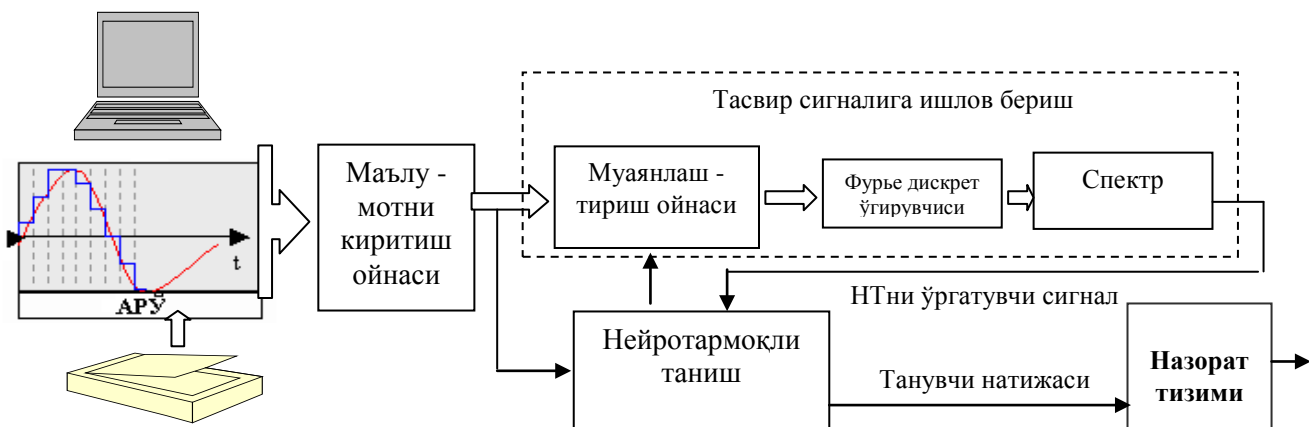
Парсингли моделлаштириш механизми бўйича имловий хатоларни назорат қилувчи тизимнинг иши дарахтсимон модел ва сўзлар кетмакетлигини парсингли кодлаш алгоритми асосида ташкил қилинган.

Уч модулдан иборат парсингли танувчи тизим фаолияти схемаси 1-расмда кўрсатилган. «Сўз башоратчиси» модули навбатдаги w_{k+1} сўзни парсинг-сўзнинг k -префиксга асосан башоратини амалга

оширади ва бошқарувни «Таггер» модулига узатади. «Таггер» навбатдаги сўзнинг t_{k+1} POS аломатини k -префикси ва охири w_{k+1} башоратланган сўз бўйича ҳисоблайди ва бошқарувни «Конструктор» модулига узатади. «Конструктор» мавжуд тузилмани, бошқарув «Сўз башоратчиси» модулига узатилгунча қайта ўтишни генерациялаш асосида кенгайтиради.

Ишда матн элементлари ишончлилигини назорат қилиш учун ассоциатив хотирали нейрон тармоғи (НТ) асосида танувчи алгоритм ишлаб чиқилган. Матн элементи тасвирлари ишончлилигини назорат қилиш учун сигналларни киритиш жараёни 2-расмда кўрсатилган.

Матн элементларини сўз-шакллар махсус луғати, монограмма, диграмма, учграмма, k -граммлар асосидаги частотали луғатлар билан солиштириш бўйича таниш ишончсизлиги эҳтимоли баҳосини берувчи услубият таклиф қилинган.



2-расм. Матн элементи сигнали ишончлилигини назорат қилиш

Берилган луғатларнинг жами ҳажми бўйича нотўғри таниш эҳтимоли қуйидагича аниқланади:

$$p(\tau) = 1 - (1 - p(a)) \prod_{i=1}^{n-1} (1 - ((1 - p(a)) + p(a \equiv a_i)(2p(a) - 1))), \quad p(a) < 10^{-3},$$

бунда τ – тимсолни таниш ишончилигини баҳоловчи махсус луғат ҳажми;

$p(a)$ – маълумотларни узатиш ва ишлов беришда матн элементи сифатида хато эҳтимоли;

$p(a \equiv a_i)$ – тўғри таниш эҳтимоли,

n – таниш учун бериладиган матн элементлари сони.

Тизимда ишончли матн элементи эталон сифатида қабул қилинади ва тўғри таниш баҳоси қуйидаги формула билан аниқланади:

$$p(\tau) = 1 - (1 - p(a)) \prod_{i=1}^{n-1} \prod_{j < i, a_j, a_i \in A} (p(a_j \equiv a_i)),$$

бунда $p(a_j \equiv a_i)$ – назорат қилиш тизимининг ишончилилик эҳтимоли.

Ўзбек тили метаматнига ишлов берувчи тизим умумий алгоритмини яратиш учун ихтиёрий тармоқланиш сонидagi дарахтни берувчи синфлар иерархиясини аниқлаш схемаси таклиф қилинган ва шу асосда назорат қилинувчи объектлар синфлаштирилиши амалга оширилган. Жорий этилган тизимда датслабки маълумотлар ББда сақланади ва ЭХТдан фойдаланиш жараёнида йиғилган янги билимлар бўйича қайта аниқланади. ББ муҳитида реал модел шаклланади ва сервер билан алоқа мижоз сўровлари тизими бўйича амалга оширилади.

Ишлаб чиқилган тизим қуйидаги функционал қисмлардан иборат: МБда объект ҳақида реал ва эталон маълумотларни сақловчи, маълумотларни солиштириш ва натижаларни сақловчи, объектнинг концептуал, инфологик ва физик моделлари; экспертлар билимларини, далиллар боғлиқлиги бўйича баҳоларни берувчи, боғланишни жадваллар бўйича ифодаловчи статик ББ; эталон динамик жараёнлар параметрларини қисман ва тўла ноаниқлигини ҳисобга олувчи кўринишидаги НТ моделларини ифодаловчи динамик ББ; моделнинг функционал таркибида воқеликларнинг сабаб-оқибатлар кетма-кетлигини шакллантирувчи алгоритмига асосланган мантиқий хулосалаш модули; юзага келган ҳолатга боғлиқ равишда МБ ва ББ иш фаолиятини мантиқий хулосалаш жараёнида мувофиқлаштирувчи мослаштириш модули; мантиқий хулосалаш жараёнини изоҳлаш амалларини бажарувчи тушинтириш модули; масала ечимини режалаштириш жараёнини мувофиқлаштирувчи бошқариш модули; тўғри, тескари ва аралаш масала қуйилишида самарали ечимни берувчи механизмлар.

Диссертациянинг **бешинчи бобида** ўзбек тили матнлари имлосини назорат ва таҳрир қилишга бўлган асосий ёндошувлари тадқиқ қилинган, ҳамда кўп босқичли морфологик таҳлил ва таркибий n-граммалардан фойдаланиш асосида имлони назорат ва таҳрир қилувчи модел, усул, алгоритмлари ишлаб чиқилган.

Адабиётлар таҳлили натижалари бўйича морфологик таҳлил асосида қурилган тизимларга хос кенг тарқалган қуйидаги усулларнинг тамойил, хусусиятлар ва камчиликлари аниқланди: матнни сўзларга ажратиш, луғатларга солиштириш, сўзлар негизини тақрибий аниқлаш, матннинг луғат билан «уйқашishi» ва сўзларни нормаллаштириш.

Юқори самарали ҳисобланган тизимлар сўз-шакллари икки босқичли моделлар асосида амалга оширилиши аниқланди, бу тамойил ўзбек тили имлосини назорат ва таҳрир қилувчи тизимни яратишда таянч механизми сифатида қабул қилинишга тавсия этилди.

Демак, икки босқичли модел ва сўз-шакллари таҳлил қилиш усуллари ва тамойиллари морфологик таҳлилни кўп босқичли модел асосида жорий этиш имкониятларини яратади ва ўзбек тили матнларидаги хатоларни назорат қилиш тизимини қуришда катта назарий ва амалий аҳамиятга эга.

Ишлаб чиқилган кўп босқичли график схемаси ўзбек тили сўз-шакллари қуришнинг ва аффиксларни негизга ва бошқа турдаги қисмларга кўшиш кетма-кетлигини акслантиради.

Ўзбек тили сўз шакллари ҳамда уларнинг фонетик ўзгаришларнинг формал тавсифлаш ва таҳлили услубияти ва алгоритмлари яратилган. Кўп босқичли жадвал сифатида ўзбек тили аффиксларининг боғланишларини кўрсатувчи, сўзни охиридан кўриб чиқишни бажарувчи, аффиксларни босқичлар бўйича рекурсив равишда ажратувчи услубиятлар ва имлони ишончли назорат қилувчи амаллар ишлаб чиқилган. Яратилган назарий ҳолатлар бўйича имлони назорат ва таҳрир қилиш тизимнинг алгоритмлари ва тузилмаси ишлаб чиқилган.

Ахборот таркибида n -граммли хатолар пайдо бўлиши модели асосида эса, уларнинг эҳтимолли баҳоларни олиш услубияти яратилган.

Хатоларнинг ўртача эҳтимоли хатолар шартли эҳтимолларини маълумотлар ансамбли бўйича ўрталаштириш асосида аниқланади, яъни:

$$P = \sum_{\alpha_i} P_{\alpha_i} \sum_{\alpha_j (\alpha_i \neq \alpha_j)} P_{\alpha_i \alpha_j},$$

бунда $P_{\alpha_i \alpha_j}$ - α_i маълумотни тўғри қабул қилиш эҳтимоли.

Маълумотлар таркибидаги n -граммли хатолар эҳтимолларини аниқлаш катта ҳажмдаги маълумотларга ишлов беришни талаб қилади ва мураккаб ҳисоблаш билан боғлиқлиги туфайли, n -граммли хатолар эҳтимолини ҳисоблаш учун статистикага ишлов беришни махсус моделлаштириш ва автоматлаштириш бўйича амалга оширилди.

Таклиф қилинган услубият бўйича $p(\alpha)$ n -граммли $\alpha = \alpha_1 \dots \alpha_n$ кетма-кетлигидан иборат хато эҳтимоли умуман учрайдиган n -граммалар сонини ҳар бир хатоликлар эҳтимоллари йиғиндисига бўлиниши билан аниқланади:

$$p(\alpha) = \frac{C(\alpha)}{\sum_{\alpha^*} C(\alpha^*)},$$

бунда α^* - α дан фарқли «кетма-кетликлар бош тўплами»га тегишли n -граммли нусха.

Қаторда α_j^* каби содир бўладиган хатоликларнинг тўла эҳтимоллиги $p(\alpha_i = \alpha_j^* | \alpha_i = \alpha_i^* \forall i \neq j)$ шартли эҳтимоллари бўйича аниқланади

$$p(\alpha_1^* \dots \alpha_n^*) = p(\alpha_i = \alpha_j^* | \alpha_i = \alpha_i^* \forall i \neq j) p(\alpha_i = \alpha_i^* \forall i \neq j).$$

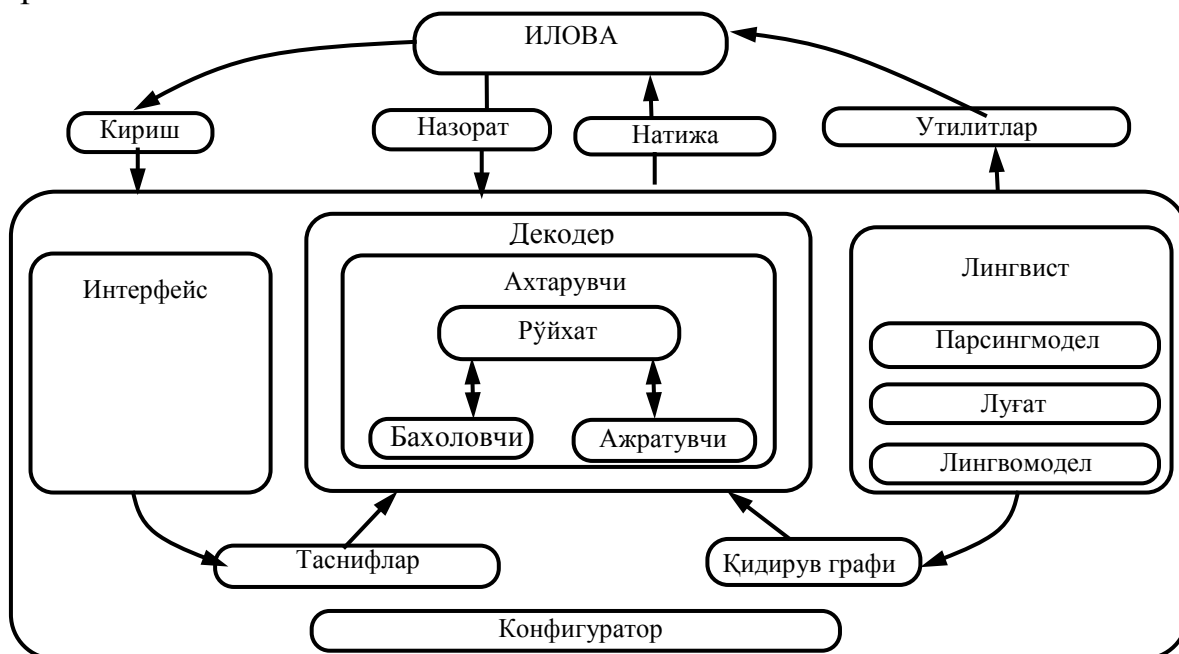
Бунинг учун, қаторда содир бўлган белги сифатидаги хатоликларнинг шартли эҳтимоллари олдинги белгиларни ҳисобга олиш бўйича қуйидаги қоидага асосан ҳисобланади:

$$p(\alpha_1 \dots \alpha_n) = \left(\prod_{k=2}^n p(\alpha_k | \alpha_1 \dots \alpha_{k-1}) \right) p(\alpha_1).$$

Марков занжири $(n-1)$ тартибда хатоликларнинг соддалаштирилган n -граммли модели олинди, бу эса, ўз навбатида, барча n -граммларнинг статистик параметрларини баҳолаш имкониятини берди.

Фреймворк Сфинкс-4 технологияси муҳитида имлони назорат қилиш тизимнинг такомиллаштирилган варианты ишлаб чиқилди. Бу, махсус масалалар ечимига чамаланган хусусий компонентларни қўллаш ва ихтиёрий n -граммани ҳосил қилувчи контекстга мослаштирувчи тизимнинг модулли тузилмани яратиш имкониятини берди. Тизимнинг қобиғи ва унинг фаолиятида барча компонента, ҳамда фреймворк дастурий модулларини қўлловчи тизим ядросини тавсифлаш усуллари ишлаб чиқилди.

Фреймворк Сфинкс-4 асосида ишлаб чиқилган ва Интерфейс, Декодер, Лингвист сифатида номланган асосий, ҳамда Конфигурация ва Утилитлар номидаги қўшимча блоклардан иборат тизим архитектураси 3-расмда берилган.



3-расм. Сфинкс-4 асосида имлони назорат қилиш тизим архитектураси

Схемада блоклар орасидаги ва блокларнинг иловалар билан бўлган алоқалари ҳам тасвирланган. Ишда тизимнинг асосий ва қўшимча Интерфейс, Декодер ва Лингвист блокларининг фреймворк Сфинкс-4 муҳитидаги вазифалари келтирилган.

Интерфейс блоқи вазифаси бир ёки ундан кўп матн элементларини (белги, ҳарф, сўз) тизимга киритиш учун қабул қилиш ва уларни аломатлар кетма-кетлиги сифатида параметрлашдан иборат.

Лингвист вазифаси стандарт тил моделини n-граммли таркибдан парсинглаштиришга ўгириш, Қидирув графини тузиш ва унда бир Парсинг-модел ҳамда ундан кўп бўлган тўпламни шакллантиришдан иборат.

Декодер таркибидаги Қидирув модули, Интерфейс ҳамда Лингвист томонидан генерацияланган Қидирувчи-графлардан олинган аломатларни натижаларни дешифрлаш ва генерация қилиш учун ишлатади.

Ўтказилган тест тажрибалари натижалари, таклиф қилинган вариантда яратилган тизим сўзни танишда нисбий хато кўрсаткичи яхшиланишини, ва тизим иши тезлиги реал вақтда морфологик таҳлил асосида имлони назорат қилувчи тизимга нисбатан икки марта оширилишини кўрсатди. Тизим иши икки процессорли режимда CPU 1015MHZ ULTRASPARC RIII WITH 2G RAM процессорларда ташкил қилинганини таъкидлаймиз. Сфинкс-4 қобиғига кирувчи модуллардан шу асосда, ҳамда мослашувчан алгоритмлар ва тилнинг статистик моделларидан фойдаланишлик туфайли, кенг тартибда қўлланиладиган матн таҳлили, матнни нутқга ва нутқни матнга генерациялаш ва бошқа матнли маълумотларга ишлов бериш каби масалаларни ечиш имконияти яратилади.

Маълумотлар ишончлилигини n-граммли модел асосида назорат қилиш тизими ишлов бериладиган оқимларнинг мақбул тартибини ва юкланишини ташкил қилиш учун маълумотларга статистик ишлов бериш усулларида фойдаланади, n-граммали грамматика ва дарахтсимон моделни қўллайди ва натижада ишлов бериладиган файллар ўлчамлари форматлар таркибидаги ортиқчаликларни йўқотиш бўйича қисқартириш имконини беради.

Ишда ортиқчаликни ихчамлаштириш мақсадида грамматика элементлари билан амаллар манипуляциясининг схемаси дарахтсимон тасвирда кўрсатилган.

Грамматиканинг n-граммасини тавсифлаш, турли тилдаги дастурлашни қувватлаш ва ахборотларга тармоқли ишлов бериш учун махсус формат қурилди ва бу муҳитда n-граммали моделлар тавсифи XML грамматикасида амалга оширилди.

Диссертациянинг **олтинчи бобида** ахборот ишончлилигини назорат қилиш тизими учун синфлар, усул ва дастурий модулларни PR-ортиқчаликдан фойдаланиш асосида ишлаб чиқиш ва уларни маълумотларга тармоқли ишлов бериш тизимларининг ахборот узатувчи пакетлар таркибида, ҳамда ЭХТда жорий этиш бўйича маълумотлар келтирилган.

Электрон хужжат айланиш тизимининг иш юритиш элементларини яратиш, сақлаш ва узатиш вазифаларини амалга оширадиган дастурий-инструментал воситалар (ДИВ) мажмуаси 1С 8.2 интеграллашган муҳитда ишлаб чиқилган. Таъкидланган асосда, ишлаб чиқилган тизимни жорий этиш хусусияти маълумотни тайёрлаш, киритишда, трансформациялаш туфайли содир бўладиган маълумотлардаги хатоларни назорат қилувчи дастурий мажмуа тарзидаги амалга оширилган CONTROL қисми ўз таркибига

олганлигидан иборат.

Ишлаб чиқилган дастурий мажмуа ахборот тизимининг вазифавий масалаларини ечишга чамаланган қисмлари таркибига киради ва талаб қилинганда, дастурий маҳсулотлар сифатида берилади.

Ишлаб чиқилган мажмуани матнли маълумотларга ишлов берувчи ихтиёрий компьютер тизимларини сервисли қўлловчи алоҳида қисм сифатида ишлатиш мумкин, ЭХТ вазифавий қисмлари таркибига киради, OSI моделининг Тадбиқий ва Тақдимот босқичларида содир бўладиган хатоларни замонавий телекоммуникация тизимларида узатиладиган протоколлар таркибида назорат қилишда қўлланилади.

Тадқиқ қилинган ЭХТда объектли-йўналтирилган дастурлаш усуллари, МБ ва ББдан самарали фойдаланиш имконини берадиган қайд қилинувчи карточкалар реквизитлари бўйича ҳужжатларни кенг кўламда қидириш; ходимлар лавозимлари, стандарт ибора ва сўз алмашуви маълумотномаларини юритиш; маълумотномани узвий таҳрирлаш; ҳужжатлар, қайд қилиш картотекаси, шахсий ишлар, мавзу рубрикалари ва бошқа мезонлар бўйича бажарувчи амалларни тизимга бириктиришда ходимларга рухсат бериш ҳуқуқини мавқелаштириш вазифаларини амалга оширувчи назорат қилиш объектлари ажрати қаралган.

CUDA NVIDIA технологияси асосида, Сфинкс-4 муҳитида имловий хатоларни назорат қилиш ва таҳрирлаш дастурий тизимнинг қобиғи ишлаб чиқилган. Бунга қуйидаги функцияларни жорий этишга қаратилган модуллар киради: модул Ф1 – сўзни киритади ва уни виртуал жадвалдан ўтказди; модул Ф2 – луғатда бинарли излаш услубини жорийлаштиради; модул Ф3 – луғатни сақлаш ва юклаш учун массивни шакллантиради; модул Ф4 – тезкор саралашни жорий этади; модул Ф5 – объектлар алмашишини амалга оширади; модул Ф6 – сўзлар чегарасини текширади; модул Ф7 – юкланган луғатда бинарли излашни амалга оширади; модул Ф8 – орфографик хатоларни назорат қилади; Ф9-Ф15 модуллари – орфографияни назорат қилиш муҳитида матнни транслитерациялашни амалга оширади; Ф16-Ф20 модуллари – киритилган матнни парсинглаш услубини жорий этади ва парсингли моделлаштиришни амалга оширади; модул Ф21 – истеъмолчига ҳужжатлар билан ишлаш ва турли шрифт хоссаларини таҳрир қилиш, қўшиш, матнни таҳрир қилиш ва бекор қилиш, ҳамда файл сифатида хотирага олиш имкониятларини беради; модуль Ф22 – матнли процессор бўйича таҳрирловчи ва ўзбек тилидаги матнни форматлайдиган MDI иловаси; модул Ф23 – иловалар ёпилганда, барча ҳужжатларнинг тўғри ёпилишини таъминлайди.

Орфографик хатоларни таҳрир қилиш вазифасини Хаффман мослашувчан кодери бўйича амалга оширувчи модуллар жорий этилган. Шу қаторда, қайтарилувчи тикланиш тамойилларини қўллашда Зив-Лемпел қоидаси бўйича кодлаш ва ихчамлаштириш модел ва алгоритмлари амалга оширилган. Лемпел-Зив-Велч статистик кодлашнинг тамойили ва қоидаси бўйича матн қаторларини сақлайдиган ва тартибланишган катакларга эга луғатни қўллашга асосланган такомиллаштирилган модул жорий этилган.

Арифметик кодлаштиришнинг оддий, ярим мослашувчан ва мослашувчан моделлари асосида ахборот таркибидаги имловий хатоларни назорат қилувчи дастурий модуль яратилган.

Ўзбек тили матнлари имлосини назорат қилувчи тизимнинг асосий синфлари, фаолиятини ташкил қилиш услублари ва умумий таркиби аниқланди. Тизим таркиблари бўйича дастурий маҳсулотларни қайд қилиш гувоҳномалари олинган. Ўзбек «spell-checker» имлони икки кўринишдаги алфавитда назорат қилишга қодир ва талаб бўйича бир алфавит туридан бошқасига алмаштиради.

Матнли маълумотлар ишончлилигини назорат қилиш тизими ва ишлаб чиқилган алгоритмлари тажрибавий текширувдан ўтказилди ва натижалари СамДУда масофадан ўқитиш тизими ва «Лаззат-Мева» ЭХТ маълумотлари бўйича таҳлил қилинди. Тадқиқот натижалари, амалиётга жуда яқин бўлган, қуйидаги шароитларда олинди: киритиладиган маълумотлар ҳажми 10^6 белги (300 электрон матнли саҳифа); хатоликлар эҳтимоли қиймати - $P = 3,7 \cdot 10^{-3}$; ҳужжатдаги сўз узунлиги ўртача 8 ўнли белгига тенг; PR-ортиқчаликларни қўллашга асосланган алгоритмлар бўйича ҳужжатлардаги сўзлар ҳимоя қилинган; электрон воситалар узилиш ва носозликлари билан боғлиқ хатолар ҳисоблашда инобатга олинмайди.

Жадвалда турли мезонлар бўйича ахборот ишончлилигини назорат қилувчи алгоритмлар самарадорлигининг қиёсий таҳлили келтирилган. Жадвал 6,7 устунларида хато аниқланмаслиги эҳтимоли пасайиши коэффициенти (ахборот ишончлилигини таъминловчи ютуқ), яъни $K_H = P/P_H$ – назарий ва $K_H^* = P/P_H^*$ – тажрибавий қийматлари келтирилган.

Жадвал 8,9 устунларида назорат мураккаблиги - $K_{ПА} = T_P/T_{ПА}$, $K_{ТА} = T_P/T_A$, ҳамда 10, 11 устунларда жорий этиш харажатлари - $K_{СПА} = C_P/C_{ПА}$, $K_{СА} = C_P/C_A$ коэффицент қийматлари келтирилган.

Маълумотларга ишлов бериш объектлари хусусиятига асосан тизимни жорий қилиш шароитлари: маълумотни клавиатурали киритиш мураккаблиги – 400 соат, киритиш харажати – 60 минг сўм, хатони визуал назорат қилиш мураккаблиги – 350 соат, хатони визуал назорат қилиш харажатлари – 45 минг сўм, сканерлаш ва таниш мураккаблиги – 150 соат, харажати – 25 минг сўм миқдориди қабул қилинган.

Чизиқли йиғиндиларга асосланган алгоритмни қўллашда минимал хато аниқланмаслик эҳтимоли - $P_H \approx 3 \cdot 10^{-5}$; назорат мураккаблиги – 0,5 с.; назорат харажати – 1 минг сўмга баробар.

Текислик бўйича йиғиндига асосланган алгоритм минимал хато аниқланмаслик эҳтимолини - $P_H \approx 4 \cdot 10^{-6}$; назорат мураккаблигини – 1,2 с.; назорат харажатини – 1,5-2 минг сўм бўлишини таъминлайди.

Модулли йиғиндига асосланган алгоритм минимал хато аниқланмаслик эҳтимолини - $P_H \approx 6 \cdot 10^{-5}$; назорат мураккаблигини – 0,8 с.; назорат харажатини – 1,2 минг сўм бўлишини таъминлайди.

Ахборот ишончилигини назорат қилиш алгоритмларининг самарадорлиги таҳлили

№	Хатоларни назорат қилиш, аниқлаш ва тузатиш алгоритмлари	Самарадорлик мезонларининг қийматлари							
		Эҳтимоллик P_H		Коэффициент $K_H = P/P_H$		Назорат мураккаблиги - K_T		Назорат харажати - K_C	
		Назарий - P_H	Тажриб. P_H^*	Назарий K_H	Тажриб. K_H^*	$K_{ТПА}$	$K_{ТА}$	$K_{СПА}$	$K_{СА}$
Сунъий ортиқчаликдан фойдаланиш									
1	Чизиқли йиғиндилар бўйича	10^{-5}	$2 \cdot 10^{-5}$	100	200	1,8	1,7	2	2
2	Модулли йиғиндилар бўйича	$10^{-5}-10^{-6}$	$5 \cdot 10^{-6}$	1000	5000	2	2	2	2
Табиий ортиқчаликдан фойдаланиш									
3	n-граммли моделлари бўйича	$10^{-5}-10^{-6}$	$4 \cdot 10^{-6}$	1000	4000	2	2	3	3
4	Морфологик модел бўйича	10^{-6}	$2 \cdot 10^{-6}$	1000	2000	3	3	3	3
Таркибий-технологик ортиқчаликдан фойдаланиш									
5	Эксперт тизимари ва билимлар базаси асосида назорат қилиш алгоритмлари	$10^{-5}-10^{-6}$	$1.5 \cdot 10^{-6}$	1000	1500	1,8	1,7	2	2
6	Мантиқий назорат мезонлари асосидаги алгоритмлар	10^{-5}	$2 \cdot 10^{-5}$	100	200	2	2	2	2
Статистик ортиқчаликдан фойдаланиш									
7	Лемпел-Зив-Велч усули ва Барроуз-Вилер ўгириши асосидаги алгоритмлар	$2 \cdot 10^{-5}-10^{-6}$	$6 \cdot 10^{-6}$	2000	6000	3	3	3	3
8	Арифметик кодлаш асосидаги алгоритмлар	$3 \cdot 10^{-6}-10^{-7}$	10^{-7}	3000	10000	3	3	3	3

Ишлаб чиқилган усуллар мавжуд усулларга нисбатан назорат мураккаблиги ва харажати 2-3 маротаба камайтириши аниқланди. Ахборот тизими ишида визуал назорат мураккаблиги ва харажатлари тизимнинг ярим автоматлаштирилган ва тўла автоматлаштирилган шароитидаги мураккаблик ва харажатлари нисбатлари олинди ва қабул қилинган мезон коэффициентларининг 2-3 маротаба пасайишини кўрсатди. Тажрибавий - P_H^* нинг қиймати унинг назарий қиймати билан χ^2 – Пирсон мезонининг 0,9 кафолатли эҳтимолида текширилди.

Иш ҳужжатларини автоматлаштириш компьютер тизими кўринишидаги ЭХТ ва маълумотларни адаптив узатиш, ишлов бериш ва таҳлил қилиш ахборот-коммуникация тармоқларида яратилган усул, алгоритм ва дастурий воситаларни жорий этишнинг иқтисодий самарадорлигини ҳисоблаш услубияти ишлаб чиқилди. Ахборот коммуникация тармоқларида «УзТелеком» Самарқанд филиали, ҳамда «Лаззат-Мева» қўшма корхонасининг маълумотларни адаптив узатиш, ишлов бериш ва таҳлил қилиш компьютер тизимларининг иқтисодий самарадорлиги ҳар ойлик маълумотларга ишлов бериш ҳисоби бўйича 5,1 млн сўмни ташкил қилди.

ХУЛОСА

1. Диссертацияда тизимли таҳлил, бошқарув ва маълумотларга ишлов бериш концепцияси асосида PR-ортиқчаликни қўллаш тамойиллари бўйича ишлаб чиқилган ахборот ишончилигини таъминловчи конструктив усул, модел ва дастурий-алгоритмик мажмалари электрон ҳужжат айланиши тизимлари самарадорлиги ва унумдорлигини оширишни таъминлайди.

2. Ахборот ишончилигини назорат қилишда кодли, аппаратурали ва дастурий усуллари назария ва амалиётининг ҳозирги ҳолатини баҳолаш натижалари ЭХТ сифатли фаолиятини таъминлаш учун мавжуд ортиқчалик турларидан фойдаланиш етарли эмаслигини кўрсатиб берди. Ахборотни узатиш ишончилигини таъминловчи амалий усул ва дастурий-алгоритмик мажмуаларини ишлаб чиқишда электрон ҳужжатлардаги PR-ортиқчаликдан фойдаланиш тамойиллари қўлланаётган технологияларни ривожлантириш воситаларини лойиҳалаш имконини яратади.

3. Чизикли, текислик ва модулар бўйича йиғинди ҳосил қилувчи, Хаффман, Лемпел-Зив-Велч, Барроуз-Вилер қоидалари бўйича ва арифметик кодлаш, статистик таниш, мантиқий назорат усуллари электрон ҳужжатлари матнларини тузиш, узатиш ва уларга ишлов беришда маълумотлар ишончилигини таъминлашнинг алгоритм ва дастурий мажмуалари имкониятларини ошириш учун PR-ортиқчаликдан фойдаланишнинг услубий асосларини ташкил қилади.

4. Ишлаб чиқилган алгоритмлар самарадорлиги намойиши бўйича улар 92% хатоларни аниқлайди ва бир, икки карралаи ва қўшма транспозицияли хатоларни тузатишга қодир; бундан ташқари, мавжуд услубларга нисбатан назорат мураккаблиги ва жорий этиш харажатларини $P_H \approx 4 \cdot 10^{-3}$ шароитида 2-3 маротаба камайтиради; маълумотлар ишончилигини эса, уч

бароваригача оширади. Ишда самарадорлик баҳолари ва маълумот ишончилигини назорат қилиш масалаларининг аналитик ечимларини хато аниқланмаслиги эҳтимоли мезони бўйича олишга эришилади.

5. Морфологик ва n-граммли таркиблашган моделлар асосида матнли ахборотни узатишда имло хатоларини назорат ва таҳрир қилувчи усул, алгоритм ва тизимлар ишлаб чиқилган. Хатоликларнинг эҳтимолли кўрсаткичлари статистикаси асосида ишлаб чиқилган n-граммаларнинг частотали таснифларни олиш услубияти хеш-кодларни парсингли кодлаш учун тизимлаштиришда қўлланилади.

6. Статистик танишда мантиқий ва арифметик функцияларни курувчи интерполяция ва экстраполяция усуллари матн элементлари тасвирларини назорат қилиш алгоритмларини ишлаб чиқишда қўлланилган. Автоматик таниш ва тасвир сигналларни назорат қилиш қисмларидан ташкил топган маълумотларга нейротармоқли ишлов берувчи тизимида матн элементи тасвирларининг сигнал таснифлари ишончилигини назорат қилиш усуллари таклиф этилган. Усул ва алгоритмлар нотиниқ семантик гипертармоқ бўйича ЭХТ метаматн объектларини синфлаштириш ва таснифлар тегишлилигини аниқлаш асосида билан ахборот ишончилигини назорат қилиш тизими таркибида жорий этилади.

7. Луғат, статистик ва хеш-кодлаш усуллари асослаган маълумотлар ишончилигини назорат қилиш модел ва алгоритмлари ишлаб чиқилган ва улар параллел ҳисоблаш NVIDIA дастурий-аппарат муҳитида сонли таҳлил стандарт кутубхоналари ҳамда CPU ва GPU орасида мақбул маълумотлар алмашувини берувчи воситаларни қўллаш самарадорлигини таъминлайди.

8. Фреймворк Сфинкс-4 муҳитида n-граммли грамматика тавсифини берувчи дарахтсимон модел бўйича имлони назорат қилиш тизим қобиғининг тавсифланиши идентификацияланган ва архитектураси ишлаб чиқилган ҳамда PR-ортиқчаликдан фойдаланадиган турли тил моделлари учун тизим дастурий модуллари жорий этилган. Тизимнинг яратилган варианты морфологик таҳлилга асосланган имлони назорат қилувчи тизимга нисбатан аниқланмаган хатолар сонини, жорий этиш ҳаражатларини ҳамда назорат мураккаблигини икки мартаба камайтиришга эришилади.

9. Ишлаб чиқилган моделлаштирувчи алгоритмлар, дастурий воситалар мажмуалари ва маълумотлар ишончилигини PR-ортиқчалик асосида назорат қилувчи тизимлар ОЎЮнинг автоматлаштирилган ўқув жараёнида, ахборот-коммуникация тармоқларида маълумотларни адаптив узатиш, ишлов бериш ва таҳлил қилишда, шунингдек корхоналар ЭХТ амалиётида қўлланилади.

10. Ишлаб чиқилган электрон хужжат айланиш тузилмасида маълумотлар ишончилигини назорат қилиши дастурий мажмуалари ҳамда ахборот-коммуникация тармоқларида маълумотларни узатиш, қайта ишлаш ва таҳлил қилиш тизимлари Алоқа, ахборотлаштириш ва телекоммуникация технологиялари Давлат Қўмитасининг «УзТелеком» Самарқанд филилали ва «Лаззат-Мева» қўшма корхонасида татбиқ этилган, диссертация натижалари бўйича олинган далолатномалар иқтисодий самарадорликни тасдиқлайди.

**НАУЧНЫЙ СОВЕТ 16.07.2013.Т/ФМ.29.01 при ТАШКЕНТСКОМ
УНИВЕРСИТЕТЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ и
НАЦИОНАЛЬНОМ УНИВЕРСИТЕТЕ УЗБЕКИСТАНА ПО
ПРИСУЖДЕНИЮ УЧЕНОЙ СТЕПЕНИ ДОКТОРА НАУК**

**ГОСУДАРСТВЕННЫЙ КОМИТЕТ СВЯЗИ, ИНФОРМАТИЗАЦИИ И
ТЕЛЕКОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ
РЕСПУБЛИКИ УЗБЕКИСТАН**

ТАШКЕНТСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

АХАТОВ АКМАЛ РУСТАМОВИЧ

**МЕТОДЫ И ПРОГРАММНО-АЛГОРИТМИЧЕСКИЕ КОМПЛЕКСЫ
ОБЕСПЕЧЕНИЯ ДОСТОВЕРНОСТИ ИНФОРМАЦИИ В СИСТЕМАХ
ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА НА ОСНОВЕ PR-
ИЗБЫТОЧНОСТИ**

**05.13.01 – Системный анализ, управление и обработка информации
(технические науки)**

АВТОРЕФЕРАТ ДОКТОРСКОЙ ДИССЕРТАЦИИ

Ташкент – 2014

Тема докторской диссертации зарегистрирована за № 20.02.2014/В2013.1.Т1 в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан.

Докторская диссертация выполнена в Самаркандском государственном университете.

Полный текст докторской диссертации размещен на веб-странице Научного совета по присуждению учёной степени доктора наук 16.07.2013.Т/ФМ.29.01 при Ташкентском университете информационных технологий и Национальном университете Узбекистана по адресу www.tuit.uz.

Автореферат диссертации на трех языках (узбекский, русский, английский) размещен на веб-странице по адресу www.tuit.uz и Информационно-образовательном портале “ZIYONET” по адресу www.ziyonet.uz

Научный консультант:

Жуманов Исраил Ибрагимович
доктор технических наук, профессор

Официальные оппоненты:

Набиев Озод Маликович
доктор технических наук, профессор

Нишанов Акрам Хасанович
доктор технических наук, профессор

Зайнидинов Хакимжон Насридинович
доктор технических наук, профессор

Ведущая организация:

Ташкентский государственный технический университет

Защита диссертации состоится «28» июня 2014 г. в 10⁰⁰ часов на заседании научного совета 16.07.2013.Т/ФМ.29.01 при Ташкентском университете информационных технологий и Национальном университете Узбекистана. (Адрес: 100202, Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-43; факс: (99871) 238-65-52; e-mail: tuit@tuit.uz).

С докторской диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского университета информационных технологий (регистрационный номер 01). Адрес: 100202, Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-65-44.

Автореферат диссертации разослан « » мая 2014 года.
(протокол рассылки № 1 от « » мая 2014 г.).

Х.К.Арипов

Председатель научного совета по присуждению учёной степени доктора наук д.т.н., профессор

М.С.Якубов

Ученый секретарь научного совета по присуждению учёной степени доктора наук д.т.н., профессор

Т.Ф.Бекмуратов

Председатель научного семинара при Научном совете по присуждению учёной степени доктора наук д.т.н., академик

АННОТАЦИЯ ДОКТОРСКОЙ ДИССЕРТАЦИИ

Актуальность и востребованность темы диссертации. В мировой практике информатизации, системы электронного документооборота (СЭД) в последние годы рассматриваются и внедряются не только как системы автоматизации процессов управления, но и как полноценные платформы для создания единого информационного пространства, что, безусловно, расширяет границы их использования и повышает к ним интерес со стороны научного сообщества. В исследованиях ведущих ученых сферы инфокоммуникационных технологий выделяется востребованность задач достоверного сбора, передачи, анализа, кодирования информации при формировании документов делопроизводства для получения эффективных технологий, повышающих мобильность и производительность СЭД.

Комплексные меры, предпринимаемые Правительством Республики Узбекистан по развитию систем регионально-территориального автоматизированного управления и созданию единого информационного пространства, направлены на широкое внедрение информационных систем, СЭД, баз данных (БД) на основе современных информационно-коммуникационных технологий. В связи с этим, разработка новых методов интеллектуальной обработки информационных ресурсов, используемых для повышения качества передачи информации, позволяющих эффективно обнаруживать и исправлять ошибки в структуре электронного документооборота с наименьшими материальными и временными затратами, отличается особой актуальностью и, вместе с тем, остается нерешенной теоретической и прикладной проблемой, имеющей важное народно-хозяйственное значение.

Требования, предъявляемые к информационным ресурсам и потокам передачи данных, как важного фактора эффективности и качества функционирования СЭД выражаются в обеспечении стабильности, целостности, сохранности и достоверности информации. Одним из важных среди факторов является критерий достоверности информации, обуславливающийся искажением передаваемых сообщений в инфокоммуникационных сетях из-за сбоев и отказов технических средств, различного рода помех в каналах связи, ошибок операторов и систем сканирования, а также распознавания.

Следовательно, особый научный интерес представляет построение эффективных систем контроля достоверности передачи и обработки информации, как приоритетной технологии обработки данных, характерной для условий автоматизированного управления и электронного документооборота на предприятиях и в организациях.

Для существующих методов, несмотря на обеспечение высокого уровня достоверности передачи информации, характерен ряд нерешенных вопросов, в качестве основных из которых можно выделить следующие: при разработке технологий электронного документооборота в структуре пакетов передачи данных значительный объем ресурса расходуется на заголовки, вместе с тем

большая часть информации заголовков остается неизменной на протяжении потока передачи целого кадра (возникающая при этом избыточность информации и механизмы обеспечения надежности доставки, заключающиеся, в основном, в послании соответствующего сообщения и повторной передаче пакетов, - приводят при обнаружении и устранении ошибок к дополнительным временным и материальным затратам); кодовые и аппаратные методы контроля передачи информации ориентированы, в основном, на устранение в управляющих полях пакетов одно- и двукратных транспозиционных ошибок, однако, при передаче данных возникают искажения в информационных полях, которые зачастую проявляются в виде текстовых ошибок более высокой кратности.

В этой связи, решение задач обеспечения достоверности передачи информации целесообразно рассматривать в двух аспектах.

При решении первого типа задач следует учитывать ошибки, вносимые оператором, сканирующими и другими устройствами, предназначенными для ввода информации. Такого вида ошибки составляют наибольший объем искажений текстовой информации и возникают на Прикладном и Представительском уровнях модели OSI (Open System Interconnection reference model).

Во втором типе задач контроля достоверности информации дополнительно учитываются вероятности искажений, возникающих на Транспортном, Сетевом, Физическом и Канальном уровнях OSI.

Востребованность диссертации характеризуется тем, что широкое внедрение IP-технологии в электронном документообороте связано с требующими пристального внимания задачами по обнаружению и устранению ошибок при подготовке и оформлении документов.

Данная исследовательская работа ориентирована на обеспечение реализации законов Республики Узбекистан “Об информатизации”, “Об электронной цифровой подписи”, “Об электронном документообороте”, “Об электронной коммерции”, “Об электронных платежах”, а также Постановления Кабинета Министров Республики Узбекистан № 126 от 04.05.2011 года «О мерах по внедрению и использованию единой защищенной электронной почты и системы электронного документооборота в исполнительном аппарате кабинета министров, органах государственного и хозяйственного управления, государственной власти на местах».

Исходя из этого, решение перечисленных задач требует проведения специальных исследований и разработок, связанных с созданием методов и алгоритмов, способных осуществлять контроль достоверности информации за счет использования вложенной избыточности, функционировать в сформированной транспортной среде и устранять существующие недостатки. Данное обстоятельство обуславливает необходимость выделения специального класса алгоритмов для обеспечения достоверности информации на основе нового типа PR-избыточности (property redundancy), определяющейся в зависимости от свойств обрабатываемого объекта.

Соответствие исследования приоритетным направлениям развития науки и технологий Республики Узбекистан. Диссертация выполнена в

соответствии с приоритетными направлениями развития науки и технологии: НТП-17 – «Разработка современных информационных систем, интеллектуальных средств управления и обучения, научно-технических БД и программных продуктов, обеспечивающих широкое развитие и внедрение информационных и телекоммуникационных технологий»; ППИ-5 – «Разработка информационных технологий, телекоммуникационных сетей, аппаратно-программных средств, методов и систем интеллектуального управления и обучения, направленных на повышение уровня информатизации общества».

Обзор международных научных исследований по теме диссертации.

Исследования по разработке средств автоматизированного контроля достоверности передачи и обработки информации с учетом особенностей естественного языка интенсивно ведутся и достигнуты определенные успехи учеными стран США, Великобритании, России, Китая, Германии, Австрии, Швеции, Норвегии, Финляндии, Турции, Израиля, Канады, Кореи, стран Латинской Америки (Бразилия, Аргентина, Чили) и др.

В публикациях международных экспертных компаний, таких как Gartner, International Data Corporation, Wobot, Forrester, CNews Analytics, ЮНИПРАВЭКС, Прогноз, Интерфейс ИТ, Интеллект Телеком и др. отмечается, что развитие технологий СЭД происходит в направлении интеграции методов обработки документов, разветвленного управления через WEB и автоматизации офисных процессов. Наряду с этим, выделяется актуальность и востребованность присоединения к ним ранее независимых технологий автоматизированного обнаружения и исправления ошибок ввода, передачи и обработки информации в электронных документах, контроля и коррекции орфографии естественных языков, как инструментария снижения влияния человеческого фактора.

Вместе с тем, по мере увеличения потребностей пользователей сетевых услуг активно выполняются исследования, посвященные разработке новых направлений и приложений, основанных на технологии передачи данных. Аналитическое изучение научно-технической литературы показало, что несомненным достоинством электронного документооборота, наряду с применением электронной цифровой подписи или технологий, гарантирующих подлинность и целостность данных, является возможность обнаружения и исправления ошибок, допускаемых при составлении, вводе, передаче и обработке электронных документов за счет методов автоматизированного контроля информации, снижающих влияние человеческого фактора.

Степень изученности проблемы. Методологию исследований по проблематике повышения достоверности передачи и обработки информации составляют методы теории кодирования и аппаратного использования искусственной избыточности информации, отраженные в большом количестве монографических работ, диссертационных исследований и научных статей. В то же время существенной особенностью технологий электронного документооборота является использование документов и

сообщений, содержащих искусственную, естественную, статистическую и структурно-технологическую избыточность.

Результаты исследований по новому направлению решения проблемы использования PR-избыточности различной природы предполагают разработку широкого спектра методов, алгоритмов и программных комплексов обеспечения достоверности передачи и обработки информационных ресурсов СЭД, научно-методические основы которых предназначены для развития существующих технологий.

Связь диссертационного исследования с планами научно-исследовательских работ отражена в следующих проектах:

государственный научно-технический проект (ГНТП)-4 – «Разработка программных средств и методов информатизации, систем автоматизации управления и регулирования технологических процессов» (2000-2002 гг.);

ГНТП-20 – «Разработка современных информационных систем, средств управления, баз данных и программных продуктов, обеспечивающих широкое развитие информационных и телекоммуникационных технологий» (2003-2005 гг.).

Цель исследования разработка конструктивных методов, моделей, алгоритмов и систем контроля достоверности информации при передаче и обработке данных на основе механизмов использования PR-избыточности, а также их программно-алгоритмическая реализация при разработке технологий электронного документооборота.

Для достижения цели сформулированы следующие **задачи исследования**:

выбор методологии создания и развития методов, программно-алгоритмических комплексов и систем контроля достоверности информации на прикладных и сетевых уровнях модели OSI за счет использования информационной PR-избыточности различной природы;

разработка методов и алгоритмов контроля достоверности информации за счет использования искусственной, естественной и семантической PR-избыточности на основе механизмов суммирования, арифметического и статистического кодирования;

создание и применение механизмов парсингового кодирования, n-граммного описания, статистического распознавания и хеш-кодирования элементов текстов для разработки методов и алгоритмов контроля достоверности информации;

разработка методов, алгоритмов и программных комплексов контроля достоверности текстов на основе моделей многоуровневого морфологического анализа и n-граммного структурного представления ошибок; моделирование основных компонентов системы контроля и коррекции орфографии; формализация оценки качества контроля;

формирование и реализация структурно-технологической PR-избыточности на основе вложенных логических критериев, баз данных и знаний в составе встроенных экспертных систем при разработке методов и алгоритмов контроля достоверности информации;

создание системы интерактивной обработки информационных ресурсов путем синтеза алгоритмов контроля достоверности текстов, экспериментальная проверка и анализ эффективности системы, расчет эффективности внедрения результатов исследований в телекоммуникационных системах и СЭД.

Объектом исследования являются информационные ресурсы и технологии сбора, формирования, ввода, передачи и обработки информации и документов систем электронного документооборота.

Предмет исследования – методы, алгоритмы и программные комплексы контроля и обеспечения достоверности информации, основанные на использование информационной PR-избыточности различной природы.

Методы исследований. В процессе исследования применены методы и алгоритмы теорий системного анализа, передачи информации, сетевых технологий, вероятности, распознавания образов, компьютерной лингвистики, а также алгоритмизации, моделирования и объектно-ориентированного программирования.

Научная новизна диссертационного исследования заключается в следующем:

разработаны концепции, методологии и программно-алгоритмические основы построения методов, моделей и алгоритмов обеспечения достоверности информации в системах электронного документооборота, выделены классы объектов, характеризующиеся PR-избыточностью, применяемой для обеспечения точности, целостности, оперативности, сжатия, доступности информационных ресурсов в СЭД;

предложены методы и алгоритмы контроля достоверности информации за счет использования искусственной избыточности на основе механизмов линейного, модульного, плоскостного суммирования и определения принадлежности к закодированным подмножествам;

разработаны методы и программный комплекс контроля достоверности информации за счет использования естественной избыточности на основе алгоритмов статистического, арифметического, парсингового кодирования, n-граммного структурированного описания, статистического распознавания и хеширования элементов текстов;

предложены методы и алгоритмы контроля и коррекции орфографических ошибок в текстах на узбекском языке на основе моделей многоуровневого морфологического анализа и описания n-граммной грамматики;

разработаны методы и алгоритмы контроля достоверности информации за счет использования структурно-технологической PR-избыточности на основе вложенных логических критериев, баз данных и знаний в составе встроенных экспертных систем;

разработаны методы синтеза алгоритмов контроля достоверности информации в системе интерактивного обнаружения и исправления ошибок при разработке технологий электронного документооборота.

Практические результаты исследования заключаются в следующем:

разработаны системы повышения достоверности информации на основе синтеза методов, использующих искусственную, естественную, статистическую и структурно-технологическую PR-избыточность, дополняющие существующие кодовые и аппаратные методы и обеспечивающие достоверность передачи текстовой информации на прикладных и сетевых уровнях модели OSI.

продемонстрирована эффективность программных комплексов контроля достоверности информации на основе критерия вероятностей необнаружения ошибок, установлено, что они до 92% обнаруживают все виды ошибок, способны исправлять однократные, двукратные и смежные транспозиционные ошибки, снижают трудоемкость и стоимость процедур контроля в 2-3 раза, повышают достоверность информации до трех порядков.

разработаны системы контроля и коррекции орфографических ошибок на узбекском языке на основе морфологических и n-граммных структурированных моделей, методов словарного, статистического и хеш-кодирования и предложена их реализация в программно-аппаратной среде параллельных вычислений NVIDIA с использованием стандартных библиотек численного анализа, оптимизированного обмена данными между CPU и GPU.

реализована архитектура системы контроля достоверности текстовой информации на базе фреймворк Сфинкс-4 для модулей с различными языковыми моделями и определено, что по сравнению с системой контроля орфографии на основе морфологического анализа при построенном варианте системы показатели числа необнаруженных ошибок, стоимости реализации и трудоемкости контроля уменьшаются до двух порядков.

Достоверность полученных результатов обосновывается тем, что проведено математическое исследование предложенных моделей, сравнительный анализ полученных формул и выкладок с реальными и экспериментальными данными на основе общепринятых критериев. Для оценки результатов исследований на их непротиворечивость протестирована эффективность программных комплексов на основе обширного контекста художественной и технической литературы, проведен сравнительный анализ эффективности программных комплексов по критериям трудоемкости, стоимости реализации и вероятности необнаружения ошибок для широкого класса алгоритмов контроля достоверности информации.

Теоретическая и практическая значимость результатов исследования. Теоретическая значимость полученных результатов исследования заключается в том, что методы, модели, алгоритмы и программные комплексы, разработанные на основе принципов использования PR-избыточности – искусственной, естественной и статистической, структурно-технологической, программно-временной в совокупности образуют концепцию создания систем повышения достоверности информации на основе синтеза прикладных методов, дополняющих существующие кодовые и аппаратные методы и обеспечивающих достоверность передачи текстовой информации в СЭД на прикладных уровнях модели OSI.

Практическая ценность работы заключается в применении PR-избыточности для программной реализации моделей многоуровневого описания грамматики и n-граммного структурированного представления текста, процедур контроля и коррекции орфографии узбекского языка, что позволяет рассчитать и эффективно использовать n-граммные показатели, сформировать правила контроля достоверности информации по механизмам парсингового распознавания и кодирования, обнаруживать и устранять ошибки на прикладном и представительском уровнях сетевой модели OSI. Внедрение программных комплексов контроля достоверности информации в структуре СЭД приводит к уменьшению временных и материальных затрат при составлении, вводе и обработке документов и ускоряет процессы обмена информацией за счет автоматического обнаружения и исправления ошибок, которые влечет за собой человеческий фактор.

Внедрение результатов исследования. Разработанные модели, алгоритмы, программные комплексы обеспечения точности, целостности, оперативности и сжатия информационных ресурсов за счет классов объектов с PR-избыточностью внедрены в системы адаптированной передачи, обработки и анализа данных Самаркандского филиала компании «УзТелеком» Государственного комитета связи, информатизации и телекоммуникационных технологий Республики Узбекистан (ДП «СамТелеком», 23.05.2010 г.). Результаты разработок системы адаптивной передачи, обработки и анализа данных внедрены в деятельности совместного предприятия «Lazzat Meva» (акт СП «Lazzat Meva» от 14.05.2010 г.) с годовой экономической эффективностью 62,2 млн. сум.

Апробация работы. Результаты исследования апробированы на 30 научно-практических конференциях, в том числе на 14 международных симпозиумах, конгрессах и семинарах, в частности, на международных: «Инновация» (Ташкент, 2004-2009); «Актуальные проблемы прикладной математики и информационных технологий – Аль-Хорезми 2009» (Ташкент, 2009); «ICI-2006», «ICI-2007» (Ташкент, 2006-2007); «ICEIC 2008» (Ташкент, 2008), X Международной научно-методической конференции ВУЗов по телекоммуникациям (Россия, Ростов-на-Дону, 2008), «WCIS-2008», «WCIS-2010», «WCIS-2012» (Ташкент, 2008-2012); III конгресс математиков тюркского мира (Алматы, 2009); и республиканских: «Современное состояние и пути развития информационных технологий» (Ташкент, 2006-2008), «Современные проблемы математики, механики и информационных технологий» (Ташкент, 2008), «Актуальные проблемы математики и информатики» (Самарканд, 2007-2010), Республиканская Ярмарка инновационных идей, технологий и проектов (Ташкент, 2008-2011).

Опубликованность результатов. По теме диссертации опубликовано 87 научных трудов, в том числе 9 в международных журналах.

Структура и объем диссертации. Диссертация состоит из введения, шести глав, заключения, списка литературы, 11 приложений и содержит 197 страниц текста, включает 47 рисунков и 12 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность и востребованность темы диссертации, сформулированы цель и задачи, выявлены объект и предмет исследования, определено соответствие исследования приоритетным направлениям развития науки и технологий Республики Узбекистан, изложены научная новизна и практические результаты исследования, обоснована достоверность полученных результатов, раскрыты теоретическая и практическая значимость полученных результатов, приведен список внедрений в практику результатов исследования, сведения по опубликованным работам и структуре диссертации.

В первой главе диссертации осуществлен теоретико-информационный анализ существующих кодовых и аппаратурных методов повышения достоверности передачи информации; определены задачи разработки методов, моделей и алгоритмов контроля достоверности передачи информации по принципам использования PR-избыточности – информационной избыточности различной природы: искусственной, естественной и статистической, структурно-технологической и программно-временной.

В работе представлена позиционная схема, отражающая принцип использования кодовой избыточности, который распространяется также на методы Рида-Соломона, БЧХ, FEC, применяемые для контроля и коррекции ошибок в подавляющем большинстве приложений в сфере цифровых телекоммуникаций. Проведенный системный анализ принципов и механизмов использования кодовой избыточности в алгоритмах обнаружения и исправления ошибок позволил определить их ориентированность на обнаружение одиночных, двойных, тройных ошибок, имеющих место в искусственно вводимых контрольных разрядах, что обусловило проведение исследований, направленных на расширение возможностей кодовых и аппаратурных методов контроля ошибок.

Гипотеза исследования и разработок, направленных на создание методов и алгоритмов контроля достоверности информации в структуре СЭД заключается в использовании PR-избыточности, т.е. избыточности различной природы – искусственной, структурно-технологической, программно-временной, статистической (естественной) – для повышения достоверности передачи и обработки информации.

Разработаны основные подходы к синтезу методов, алгоритмов и систем контроля достоверности информации. Последовательность выполнения задач проведенных исследований проиллюстрирована в виде схемы построения структурно-функциональных компонент автоматизированных систем контроля достоверности информации, в частности текстовой (АСКДТИ).

Принято, что при решении поставленных задач основным критерием эффективности разрабатываемых методов и алгоритмов выступает минимизация вероятности необнаружения ошибок за счет использования PR-избыточности до допустимых пределов достоверности информации.

Для оценки вероятности приема кодового слова S_0 предложена формула

$$P(S_0/S_1') = \sum_{i=1}^{2^m-1} (1 - P(S_0))^{-1} P(S_i/S_i'),$$

где $P(S_0)$ – вероятность обнаружения ошибки в исходном кодовом слове;

$P(S_i/S_1')$ – условная вероятность приема S_i -го кода при принятом S_1' ;

M – количество кодовых разрядов в слове.

Предложено оценивать вероятность неправильного приема кодового слова следующим образом:

$$P_H(S) = \sum_{i=0}^{2^m-1} P(S_i) \sum_{\substack{j=0 \\ j \neq i}}^{2^m-1} P(S_j) = \sum_{j=0}^{2^m-1} P(\Lambda_j),$$

где Λ_j – подмножество элементов неправильно принятых кодовых слов.

Вероятность необнаруженных ошибок предлагается оценивать, как:

$$P_{ocm}(S) = 1 - (1 - P_{\alpha_i})^m - \sum_{j=1}^{2^m-1} P(\Lambda_j).$$

Предложена схема формализации вероятностных процессов передачи информации и построены диаграммы вероятностных переходов для получения показателей оценки достоверности пакетной передачи данных, а также достоверности информации в узлах многоуровневой модели OSI.

Разработана методика определения минимального количества дополнительной информации, требуемой для обнаружения ошибок передачи данных. Для любого кодового метода, используемого в пакетах передачи данных для обнаружения ошибок в сообщении S_i , при условии, что известен принятый знак S'_k , количество информации определяется как

$$\begin{aligned} J(S_i/S'_k) &= m + \frac{1}{2^m} \sum_{k=1}^{2^m} \sum_{i=1}^{2^m} P(S_i/S'_k) \log(S_i/S'_k) = m + \sum_{i=1}^{2^m} P(S_i/S'_k) \log(S_i/S'_k) = \quad (1) \\ &= m + Q' \log Q' + P_H' \log P_H(S). \end{aligned}$$

Дальнейшим преобразованием, приняв $Q'(S) = Q(S)[1 - P_0(S)]^{-1}$ и $P_H'(S) = P_H(S)[1 - P_0(S)]^{-1}$ выражение (1) приводится к виду:

$$\begin{aligned} J(S_i/S'_k) &= m + \frac{Q(S)}{1 - P_0(S)} \log Q(S) - \frac{Q(S)}{1 - P_0(S)} \log[1 - P_0(S)] + \frac{P_H(S)}{1 - P_0(S)} \times \\ &\times \log P_H(S) - \frac{P_H(S)}{1 - P_0(S)} \log[1 - P_0(S)] - \frac{P_H(S)}{1 - P_0(S)} \log(2^m - 1) = m - \frac{P_H(S)}{1 - P_0(S)} \times \\ &\times \log(2^m - 1) - \log[1 - P_0(S)] + \frac{Q(S)}{1 - P_0(S)} \log Q(S) + \frac{P_H(S)}{1 - P_0(S)} \log P_H(S) \end{aligned}$$

где $P_0(S)$ - вероятность обнаружения ошибок в сообщении S ;

$Q(S)$ - вероятность правильного приема сообщения S .

Значение $J(S_i/S'_k)$ определяет минимум дополнительной информации о сообщении, требуемой для устранения потерь из-за необнаружения ошибок.

Предложено выражение упрощенной оценки требуемого объема избыточности для оценки эффективности произвольного кодового метода

$$R(S_i) = 1 - \frac{1}{(1-P)} \log \frac{(2^m - 1)}{nP}.$$

Показано, что по разработанной методике оценивается эффективность программных способов контроля достоверности информации, основанных на использовании PR-избыточности.

Во **второй главе** диссертации разработаны методы и алгоритмы контроля достоверности информации в составе пакетов передачи данных по принципам использования искусственной избыточности.

Разработаны методы и алгоритмы контроля достоверности информации, основанные на использовании искусственной избыточности с помощью механизмов линейного, плоскостного и модульного суммирования, устраняющих недостатки традиционных программных методов. Алгоритмы контроля достоверности информации путем суммирования могут установить не только наличие ошибки в принятой последовательности, но и указать номер позиции, в которой произошли ошибки, для их автоматического исправления. Они основаны на следующих правилах: π_1 – контроль достоверности информации; π_2 – обнаружение и исправление ошибок типа Q_1 ; π_3 – обнаружение и исправление ошибок типа Q_2 ; π_4 – обнаружение ошибок типа Q_3 ; π_5 – обнаружение ошибок типа Q_4 .

Разработана методика проведения общего и частного решения задач оценки эффективности контроля достоверности информации по критерию необнаружения ошибок для различных алгоритмов.

Суммарная вероятность необнаруженных ошибок для правил контроля достоверности информации по линейному суммированию определяется как

$$P_H = P_1 + P_2 + P_3 + P_4 = \{67[mnp^2(1-p)^{m+n-2} + C_m^2 C_n^2 p^4(1-p)^{m+n-4} + C_m^k C_n^3 p^{3k}(1-p)^{m+n-k-3}] + 5,76 \cdot 10^{-6}\}.$$

В работе исследованы графики эффективности алгоритмов, построенные для критериев трудоемкости контроля и стоимости реализации, которые показали, что при применении алгоритмов линейного и плоскостного суммирования (Π_3 и Π_6) по сравнению с другими методами трудоемкость и стоимость контроля снижается в 2-3 раза.

Графики зависимости вероятности необнаружения ошибок алгоритмами контроля P_H от вероятности ошибок P показывают, что при вероятности ошибок в пределах $10^{-3} - 10^{-4}$ разность способности обнаружения ошибок исследованных алгоритмов достигает до двух порядков, трудоемкость и стоимость контроля снижается в 2-3 раза за счет автоматического исправления ошибок.

Наибольший эффект по показателю достоверности информации достигается при применении алгоритмов: Π_2 – поразрядного суммирования, Π_3 поцифрового суммирования, Π_6 – плоскостного суммирования.

Проведен анализ эффективности алгоритмов цифрового, числового модульного суммирования на базисе модулей 8, 9, 10, 11 и произвольного модуля q . Наибольший эффект по показателю достоверности информации достигается при применении алгоритмов почислового модульного суммирования по модулям 8, 10 и 11.

Установлено, что разработанные алгоритмы обнаруживают до 92% всех видов ошибок и исправляют одно- и двукратные смежные транспозиционные ошибки.

Решена задача определения вероятности ошибок первого и второго родов алгоритма. При контроле одиночных ошибок получены следующие формулы:

Первого рода:

$$P_1 \approx 2P_{\alpha_i}(1 - P_{\alpha_i}) \cdot 10^{-2} + P_{\alpha_i}^2 \cdot 10^{-2} + 2P_{\alpha_i}(1 - P_{\alpha_i}) \cdot 10^{-4}.$$

Второго рода:

$$P_2 \approx (1 - P_{\alpha_i})^2 \cdot 1.5 \cdot 10^{-5} + P_{\alpha_i}(1 - P_{\alpha_i}) \cdot 1.5 \cdot 10^{-5}.$$

Суммарная вероятность необнаруженных ошибок алгоритма равна:

$$P_H = P_1 + P_2 \approx 2P_{\alpha_i}(1 - P_{\alpha_i}) \cdot 10^{-2} + (1 - P_{\alpha_i})^2 \cdot 1.5 \cdot 10^{-5}.$$

По представленной методике решения задач по критерию вероятности необнаружения ошибок проанализирована эффективность алгоритмов, построенных для контроля и обнаружения ошибок всех остальных видов.

В третьей главе диссертации разработаны методы и алгоритмы контроля достоверности информации на основе использования статистической, структурно-технологической, программно-временной избыточности.

Определено, что алгоритмы, основанные на механизмы использования логических, технологических, статистических связей и свойств информации создают благоприятные условия для контроля данных и обеспечивают высокую достоверность передачи информации.

Сформированы принципы построения программно-алгоритмического комплекса контроля информации на основе баз данных и баз знаний (БД и БЗ), систем управления БД (СУБД) и экспертных систем.

Разработаны механизмы применения для контроля достоверности информации статистической избыточности, которая традиционно используется в задачах кодирования и сжатия передаваемой информации.

Предложена обобщенная модель оценки вероятности необнаружения ошибок при использовании исследованных алгоритмов контроля изображений элементов текста:

$$P_H = P_1W_1 + P_2W_2,$$

где W_1 , W_2 - коэффициенты потери, связанные соответственно с ошибками первого и второго рода, которые задаются в вещественном интервале $[0,1]$.

При этом все элементы контролируемого текста разделяются на два класса: достоверные данные и искаженные.

Усреднение вероятности при большом числе N_1 элементов первого класса дает

$$x_1 = N_1 p_1 (1 - P_{\text{л}}) W_1 + N_1 (1 - p_1) P_{\text{л}} W_2,$$

где $P_{\text{л}}$ – вероятность того, что элемент отнесен к неверному классу.

Усреднение же по N_2 элементов второго класса приводит к

$$x_2 = N_2 p_1 P_{\text{л}} W_1 + N_2 (1 - p_1) (1 - P_{\text{л}}) W_2.$$

Вычисление взвешенной разности дает

$$\frac{x_1}{N_1} - \frac{x_2}{N_2} = p_1 (1 - 2P_{\text{л}}) W_1 - (1 - p_1) (1 - 2P_{\text{л}}) W_2.$$

Для любого $P_{\text{л}} \neq 1/2$, оцениваются сумма и разность $p_1 W_1$ и $(1 - p_1) W_2$, позволяющие получить W_1 и W_2 в виде

$$W_1 = \frac{x_1 N_2 - x_2 N_1}{N_1 N_2 p_1 (1 - 2P_{\text{л}})}; \quad W_2 = \frac{x_1 N_2 - x_2 N_1}{N_1 N_2 (1 - p_1) (1 - 2P_{\text{л}})}.$$

В работе разработаны алгоритмы контроля достоверности информации, основанные на использовании механизмов сжатия методов Хаффмана и арифметического кодирования, преобразование Барроуза-Вилера, а также алгоритм Лемпеля-Зива и их модификации.

В предложенных методах ключевым моментом является определение нижних и верхних вероятностных границ контроля разрешенных кодовых комбинаций и оценка наиболее вероятных элементов сообщений, предоставляемых для контроля.

В качестве иллюстрации этого явления построены кривые плотности распределения вероятностей переданной информации α в зависимости от принятой β , как график функций $\varpi(\alpha)$ и $\varpi(\beta/\alpha)$ – условного распределения для нормального закона.

Практическая реализация разработок показала, что чем уже пределы возможных отклонений от контролируемого значения, тем больше ошибок первого рода можно обнаружить методом двумерной регрессии или чем теснее взаимосвязь между контролируруемыми параметрами, тем значение параметра ближе к истинному.

Получено общее и частное решения задач, заключающиеся в нахождении оптимальных границ условного момента, минимизирующих общую вероятность ошибок при контроле информации вида:

$$P_H = P_1 + P_2 = \frac{P}{B} \int_{x(\alpha)}^{y(\alpha)} W(\alpha) d\alpha + (1 - P) \left(1 - \int_{x(\alpha)}^{y(\alpha)} W(\alpha) d\alpha \right).$$

При этом, получены оценки достоверности информации на основе нахождения порога различия функций распределений $W(\alpha)$ и $W(\beta)$ по условной вероятности распределения $W(\beta/\alpha)$.

Задача контроля достоверности информации также решена на основе использования структурной PR-избыточности в контейнерах изображений пакета передачи данных.

Введены следующие обозначения процедур включения и извлечения: $E(V, W) = V_W$ – процедура включения структурной избыточности в контейнер изображений; $D(V, V_W) = W'$ (или $D(V_W) = W'$) – процедура извлечения структурной избыточности в случае необходимости; α – масштабирующий коэффициент; C – бинарный признак подобия двух сигналов: $C = 1$, если коэффициент взаимной корреляции больше некоторого порога δ ; в противном случае $C = 0$.

Обоснованы следующие варианты работы алгоритмов контроля достоверности информации на основе структурной избыточности.

В первом варианте алгоритм контроля - $A_1(E, D, C)$ основывается на строго обратимом процессе. При этом предполагается, что для любого V_W существует отображение E^{-1} такое, что выполняются условия: $E^{-1}(V_W) = (V_F, W_F)$ и $E(V_F, W_F) = V_W$, причем, E^{-1} вычислительно осуществимо, W_F принадлежит к классу структурной избыточности, позволяющей извлечь истинное и ложное изображения и $C(D(V_W, V_F), W_F, \delta) = 1$.

Во втором варианте алгоритм $A_2(E, D, C)$ основывается на слабо обратимом процессе. При этом ставится требование, чтобы выполнялось условие $E(V_F, W_F) = V_W$. Отмечено, что даже $E(V, W) = V_W$ может не выполняться в силу различного рода искажений V_W . С другой стороны, это требование слишком слабо для определения обратимости, в связи с чем оно заменено на требование, чтобы $E(V_F, W_F) = V_{W'}$, где $C(V_{W'}, V_W, \delta) = 1$.

Другим условием учета слабой обратимости процесса в алгоритме $A_2(E, D, C)$ является то, что для любого V_W существует отображение E^{-1} такое, где $E^{-1}(V_W) = (V_F, W_F)$ и $E(V_F, W_F) = V_{W'}$. При этом E^{-1} вычислительно осуществимо, $C(V_{W'}, V_F, \delta) = 1$, $C(V_{W'}, V_W, \delta) = 1$ и $C(D(V_{W'}, V_F), W_F, \delta) = 1$.

Разработаны алгоритмы контроля достоверности информации, основанные на использовании структурно-технологической PR-избыточности и встроенных экспертных систем, баз знаний, процедур логического контроля информации и структур данных. Предложены принципы выбора характеристик, критерии и места реализации решающих правил в составе рассматриваемой СЭД. На основе сформулированных принципов и критериев построены алгоритмы, программные модули систем контроля достоверности информации в СЭД.

Предложен комбинированный подход к использованию программно-временной избыточности для реализации средств автоматического контроля и восстановления данных, а также для функционирования средств защиты на основе PR-избыточности. Проведена классификация организационно-оперативных методов контроля достоверности информации в СЭД.

Система контроля достоверности информации построена на основе технологии объектно-ориентированного программирования и включает тринадцать функциональных и восемнадцать прикладных программных модулей, которые устойчивы к различным возмущениям и способны сохранять достаточное качество результатов во всех реальных условиях.

Проведен сравнительный анализ эффективности разработанных методов и алгоритмов контроля достоверности информации. Сближение экспериментальных значений - P^*_H к аналитическим, полученным в теоретических разделах диссертации проверены по критерию χ^2 – Пирсона при гарантийной вероятности 0.9, что подтверждает хорошее соответствие экспериментальных результатов теоретическим.

В четвертой главе диссертации разработаны методы и алгоритмы контроля достоверности информации по принципам использования статистической избыточности на основе n -граммных языковых моделей, моделей парсингового кодирования и распознавания элементов текста.

Смоделированы процессы построения закономерностей распределения n -граммной статистики и разработаны алгоритмы: выявления искаженных элементов текста, кластеризации, поиска, структуризации, а также получения частотных характеристик, реализуемые при построении системы контроля и коррекции ошибок.

Выявлено, что работа системы контроля на основе моделей поиска, распознавания и кластеризации, связана с привлечением большого объема памяти. Для устранения этих недостатков при решении задач контроля орфографии текстов рекомендовано использование механизмов систем параллельных вычислений, основанных на модели GPU и технологии CUDA.

Оценка минимизированной таким образом потребляемой памяти представляется функцией

$$f(x) = -\frac{nx}{\ln(1 - P^{1/x})}$$

где n – длина кодов,

x – переменная, обозначающая количество хеш-функций.

Обоснованы принципы, предложены модели и алгоритмы построения систем контроля достоверности информации, позволяющие сочетать: механизмы словарного, статистического и хеш-кодирования; способы, алгоритмы, процедуры параллельных вычислений.

При построении программно-технического комплекса (ПТК) контроля достоверности информации (КДИ) использовано унифицированное программно-аппаратное решение для параллельных вычислений на видеочипах NVIDIA. Особенностью системы на основе параллельных вычислений является то, что при ее исполнении решается задача устранения превосходства числа логических потоков над физическими устройствами.

Для оценки эффективности алгоритмов контроля достоверности информации на основе парсингового моделирования предложено использовать полную вероятность выделения и распознавания строки слов

$$W \doteq \arg \max_W P(W | A)P(A),$$

где W – исследуемое слово в виде множества элементов текста $W = w_1, w_2, \dots, w_n$;

$P(W / A)$ – условная вероятность слова W в строке A ;

$P(A)$ – априорная вероятность строки $A = \{a_1, a_2, \dots, a_n\}$ в тексте.

Работа системы контроля орфографических ошибок на основе алгоритмов парсингового кодирования организована по предложенной древовидной модели.



Рис. 1. Взаимосвязь модулей парсингового распознавания текста

«Предсказатель слова» предсказывает следующее слово w_{k+1} , данное k -префиксом слово-парсинга, и затем передает управление модулю «Таггер». «Таггер» предсказывает POS признак t_{k+1} следующего слова, данного k -префиксом слово-парсинга и последнего предсказанного слова w_{k+1} , а затем передает управление модулю «Конструктор». «Конструктор» наращивает структуру, повторно генерируя переходы, пока управление не перейдет к модулю «Предсказатель».

В работе разработан алгоритм распознавания на основе нейронной сети (НС) ассоциативной памяти для контроля достоверности элементов текста. Процесс ввода и контроля достоверности сигнала изображения элемента текста на основе НС проиллюстрирован на рис. 2.

Предложена методика оценки вероятности недостоверного распознавания путем сравнения распознаваемого элемента текста со специальными справочниками словоформ, частотными словарями монограмм, диграмм, триграмм, к-грамм.

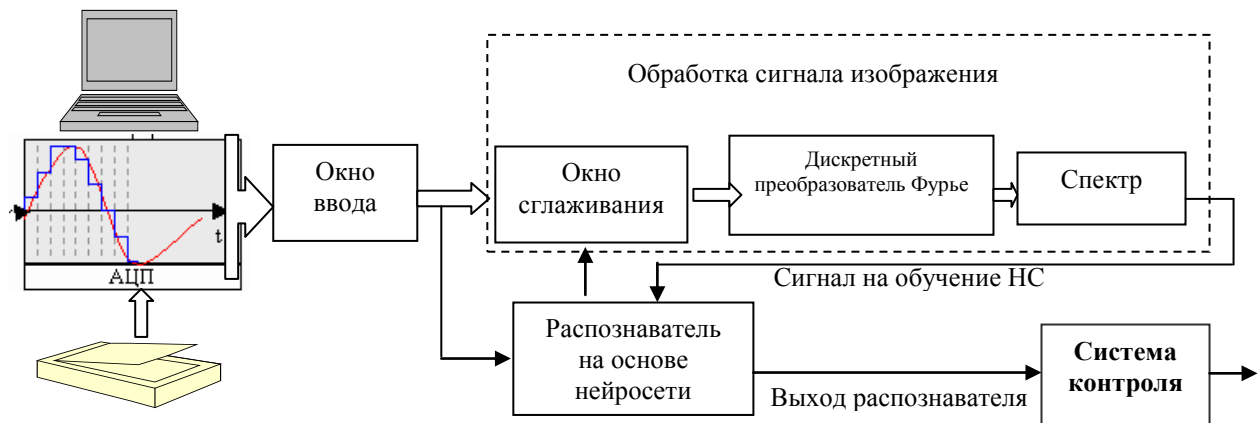


Рис. 2. Контроль достоверности сигнала элемента текста

Вероятность неправильного распознавания по общему объему справочника определяется как

$$p(\tau) = 1 - (1 - p(a)) \prod_{i=1}^{n-1} (1 - ((1 - p(a)) + p(a \equiv a_i)(2p(a) - 1))), \quad p(a) < 10^{-3},$$

где τ – объем специального справочника проверки достоверности распознавания образа;

$p(a)$ – вероятность ошибки элемента текста при передаче и обработке информации;

$p(a \equiv a_i)$ – вероятность правильного распознавания, n – количество распознаваемых элементов текста.

Для оценки верности распознавания системы, где в качестве эталонного образца используется достоверный элемент текста, получена формула:

$$p(\tau) = 1 - (1 - p(a)) \prod_{i=1}^{n-1} \prod_{j < i, a_j, a_j \in A} (p(a_j \equiv a_i)),$$

где $p(a_j \equiv a_i)$ – вероятность появления информации на выходе рассматриваемой системы контроля достоверности информации.

При построении обобщенного алгоритма системы обработки метатекста узбекского языка предложена схема определения иерархии классов, которая представляет дерево с произвольным количеством ветвлений и осуществлена классификация контролируемых объектов. В реализованной системе первичные данные хранятся в БЗ и уточняются в процессе приобретения новых знаний при эксплуатации СЭД. Реальная модель формируется в среде БЗ, а связь с сервером осуществляется через запросы пользователя системы.

Разработанная система состоит из следующих функциональных частей: БД, в которой хранятся эталонные и фактические данные об объекте, результаты их сравнения, концептуальная, инфологическая и физические модели объекта; БЗ статическая – экспертные знания, формулы факты, зависимости, таблицы; БЗ динамическая – комбинированные модели НС в виде эталонных динамических процессов, с учетом частичной или полной неопределенности параметров; модуль логического вывода по алгоритму порождения причинно-следственной сети событий в функционально-структурной модели; модуль адаптации, координирующий работу БД и БЗ в процессе логического вывода в зависимости от сложившейся ситуации; модуль объяснения, который представляет собой интерпретацию процесса логического вывода; планировщик, координирующий процесс решения задачи; решатель, позволяющий находить эффективные решения в прямой, обратной и смешанной постановке задач.

В пятой главе диссертации исследованы основные подходы к контролю и коррекции орфографии текстов на узбекском языке, разработаны модели, методы и алгоритмы контроля и коррекции орфографических ошибок на основе многоуровневого морфологического анализа и использования n-граммного представления словоформ.

В результате проведенного анализа литературы выявлены принципы, особенности и недостатки наиболее известных методов построения систем на основе морфологического анализа: разбиение текста на слова, сопоставление со словарем, приближенное выделение основ слов, «слияние» текста со словарем и нормализация слов.

Определено, что считающиеся наиболее эффективными системы основаны на принципах использования двухуровневых моделей анализа словоформ, которые рекомендуются в качестве базового механизма при построении систем контроля и коррекции орфографии узбекского языка.

Следовательно, при построении систем контроля ошибок в текстах на узбекском языке наибольшую теоретическую и практическую значимость представляет реализация многоуровневой модели морфологического анализа на основе синтеза принципов и возможностей двухуровневых моделей и способов анализа словоформ.

Разработанная в работе многоуровневая графическая схема построения словоформ узбекского языка отражает последовательность присоединения аффиксов к основам слов.

Получены методики и построены алгоритмы формального описания и анализа узбекских словоформ, их фонетических изменений и описана многоуровневая таблица взаимосвязей узбекских аффиксов, позволяющая четко разобрать с конца заданное слово, рекурсивным подходом разложить его по уровням аффиксов и тем самым убедиться в корректности орфографии. На основе полученных теоретических положений разработаны алгоритмы и структура системы контроля и коррекции орфографии.

Разработана методика оценки вероятностей появления ошибок в информации на основе n -граммной модели.

Средняя вероятность ошибки находится путем усреднения условных вероятностей ошибки по всему ансамблю сообщений:

$$P = \sum_{\alpha_i} P_{\alpha_i} \sum_{\alpha_j (\alpha_i \neq \alpha_j)} P_{\alpha_i \alpha_j},$$

где $P_{\alpha_i \alpha_j}$ - вероятность правильного приема α_j - го сообщения.

В связи с тем, что определение вероятностей n -граммных ошибок связано с обработкой большого объема статистических данных и трудоемкими вычислениями, проведено специальное моделирование и осуществлена автоматизация процессов вычисления статистики вероятностей n -граммных ошибок.

Предложенная методика определяет значение вероятности $p(\alpha)$ появления n -грамм $\alpha = \alpha_1 \dots \alpha_n$, равное отношению $C(\alpha)$ к общему числу экземпляров всех встреченных в совокупности n -грамм:

$$p(\alpha) = \frac{C(\alpha)}{\sum_{\alpha^*} C(\alpha^*)},$$

где α^* - экземпляр n -граммы из «генеральной совокупности всех последовательностей», отличный от α .

На основе условной вероятности $p(\alpha_i = \alpha_j^* / \alpha_i = \alpha_i^* \forall i \neq j)$ появления ошибок α_j^* получена полная вероятность строки

$$p(\alpha_1^* \dots \alpha_n^*) = p(\alpha_i = \alpha_j^* | \alpha_i = \alpha_i^* \forall i \neq j) p(\alpha_i = \alpha_j^* \forall i \neq j).$$

Для вычисления условной вероятности появления ошибочных символов в строке с учетом всех предшествующих символов предлагается правило:

$$p(\alpha_1 \dots \alpha_n) = \left(\prod_{k=2}^n p(\alpha_k / \alpha_1 \dots \alpha_{k-1}) \right) p(\alpha_1).$$

Используя Марковскую цепь $(n-1)$ -го порядка получена простая n -граммная модель ошибок, которая позволила оценить статистические параметры рассматриваемых n -грамм.

Разработан модифицированный вариант системы контроля орфографии на базе технологии фреймворк Сфинкс-4, которая позволяет создать модульную структуру системы, включает отдельные компоненты, посвященные решениям специфических задач и адаптации путем поддержки контекста свободных n -грамм. Разработана оболочка системы и способы описания ядра функционирования системы с поддержкой всех компонент и программных модулей фреймворк.

Архитектура разработанной системы на базе фреймворк Сфинкс-4 приведена на рис. 3 и включает основные блоки: Интерфейс, Декодер, Лингвист, а также вспомогательные блоки Конфигурация и Утилиты.

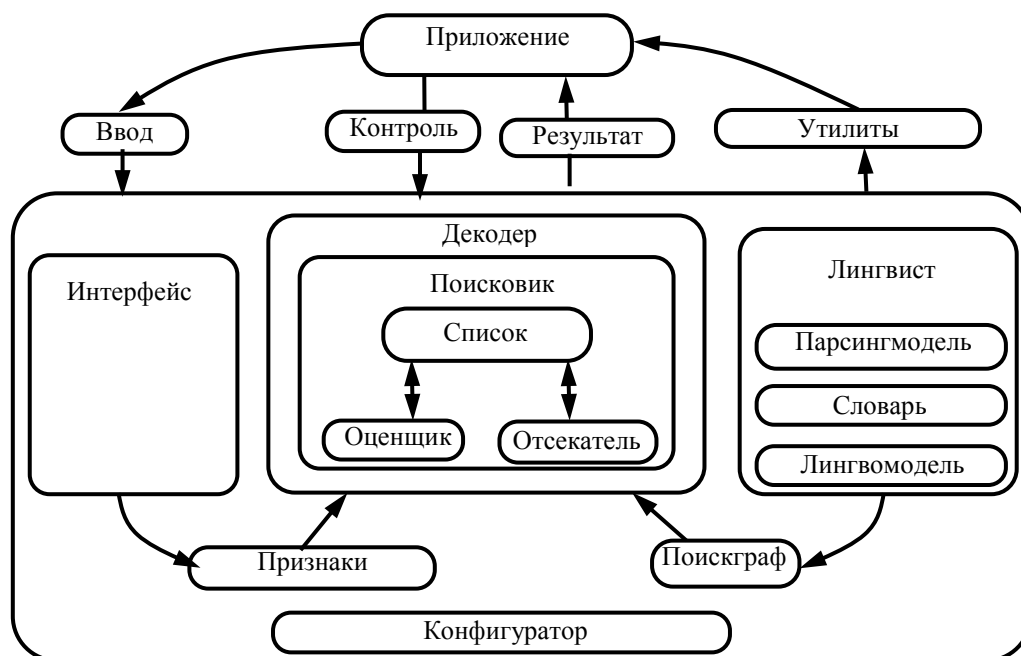


Рис. 3. Архитектура системы контроля орфографии на базе Сфинкс-4

На схеме также изображены связи между блоками и связи блоков с основным приложением. В работе приведено описание функций основных и вспомогательных блоков системы Интерфейс, Декодер и Лингвист и их взаимодействие в фреймворк Сфинкс-4.

Функцией блока Интерфейс является прием одного или более элементов текста (символов, букв, слов) на вход и параметризация их в последовательность признаков.

Лингвист предназначен для перевода стандартной языковой модели вместе с информацией о парсинговой структурированности n -грамм из одного или более наборов Парсинг-модели в Поиск-граф.

Поисковик в Декодере использует как признаки, полученные из Интерфейса, так и Поиск-граф, генерируемый в Лингвист, чтобы выполнить фактическую расшифровку, генерацию результатов.

Проведенные тестовые испытания показали, что построенный вариант системы улучшает показатель ошибки на слово при распознавании и скорость работы системы, измеренную в режиме реального времени, – по сравнению с системой контроля орфографии на основе морфологического анализа в два раза. Следует отметить, что работа системы была организована в двухпроцессорном режиме с процессорами CPU 1015MHZ ULTRASPARC RIII WITH 2G RAM. Благодаря такому использованию модулей, входящих в оболочку Сфинкс-4, адаптированных алгоритмов и стохастических моделей языка возникает возможность поддерживать широкое разнообразие задач, связанных с анализом текста, генерацией текста в речь, речи в текст и других задач обработки текстовой информации.

Для организации удобного порядка потоков обработки и загрузки данных для систем контроля достоверности информации на основе n -граммных моделей разработаны способы представления статистической информации, используемой в n -граммной грамматике и древовидная модель ее представления, которая позволяет сократить размеры обрабатываемых файлов путем устранения избыточности в структуре формата.

В работе проиллюстрирована схема процедур манипуляции над элементами грамматики для сокращения избыточности в древовидном представлении.

Построенный формат описания n -граммной грамматики позволяет поддерживать различные языки программирования и осуществлять сетевую обработку информации за счет того, что описание n -граммных моделей осуществлено на основе грамматики XML.

В **шестой главе** диссертации приведены сведения по разработке классов, методов и программных модулей систем контроля достоверности информации на основе использования PR-избыточности и реализации их в структуре пакетов передачи данных систем сетевой обработки информации и в СЭД.

Разработан комплекс программно-инструментальных средств (ПИС) СЭД, предназначенных для создания, редактирования, хранения и передачи элементов делопроизводства предприятия, созданных в интегрированной среде 1С 8.2. Особенностью реализации системы является то, что в редакторах, связанных с вводом, набором, трансформацией информации, функционирует подсистема контроля достоверности данных CONTROL, состоящая из комплекса программных модулей, разработанных в Delphi.

Комплекс программ привязывается к основным подсистемам и функциональным задачам применяемой информационной системы и представляется в виде программных продуктов.

Следует отметить, что подсистему контроля достоверности информации можно использовать локально в качестве сервисной поддержки любой компьютерной системы обработки текстовой информации и включить в состав функциональных подсистем СЭД, а также в структуру протоколов современных телекоммуникационных систем для борьбы с ошибками на Прикладном и Представительском уровнях OSI.

В принятой СЭД выделены основные объекты контроля, которые выполняют: полномасштабный поиск документов по реквизитам регистрационной карточки; ведение справочника должностей сотрудников; ведение справочника стандартных фраз и словесных оборотов; регулярную корректировку справочника организаций; авторизацию прав доступа сотрудников к системе в привязке к их функциональным ролям в разрезе документов, картотек регистрации, дел, тематических рубрикаторов и других критериев, что позволяет эффективно использовать метод объектно-ориентированного программирования, БД и БЗ.

Разработаны программные оболочки систем контроля и коррекции орфографических ошибок в среде Сфинкс-4 и на основе технологии CUDA NVIDIA, для которых общими являются модули, реализующие следующие функции: модуль Ф1 – получает на вход слово и пропускает его через разбор по виртуальной таблице; модуль Ф2 – реализует метод бинарного поиска в словаре; модуль Ф3 – формирует массив для хранения и загрузки словаря; модуль Ф4 – реализует быструю сортировку; модуль Ф5 – реализует замену объектов; модуль Ф6 – проверяет границы слов; модуль Ф7 – осуществляет бинарный поиск в загруженном словаре; модуль Ф8 – проверяет орфографию; модули Ф9-Ф15 – реализуют процедуры по транслитерации текста в среде проверки орфографии; модули Ф16-Ф20 – реализуют методы парсинга входного текста и выполняют процедуры парсингового моделирования; модуль Ф21 – позволяет пользователям работать с документом и редактировать различные шрифтовые свойства, добавлять, редактировать и удалять текст и сохранять его в файл; модуль Ф22 – MDI приложение, реализующее текстовый процессор для редактирования и форматирования текста на узбекском языке; модуль Ф23 – обеспечивает правильное закрытие всех документов, если все приложение закрывается.

Для выполнения функций коррекции орфографических ошибок реализованы модули, основанные на адаптивном кодере Хаффмана. При обратимом преобразовании использованы модель и алгоритм кодирования и сжатия по правилу Зива-Лемпеля. Реализован модифицированный модуль на основе статистического кодирования Лемпеля-Зива-Велча, принцип применения которого заключается в использовании словаря, хранящего строки текста и содержащего пронумерованные гнезда. Разработан модуль контроля на основе простой, полуадаптивной и адаптивной моделей арифметического кодирования.

Выделены основные классы, предложены методы и общая структура организации системы контроля орфографии узбекского языка, получены свидетельства о регистрации программных продуктов. Узбекский spell-checker обладает способностью проверять орфографию в обоих представлениях алфавита (латиница, кириллица) и при необходимости переводить из одного представления в другое.

Экспериментальная проверка и анализ эффективности разработанных алгоритмов и систем контроля достоверности текстовой информации осуществлены по данным о функционирующих системах мониторинга деятельности СамГУ и СЭД «Лаззат-Мева». Результаты получены при условиях, заданных близко к практическим: объем вводимой информации равен 10^6 знакам (около 300 страниц электронного текста); значение вероятности ошибок – $P = 3,7 \cdot 10^{-3}$; слово в документе принято в среднем равным 8 десятичным знакам; защищена строка документа в целом при применении алгоритмов, основанных на использовании PR-избыточности информации; ошибками обработки, обусловленными сбоями электронных устройств, при расчетах пренебрегают.

Результаты сравнительного анализа эффективности разработанных алгоритмов по различным критериям эффективности приведены в таблице. В графах 6, 7 приведены значения коэффициента снижения вероятностей необнаруженных ошибок (коэффициент выигрыша в достоверности информации) соответственно $K_H = P/P_H$ – теоретическое и $K_H^* = P/P_H^*$ – экспериментальное. В графах 8, 9 приведены значения коэффициента трудоемкости $K_{ПА} = T_P/T_{ПА}$, $K_{ТА} = T_P/T_A$, а в графах 10, 11 коэффициенты стоимости $K_{СПА} = C_P/C_{ПА}$, $K_{СА} = C_P/C_A$.

Для существующих условий обработки данных на объектах внедрения принято, что трудоемкость клавиатурного ввода информации равна 400 часов, стоимость ее ввода – 60 тысяч сумов, трудоемкость ручного контроля ошибок – 350 часов, стоимость ручного контроля ошибок – 45 тысяч сумов, трудоемкость сканирования и распознавания – 150 часов, а стоимость – 25 тысяч сумов.

Установлено, что применение алгоритма с использованием линейного суммирования обеспечивает: минимальную вероятность необнаруженных ошибок – $P_H \approx 3 \cdot 10^{-5}$; трудоемкость контроля – 0,5 ч.; стоимость контроля – 1 тыс. сумов.

Применение алгоритма с плоскостным суммированием обеспечивает: минимальную вероятность необнаруженных ошибок – $P_H \approx 4 \cdot 10^{-6}$; трудоемкость контроля – 1,2 ч.; стоимость контроля – 1,5-2 тыс. сумов.

Применение алгоритма по модульному суммированию обеспечивает: минимальную вероятность необнаруженных ошибок – $P_H \approx 6 \cdot 10^{-5}$; трудоемкость контроля – 0,8 ч.; стоимость контроля – 1,2 тыс. сумов.

Таблица

Сравнительный анализ эффективности алгоритмов систем контроля достоверности информации

№	Названия алгоритмов контроля, обнаружения и исправления ошибок	Значения критериев эффективности							
		Вероятность P_H		Коэффициент $K_H = P/P_H$		Трудоемкость контроля - K_T		Стоимость контроля - K_C	
		теор.- P_H	эксп P_H^*	теор K_H	эксп K_H^*	$K_{ТПА}$	$K_{ТА}$	$K_{СПА}$	$K_{СА}$
Использование искусственной избыточности									
1	По линейному суммированию	10^{-5}	$2 \cdot 10^{-5}$	100	200	1,8	1,7	2	2
2	По модульному суммированию	$10^{-5}-10^{-6}$	$5 \cdot 10^{-6}$	1000	5000	2	2	2	2
Использование естественной избыточности									
3	По n-граммным моделям	$10^{-5}-10^{-6}$	$4 \cdot 10^{-6}$	1000	4000	2	2	3	3
4	По морфологической модели	10^{-6}	$2 \cdot 10^{-6}$	1000	2000	3	3	3	3
Использование структурно-технологической избыточности									
5	Алгоритмы контроля на основе экспертных систем и баз знаний	$10^{-5}-10^{-6}$	$1.5 \cdot 10^{-6}$	1000	1500	1,8	1,7	2	2
6	Алгоритмы, построенные по критериям логического контроля	10^{-5}	$2 \cdot 10^{-5}$	100	200	2	2	2	2
Использование статистической избыточности									
7	Алгоритмы на основе метода Лемпеля-Зива-Велча и преобразования Барроуза-Вилера	$2 \cdot 10^{-5}-10^{-6}$	$6 \cdot 10^{-6}$	2000	6000	3	3	3	3
8	Алгоритмы на основе арифметического кодирования	$3 \cdot 10^{-6}-10^{-7}$	10^{-7}	3000	10000	3	3	3	3

Определено, что по сравнению с традиционными, разработанные методы снижают трудоемкость и стоимость процедур контроля в 2-3 раза. Отношения трудоемкости и стоимости ручного режима работы к трудоемкости и стоимости полуавтоматического и автоматического режимов свидетельствуют о снижении значений принятых коэффициентов почти в 2-3 раза. Сближение экспериментальных значений – P_H^* к аналитическим проверены по критерию χ^2 – Пирсона при гарантийной вероятности 0,9, что подтверждает соответствие экспериментальных результатов теоретическим.

Разработана методика расчета экономической эффективности и внедрения методов, алгоритмов и программно-инструментальных средств в СЭД и системах адаптированной передачи, обработки и анализа данных в инфокоммуникационных сетях. Расчетная экономическая эффективность компьютерной системы адаптированной передачи, обработки и анализа данных в инфокоммуникационных сетях в Самаркандском филиале «УзТелеком», а также в совместном предприятии «Лаззат-Мева» по результатам обработки данных, представляемых в ежемесячных отчетах составляет 5,1 млн. сум.

ЗАКЛЮЧЕНИЕ

1. Разработанные в диссертации на основе концепций системного анализа, управления и обработки информации конструктивные методы, модели и программно-алгоритмические комплексы обеспечения достоверности информации по принципам использования PR-избыточности позволяют повысить эффективность и производительность СЭД.

2. Оценка современного состояния теории и практики кодовых, аппаратных и программных методов контроля достоверности передачи информации показала недостаточность использования существующих типов избыточности для обеспечения качественного функционирования СЭД. Принципы использования PR-избыточности электронных документов при разработке прикладных методов и программно-алгоритмических комплексов обеспечения достоверности передачи информации позволили спроектировать инструментарий развития существующих технологий.

3. Методы: линейного, плоскостного и модульного суммирования; кодирования по правилам Хаффмана, Лемпеля-Зива-Велча, Барроуза-Вилера, арифметического кодирования, статистического распознавания, логического контроля образуют методические основы использования PR-избыточности для расширения возможностей алгоритмов и программных комплексов обеспечения достоверности информации при составлении, передаче и обработке текстов электронных документов.

4. Установлено, что разработанные алгоритмы до 92% обнаруживают все виды ошибок, способны исправлять однократные, двукратные и смежные транспозиционные ошибки и по сравнению с существующими методами снижают трудоемкость и стоимость процедур контроля в 2-3 раза при вероятности ошибок $P_H \approx 4 \cdot 10^{-3}$, а также повышают достоверность

информации до трех порядков. Оценки эффективности и решения задач контроля достоверности информации получены по критерию вероятностей необнаружения ошибок.

5. Для решения задач контроля и коррекции орфографических ошибок в текстах на узбекском языке разработаны методы, алгоритмы и системы, включающие в себя морфологические и n-граммные структурированные модели. Разработанная методика получения частотных характеристик n-грамм на основе статистики вероятностных показателей искажений применяется при систематизации хеш-кодов для парсингового кодирования.

6. Интерполяционные и экстраполяционные методы построения логической и арифметической функции статистического распознавания применены при разработке алгоритмов контроля достоверности изображений элементов текста. Предложены методы контроля достоверности сигнальных характеристик изображений элементов текста в нейросетевой системе обработки информации, которая включает звенья автоматического распознавания и контроля сигналов изображений. Методы и алгоритмы реализуются в структуре системы контроля достоверности информации на основе определения принадлежности признаков и классификации объектов метатекста СЭД по нечеткой семантической гиперсети.

7. Модели, алгоритмы контроля достоверности информации, основанные на методы словарного, статистического и хеш-кодирования, обеспечивают эффективное применение программно-аппаратной среды параллельных вычислений NVIDIA с использованием стандартных библиотек численного анализа, оптимизированного обмена данными между CPU и GPU.

8. Разработаны архитектура и идентифицированы способы описания программной оболочки системы контроля орфографии на основе древовидного представления n-граммной грамматики в среде фреймворк Сфинкс-4, реализующей программные модули для различных языковых моделей, использующих PR-избыточность. Созданный вариант системы снижает число необнаруженных ошибок, стоимость реализации и трудоемкость по сравнению с системой контроля орфографии на основе морфологического анализа до двух порядков.

9. Разработанные моделирующие алгоритмы, комплексы программных средств и системы контроля достоверности информации на основе использования PR-избыточности нашли практическое применение в системах автоматизированной организации учебной среды в ВУЗах; адаптированной передачи, обработки и анализа данных в инфокоммуникационных сетях; СЭД предприятий и учреждений.

10. Разработанные программные комплексы контроля достоверности информации в СЭД и компьютерная система адаптированной передачи, обработки, анализа данных внедрены в производственных условиях в Самаркандском филиале компании «УзТелеком» Государственный Комитет связи, информатизации и телекоммуникационных технологий и в совместном предприятии «Лаззат-Мева», полученные соответствующие акты о внедрении подтверждают экономическую эффективность результатов диссертации.

**SCIENTIFIC COUNCIL 16.07.2013.T/FM.29.01 at TASHKENT
UNIVERSITY of INFORMATION TECHNOLOGIES and NATIONAL
UNIVERSITY of UZBEKISTAN on AWARD of SCIENTIFIC DEGREE of
DOCTOR of SCIENCES**

**STATE COMMITTEE FOR COMMUNICATION, INFORMATIZATION
AND TELECOMMUNICATION TECHNOLOGIES
OF THE REPUBLIC OF UZBEKISTAN**

TASHKENT UNIVERSITY of INFORMATION TECHNOLOGIES

AKHATOV AKMAL

**METHODS, SOFTWARE AND ALGORITHMIC COMPLEXES OF
PROVIDING INFORMATION AUTHENTICITY IN ELECTRONIC
DOCUMENT INTERCHANGE SYSTEMS ON THE BASIS OF PR-
REDUNDANCY**

**05.13.01 – System analysis, management and information processing
(technical sciences)**

ABSTRACT OF DOCTORAL DISSERTATION

Tashkent – 2014

The subject of doctoral dissertation is registered on № 20.02.2014/B2013.1.T1 at the Supreme Attestation Commission of the Cabinet of Ministers of the Republic of Uzbekistan.

Doctoral dissertation is carried out at Samarkand State University.

The full text of doctoral dissertation is placed on web page of Scientific council on award of scientific degree of doctor of sciences 16.07.2013. T/FM.29.01 at the Tashkent university of information technology and National University of Uzbekistan to the address www.tuit.uz.

Abstract of dissertation in three languages (Uzbek, Russian, English) is placed on web page to address www.tuit.uz and Information-educational portal "ZIYONET" to address www.ziyonet.uz

Scientific consultant: **Jumanov Israil Ibragimovich**
doctor of technical sciences, professor

Official opponents: **Nabiyev Ozod Malikovich**
doctor of technical sciences, professor

Nishanov Akram Xasanovich
doctor of technical sciences, professor

Zaynidinov Xakimjon Nasridinovich
doctor of technical sciences, professor

Leading organization: Tashkent State Technical University

Defense will take place «28» June 2014 at 10⁰⁰ at the meeting of scientific council number 16.07.2013.T/FM.29.01 at Tashkent University of Information Technologies and National University of Uzbekistan. (Address: 100202, Tashkent, 108, Amir Temur str. Ph.: (99871) 238-64-43; fax: (99871) 238-65-52; e-mail:tuit@tuit.uz).

Doctoral dissertation can be reviewed in Information-resource centre at Tashkent university of information technology (registration number 01). Address: 100202, Tashkent, Amir Temur str., 108. Ph.: (99871) 238-65-44.

Abstract of dissertation sent out on « » May 2014 year
(mailing report № 1 on « » May 2014)

X.K.Aripov
Chairman of scientific council on award of scientific degree of doctor of sciences D.T.S., professor

M.S.Yakubov
Scientific secretary of scientific council on award of scientific degree of doctor of sciences D.T.S., professor

T.F.Bekmuratov
Chairman of scientific seminar under scientific council on award of scientific degree of doctor of sciences, D.T.S. academician

ANNOTATION OF DOCTORAL DISSERTATION

Topicality and demand of the theme of dissertation. In world practice of informatization, systems of electronic document interchange (EDIS) are considered and introduced last years not only as systems for automation of manage processes but also as high-grade platforms for creation of uniform information field, so borders of their use, certainly, expands and scientists of the world raises interest to them. In researches of leading scientists of infocommunication technologies the demand of problems of authentic gathering, transfer, analysis, coding of information during formation of office-work documents is allocated for getting of effective technologies raising mobility and productivity of EDIS.

Complex measures undertaken by the Government of Republic of Uzbekistan on development of systems of region-territorial automated management and to creation of uniform information field are directed on wide introduction of information systems, EDIS, databases (DB) on the basis of modern information-communication technologies. In this connection, working out of new methods to intelligent processing of information resources used for improvement of data transfer quality, allowing effectively to find out and correct errors in structure of electronic document interchange with least material and time expenses, differs a special urgency and, at the same time, remains the unresolved theoretical and applied problem having important economic value.

Requirements to information resources and streams of data transmission as the important factor of efficiency and quality of EDIS functioning are expressed in providing of stability, integrity, safety and authenticity of the information. One of important among factors is the criterion of authenticity of the information, caused by distortion of transferred messages in infocommunication networks because of failures and refusals of means, any handicaps in communication channels, errors of operators, scanning and recognition systems.

Hence, construction of effective systems to control of information authenticity during transfer and processing represents special scientific interest as priority technology of data processing, characteristic for conditions of automated management and electronic document interchange at the enterprises and organizations.

For existing methods, despite of providing the high level of information transfer authenticity, some unsolved questions is typical, and as basic of them it is possible to allocate the following: at development technologies of electronic documents interchange in structure of modern data transfer packages the significant volume of resource is spent for headings, at the same time the most part of information in headings remains constant from package to package during stream of transfer of the whole frame (redundancy of the information arising at it and mechanisms to provide reliability of deliveries consisting, basically, in sending the

appropriate message and repeated transfer of packages, - result to additional time and material expenses while errors detection and elimination); code and hardware methods of information transfer control are focused, mainly, on elimination the transposition mistakes in managing fields of packages; however, at data transfer there are also distortions in information fields, which frequently reveal as multiple text mistakes.

In this relation, the decision of tasks for providing of information transfer authenticity is reasonable to consider in two aspects.

Solutions of the first type tasks should taken into account errors of the man-operator, scanning and other devices intended for input of the information. These kinds of mistakes make greatest volume of distortions in text and arise at Applied and Representation levels of OSI model (Open System Interconnection reference model).

Solutions of second type tasks, devoted to control of the information reliability, take into consideration the probability of distortions which occur at stages of Transport, Network, Physical and Line of model OSI.

Demand of the dissertation is characterized by the fact that introduction of a wide range of IP-enabled technologies in electronic document interchange is connected to requiring close attention tasks of detection and correction of errors during preparation and processing of documents.

This research work is focused on providing realization of laws of the Republic of Uzbekistan «On informatization», «On electronic digital signature», «On electronic document», «On electronic commerce», «On electronic payments», Decree of the Cabinet of Ministers of the Republic of Uzbekistan № 126 on 05.04.2011 «About measures on installation and use of a single secured e-mail and system of electronic document interchange in the executive office of cabinet of ministers, bodies of government and economic governance, local government».

Following that, the solution of listed tasks requires carrying out the special researches and development connected to creation of methods and algorithms, capable to control the information authenticity in structure of data transfer packages at the expense of use the enclosed redundancy, and able to function in transport environment, eliminating existing lacks. This fact causes necessity of allocation of a special class algorithms for providing of information authenticity on the basis of new type of PR-redundancy (property redundancy), defined by depending on properties of processed object.

Conformity of research to priority directions of development of science and technologies of the Republic of Uzbekistan. The dissertation is carried out according to priority directions of science and technologies development: STP-17 – «Working out of modern information systems, means of intellectual control and training, scientific and technical DB and software products providing wide development and introduction of information and telecommunication technologies»; ASP-5 - «Working out of

information technology, telecommunication networks, hardware-software means, methods and systems of intellectual management and training, directed on increase level of society informatization».

Review of international scientific researches on dissertation subject.

Researches on working out of means to automated control of data transfer and processing authenticity taking into account features of a natural language are especially intensively conducted and certain successes are reached by scientists of USA, Great Britain, Russia, Germany, China, Austria, Sweden, Norway, Finland, Turkey, Israel, Canada, Korea, countries of Latin America (Brazil, Argentina, Chile) etc.

Publications of international expert companies, such as Gartner, International Data Corporation, Wobot, Forrester, CNews Analytics, YUNIPRAVEKS, PROGNOZ, Interface IT, Intellect Telecom and others notice that development of EDIS technologies is carried out in a direction of integrating methods of documents processing, branched management by WEB and automation of office processes. Along with it, experts allocate the urgency and demand of joining to them before independent technologies of automated detection and correction of errors during input, transfer and information processing in electronic documents, control and correction of spelling of natural languages, as toolkit of decrease human's factor influence.

Along with it, while revealing requirements of network services users, researches devoted to development of new directions and applications, based on technology of data transfer are actively carried out. Analytical studying of the scientific and technical literature show that doubtless advantage of electronic document interchange, along with application of the electronic digital signature or technologies guaranteeing authenticity and integrity of the data, is possibility of detection and correction of errors which are made during preparation and transfer of electronic documents, at the expense of methods of automation information control and decrease influence of the human factor.

Degree of study of problem. The methodology of researches on problematic, devoted to increase the reliability of information transfer, consists of advanced methods of the coding theory and using by hardware the artificial redundancy of information, which were described in a plenty of monographics, theses of researches and scientific clauses. At the same time, the essential feature of transfer and processing of the text information in EDIS is to use in them documents and messages containing artificial, natural, statistical and structural-technological redundancy.

Results of researches in the new direction to solving the problem of use of redundancy with various nature assume working out of a wide spectrum of methods, algorithms and program complexes to providing of authenticity of EDIS information resources during transfer and processing which scientifically-methodical bases are intended for development of existing technologies.

Connection of dissertational research with the plans of scientific-research works is reflected in following projects:

State scientific technical program (SSTP)-4 – «Development of software and methods of informatization, systems of management automation and regulation of technological processes» (2000-2002),

SSTP-20 «Development of modern information systems, tools of manage, databases and software ensuring wide development of information and telecommunication technologies» (2003-2005).

Purpose of research is development of constructive methods, models, algorithms and systems of information authenticity control during transfer and processing of the data on the basis of mechanisms used PR-redundancy of various nature, and software and algorithmic realization of results for developing technologies of electronic document interchange.

To achieve this goal the following **tasks of research** is formulated:

choice of methodology to creation and development of methods, software and algorithmic complexes and systems for control of the information authenticity at applied and network levels of model OSI at the expense of use information PR-redundancy of various nature;

development of methods and algorithms of information authenticity control at the expense of use the artificial, natural and semantic PR-redundancy based on mechanisms of summation, arithmetic and statistical coding;

creation and application of mechanisms of parsing codes, n-grams description, statistical recognition and hash-coding of text elements for development of methods and algorithms of information authenticity control;

development of methods, algorithms and software complexes to texts authenticity control on the basis of the multilevel morphological analysis model and n-gram structural representation of mistakes; modeling of the basic components and construction of systems to spelling control and correction; formalization of an estimation of spelling control quality;

formation and realization of structural-technological PR-redundancy based on enclosed logic criteria, databases and knowledgebase in structure of the built-in expert systems during development of methods and algorithms of the information authenticity control;

creation of system of interactive processing of information resources on the basis of synthesis of text authenticity control algorithms, experimental check and analysis of system effectiveness, calculating economic efficiency of realization of researches results in telecommunication systems and EDIS.

Object of research are information resource and technologies of gathering, formalization, entering, transforming and processing of information and documents in electronic document interchange systems.

Point of research - methods, algorithms and systems for controlling and maintenance the text information authenticity based on use of various nature information PR-redundancy.

Methods of research. In research methods of the system analysis, theories of information transfer, network technologies, probability, recognition of images, computer linguistics and also methods of algorithming, simulation and object-oriented programming are used.

Scientific novelty of dissertational research consists in the following:

concept, methodology and software and algorithmic bases to construction methods, models and algorithms for the information authenticity control in systems of electronic document interchange are developed, classes of objects characterized by PR-redundancy, applied to provide accuracy, integrity, efficiency, compression, availability of information resources in EDIS are allocated;

methods and algorithms are offered for control of information reliability at the expense of use the artificial redundancy on the basis of linear, modular, plane summing mechanisms and definition of belonging to the coded subsets;

methods and software complexes are developed for control of information reliability at the expense of use the natural redundancy on the basis of algorithms in which procedures of statistical, arithmetic, parsing coding, n-gram structured description, statistical pattern recognition and hashing of text elements are realized;

for control and correction of spelling mistakes in texts on Uzbek language methods and algorithms are offered on the basis of models of multilevel morphological analysis and n-grams Grammatik description;

on the basis of enclosed logic criteria, database and knowledgebase in structure of the built-in expert systems methods and algorithms are developed for the control of information authenticity at the expense of use the structural-technological PR-redundancy;

methods are offered to synthesis algorithms of text information reliability control in interactive system of errors detection and correction for developing technologies of electronic document interchange.

Practical results of research consist in the following:

systems of increase information authenticity on the basis of synthesis methods using artificial, natural, statistical and structurally-technological PR-redundancy are developed to supplement existing code both hardware methods and to provide text information authenticity at applied and network levels of OSI model.

efficiency of developed software complexes to information authenticity control on the basis of criterion of undetected errors probabilities is shown, it is established that they find out 92 % all kinds of errors and capable to correct single, double and adjacent transpositional mistakes, reduce labour input and cost of control in 2-3 times, raise information authenticity up to three orders.

systems to control and correction of spelling errors in Uzbek texts on the basis of morphological and n-gram structured models, and also

methods of dictionary, statistical and hesh-coding are developed and their realization is offered in the hardware-software environment of parallel calculations NVIDIA with use of standard libraries of numerical analysis, optimized data exchange between CPU and GPU.

the architecture of the system to control of text information authenticity is realized on the base of framework Sfinks-4 for modules with various language models and it is defined that the constructed variant of system considerably reduces number of undetected errors and realization cost both labour input are decrease twice in comparison with the system of spelling on the basis of morphological analysis.

Reliability of obtained results is proved by mathematical research of offered models, comparative analysis of obtained formulas and calculations with real and experimental data on the basis of the standard criteria. For estimation of researches results on their logical consistency the overall performance of program complexes is tested on the basis of voluminous context of the popular art and technical literature, the comparative analysis of efficiency of program complexes by criteria of labour input, realization costs and probability of undetected errors is carried out for a wide class of algorithms to control of information authenticity.

Theoretical and practical value of results of research. The theoretical value of work consists that the developed methods, model and the algorithms on the basis of principles of use the PR-redundancy of various nature – artificial, natural and statistical, structural-technological, programm-temporary, in aggregate form the concept and software and algorithmic bases of designing the systems to increase the information reliability by synthesis of program methods. Such methods successfully add existing code and hardware methods and ensure authenticity of text information in EDIS during transmission at applied levels of OSI model.

The practical importance of work consists in applying PR-redundancy for software realization of models of multilevel grammar description and n-gramm structured representation of the text, procedures of control and correction of Uzbek spelling that allows to calculate and effectively to use n-gramm parameters, to generate rules of information authenticity control on mechanisms of parsing recognition and coding, to find out and eliminate errors at applied and representation levels of network model OSI. Introduction of software complexes for control of the information authenticity in the EDIS structure leads to reduction time and material expenses on input of the document and accelerates processes of information interchange due to automatic detection and correction of errors which are caused by the human factor.

Realization of results. Developed models, algorithms, software complexes to providing accuracy, integrity, operation and compression of information resources at the expense of classes of objects with PR-redundancy are introduced in systems of adapted transfer, processing and analysis of the data at Samarkand branch of company «UzTelecom» of

State committee of communication, information and telecommunication technologies of the Republic of Uzbekistan (BE «SamTelekom», 23.05.2010). Results of system of adaptive transfer, processing and analysis of the data are introduced in activity of joint venture "Tasty-Fruit" (the certificate of JV «Lazzat-Meva», 14.05.2010) with annual economic efficiency of 62,2 million sum.

Approbation of work. Results of research are approved at 30 scientifically-practical conferences, including 14 international symposiums, congresses and seminars, in particular, on: International scientific and practical conference «Innovation» (Tashkent, 2004-2009); International conference «Urgent problems of applied mathematics and information technologies - Al-Khorezmi-2009» (Tashkent, 2009), International Conference In Central Asia On Internet - «ICI-2006», «ICI-2007» (TASHKENT, 2006-2007); International Conference on Electronics, Information and Communication – «ICEIC 2008» (Tashkent, 2008), X International scientific conference of universities of telecommunications (Russia, Rostov-on-Don, 2008), World Conference on Intelligent Systems for Industrial Automation – «WCIS-2008», «WCIS-2010», «WCIS-2012» (Tashkent, 2008-2012); The Third Congress of the World Mathematical Society of Turkic Countries (Kazakhstan, Almaty, 2009); Republican Scientific conference «Current status and the development of information technologies» (Tashkent, 2006-2008), Scientific conference «Current problems of mathematics, mechanics and information technologies» (Tashkent, 2008); Scientific and practical conference «Actual problems of mathematics and computer science» (Samarkand, 2007-2010), and also at annual Republican Fairs of innovation ideas, technologies and projects (Tashkent, 2008-2011).

Publication of results. It is published 87 proceedings, including 9 scientific papers in the international journals on a dissertation subject.

Structure and volume of dissertation. The dissertation consists of Introduction, six Chapters, Conclusion, References and 11 appendix, contains 197 pages of the text, 47 figures and 12 tables.

MAIN CONTENTS of DISSERTATION

In introduction the urgency and demand of the theme of dissertation is proved, the purpose and problems, and also object and an object of research are formulated, conformity of research to priority directions of development of science and technologies in the Republic of Uzbekistan is stated, scientific novelty and practical results of research are stated, reliability of obtained results is proved, the theoretical and practical importance of obtained results is reveals, the list of introductions in practice of research results, data on published works and dissertation structure are given.

In the first chapter of dissertation theoretical-methodical bases of code and apparatus methods of information transfer are investigated; tasks of designing

methods, models and algorithms are defined for control the information transfer reliability by principles of use of PR-information redundancy – redundancy of a various nature: artificial, natural, statistical, structural-technological and program-temporary, which reflects the concept of construction the system for raising the authenticity of information transfer.

In the work the positional scheme are showed to reflect a principle of use of code redundancy which is distributed also to methods of Reed-Solomon, BCH, FEC, used for the control and correction of mistakes in overwhelming majority of applications in the field of digital communications. The system analysis of principles and mechanisms used code redundancy in algorithms for detection and correction of mistakes allowed to determine, that they are focused on detection of single, double, threefold mistakes that occur in artificially entered control digits. This fact has caused carrying out researches directed on expansion of code opportunities and apparatus methods for the mistakes control.

Hypothesis of the research and development directed on creation the methods and algorithms of information reliability control in structure of data transfer packages is to use the PR-redundancy, i.e. redundancy of various nature - artificial, structural-technological, program-temporary, statistical (natural) to increase authenticity of text information during transfer and processing.

The main approaches are developed to synthesis of methods, algorithms and monitoring systems for control of text information authenticity. The sequence to performance the tasks of researches is illustrated in the form of scheme for constructing structural and functional components of the automated systems for control of information authenticity, particularly for texts (ASCTIA).

It is accepted that for solving putting problems the main criterion of efficiency for designed methods and algorithms is the minimization of probability of undetected mistakes at the expense of use the PR-redundancy up to allowable limits of information transfer reliability.

To estimate the probability of reception of code word S_0 the formula is offered as:

$$P(S_0/S_1') = \sum_{i=1}^{2^m-1} (1 - P(S_0))^{-1} P(S_i/S_i'),$$

where $P(S_0)$ is the probability that in an initial code word is a mistake;

$P(S_i/S_1')$ is conditional probability of reception S_i -th code word when S_1' has been accepted;

m is the number of code bits in the word.

Expression to estimate the probability of incorrect reception of a code word:

$$P_H(S) = \sum_{i=0}^{2^m-1} P(S_i) \sum_{\substack{j=0 \\ j \neq i}}^{2^m-1} P(S_j) = \sum_{j=0}^{2^m-1} P(\Lambda_j),$$

where Λ_j is a subset of elements of incorrectly accepted code words.

The probability of undetected mistakes is estimated as:

$$P_{rem}(S) = 1 - (1 - P_{\alpha_i})^m - \sum_{j=1}^{2^m-1} P(\Lambda_j).$$

The scheme are offered for formalization of probability processes during information transfer and diagrams of probability transitions are constructed for getting the parameters of authenticity at batch data transfer and parameters of information authenticity in units of multilevel OSI model.

The technique are developed for determining the minimum of additional information required to detect the data transmission errors. For any code method used in data transfer packages for detect errors in the message S_i , provided that the accepted mark S'_k is known, the amount of information is defined as follows

$$\begin{aligned} J(S_i/S'_k) &= m + \frac{1}{2^m} \sum_{k=1}^{2^m} \sum_{i=1}^{2^m} P(S_i/S'_k) \log(S_i/S'_k) = m + \sum_{i=1}^{2^m} P(S_i/S'_k) \log(S_i/S'_k) = (1) \\ &= m + Q' \log Q' + P'_H \log P'_H(S). \end{aligned}$$

Transformations of expression (1), having accepted $Q'(S) = Q(S)[1 - P_0(S)]^{-1}$ and $P'_H(S) = P_H(S)[1 - P_0(S)]^{-1}$ reduce to:

$$\begin{aligned} J(S_i/S'_k) &= m + \frac{Q(S)}{1 - P_0(S)} \log Q(S) - \frac{Q(S)}{1 - P_0(S)} \log[1 - P_0(S)] + \frac{P_H(S)}{1 - P_0(S)} \times \\ &\times \log P_H(S) - \frac{P_H(S)}{1 - P_0(S)} \log[1 - P_0(S)] - \frac{P_H(S)}{1 - P_0(S)} \log(2^m - 1) = m - \frac{P_H(S)}{1 - P_0(S)} \times \\ &\times \log(2^m - 1) - \log[1 - P_0(S)] + \frac{Q(S)}{1 - P_0(S)} \log Q(S) + \frac{P_H(S)}{1 - P_0(S)} \log P_H(S) \end{aligned}$$

where $P_0(S)$ is probability of detecting the mistakes in message S ;

$Q(S)$ is probability of correct reception of the message S .

The value of $J(S_i/S'_k)$ determines a minimum of additional information about the message, necessary for elimination of losses due to undetected mistakes.

One more result in present chapter is obtaining the expression to simplified estimation of redundancy required for estimation of efficiency of all code methods

$$R(S_i) = 1 - \frac{1}{(1 - P)} \log \frac{(2^m - 1)}{nP}.$$

In the work it is shown, that the developed technique allow to estimate the efficiency of program ways for control of information authenticity based on use of PR-redundancy.

In the second chapter of dissertation methods and algorithms are developed to control the authenticity of information in structure of data transfer packages by principles of using the artificial redundancy.

Methods and algorithms are developed for control of the information authenticity on the basis of using the artificial redundancy by mechanisms of linear, plane and modular summation which eliminate lacks of traditional program methods. Algorithms of control of information authenticity by summation can establish not only presence of error in the accepted sequence, but also specify a position number of mistakes for their automatic correction. Algorithms are based on the following rules: π_1 for control of information transfer authenticity; π_2 for detecting and correction mistakes of type Q_1 ; π_3 for detecting and correction mistakes of type Q_2 ; π_4 for detecting mistakes of type Q_3 ; π_5 for detecting mistakes of type Q_4 .

The technique for realization the general and private solution are developed for tasks of estimation the efficiency of information authenticity control by criterion of undetected mistakes for various algorithms.

The total probability of undetected mistakes for rules to control the information authenticity by linear summation is determined by formula

$$P_H = P_1 + P_2 + P_3 + P_4 = \{67[mnp^2(1-p)^{m+n-2} + C_m^2 C_n^2 p^4 (1-p)^{m+n-4} + C_m^k C_n^3 p^{3k} (1-p)^{m+n-k-3}] + 5,76 \cdot 10^{-6}\}.$$

In the work the diagrams of algorithms efficiency are investigated by criteria of control laboriousness and cost of implementation. When using algorithms of linear and plane summation (Π_3 and Π_6) in comparison with other methods laboriousness and cost of the control is reduced in 2-3 times.

The diagrams of dependence of undetected mistakes probability P_H by control algorithms from probability of mistakes P show, that when probability of mistakes is in limits $10^{-3} - 10^{-4}$ the difference of ability of mistakes detection for investigated algorithms achieves up to two orders, the laboriousness and cost of control is reduced in 2-3 times at the expense of automatic correction of mistakes.

The greatest effect on parameter of information authenticity is achieved for algorithms: Π_2 – digit-by-digit summation, Π_3 – numeral summation, Π_6 – plane summation.

The analysis of efficiency of algorithms of digital and numerical modular summation on basis of modules 8, 9, 10, 11 and modules q showed that the greatest effect on parameter of information authenticity is achieved for algorithms of numeral modular summation on modules 8, 10 and 11.

It is established, that the developed algorithms find out up to 92% of all kinds of mistakes and correct single and double adjacent transposition mistakes.

The task is solved for definition the probability of mistakes of the first and second sorts of algorithms. At realization the algorithm to control of single mistakes following formulas are got:

For mistakes of first sort:

$$P_1 \approx 2P_{\alpha_i} (1 - P_{\alpha_i}) \cdot 10^{-2} + P_{\alpha_i}^2 \cdot 10^{-2} + 2P_{\alpha_i} (1 - P_{\alpha_i}) \cdot 10^{-4}.$$

For mistakes of second sort:

$$P_2 \approx (1 - P_{\alpha_i})^2 \cdot 1.5 \cdot 10^{-5} + P_{\alpha_i} (1 - P_{\alpha_i}) \cdot 1.5 \cdot 10^{-5}.$$

Total probability of undetected mistakes for algorithm is

$$P_H = P_1 + P_2 \approx 2P_{\alpha_i} (1 - P_{\alpha_i}) \cdot 10^{-2} + (1 - P_{\alpha_i})^2 \cdot 1.5 \cdot 10^{-5}.$$

By the submitted technique of solving tasks by criterion of undetected mistakes probability the efficiency of algorithms constructed for the control and detection of mistakes of all other kinds is analyzed.

In the third chapter of dissertation methods and algorithms are developed to control of information authenticity by principles of use the statistical, structural-technological, program-temporary redundancy.

It is determined, that the algorithms based on mechanisms of use the logic, technological, statistical connections and properties of information create favorable conditions for control of data and provide high reliability of information transfer.

The principles are formulated to constructing the software and algorithmic complex of information control on the basis of databases and knowledgebase (DB and KB), database management systems (DBMS) and expert systems.

The mechanisms to control the information authenticity are developed by application the statistical redundancy which is traditionally used in tasks of coding and compression of transmitting information.

For algorithms, which designing for control of text elements images the generalized model of estimation of undetected mistakes probability are offered in follow view:

$$P_H = P_1 W_1 + P_2 W_2,$$

where W_1, W_2 are losses coefficients, connected to mistakes of the first and second sort, which are set in an interval of real numbers $[0,1]$.

Thus, all elements of the controllable text are shared on two classes: authentic and deformed data.

When number of the first class elements N_1 is large the averaging of probability gives

$$x_1 = N_1 p_1 (1 - P_F) W_1 + N_1 (1 - p_1) P_F W_2,$$

where P_F is probability of that the element is referred to an incorrect class.

Averaging on elements N_2 of the second class results to

$$x_2 = N_2 p_1 P_F W_1 + N_2 (1 - p_1) (1 - P_F) W_2.$$

Calculation of the weighed difference gives

$$\frac{x_1}{N_1} - \frac{x_2}{N_2} = p_1 (1 - 2P_F) W_1 - (1 - p_1) (1 - 2P_F) W_2.$$

Estimating sum and difference of p_1W_1 and $(1-p_1)W_2$ for any $P_n \neq 1/2$ allow to derive W_1 and W_2 :

$$W_1 = \frac{x_1N_2 - x_2N_1}{N_1N_2p_1(1-2P_F)}; W_2 = \frac{x_1N_2 - x_2N_1}{N_1N_2(1-p_1)(1-2P_F)}.$$

One more result of issue are the developed algorithms to control of information authenticity based on use of compression mechanisms such as methods of Huffman, arithmetic coding, transformation of Burrows-Wheeler, Lempel-Ziv and their modification.

In the offered methods a key moment is the definition of bottom and top probability borders for the control of permitted code combinations and the estimation of most probable elements of messages given for the control.

To illustrate this phenomenon the curve of function for probabilities distribution density of transmitted information α depending on received information β are constructed. These curves are the diagrams of functions $\varpi(\alpha)$ and $\varpi(\beta/\alpha)$ for conditional distribution at normal law.

Practical realization of developments show that then limits of possible deviations are more narrow from controllable value, than more mistakes of the first sort are detectable by method of two-dimensional regress or then interrelation between controllable parameters more closely, than value of parameter is rapprochement to true value.

The general and private solution have been derived of tasks consisting in finding of optimum borders of conditional moment to minimize total probability of mistakes during control of the information which calculated by formula

$$P_H = P_1 + P_2 = \frac{P}{B} \int_{x(\alpha)}^{y(\alpha)} W(\alpha) d\alpha + (1-P) \left(1 - \int_{x(\alpha)}^{y(\alpha)} W(\alpha) d\alpha\right).$$

Thus, the estimations of information authenticity are obtained on the basis of finding the threshold of difference for functions of distributions $W(\alpha)$ and $W(\beta)$ by conditional probability of distribution $W(\beta/\alpha)$.

The task of information authenticity control are solved also on the basis of structural redundancy from the containers of images in data transfer package.

The following designations for procedures of including and extracting are entered: $E(V, W) = V_W$ is a procedure of inclusion the structural redundancy in container of images; $D(V, V_W) = W'$ (or $D(V_W) = W'$) is a procedure of extraction of structural redundancy when needed; α is a scaling coefficient; C is a binary attribute of similarity for two signals: $C = 1$ if coefficient of cross-correlation is more than some threshold δ ; else $C = 0$.

The following variants are proved to working algorithms of control the information authenticity on the basis of structural redundancy.

In the first variant the algorithm of control $A_1(E, D, C)$ is based on strictly convertible process. Thus it is supposed that for any V_W is a display E^{-1} for which the follow conditions are true: $E^{-1}(V_W) = (V_F, W_F)$ and $E(V_F, W_F) = V_W$; E^{-1} is computationally feasible; W_F belongs to a class of structural redundancy allowing to take true and false images; $C(D(V_W, V_F), W_F, \delta) = 1$.

In the second variant the algorithm $A_2(E, D, C)$ is based on poorly convertible process. Thus it is required condition $E(V_F, W_F) = V_W$. However, note that even $E(V, W) = V_W$ can not be true by virtue of various sort of distortions V_W . On the other hand, this requirement is too weak for definition of convertibility. In this connection it is replaced with the requirement $E(V_F, W_F) = V_{W'}$, where $C(V_{W'}, V_W, \delta) = 1$.

Another condition of registration the weak convertibility of process in algorithm $A_2(E, D, C)$ is that for each V_W there is a display E^{-1} , where $E^{-1}(V_W) = (V_F, W_F)$ and $E(V_F, W_F) = V_{W'}$. Thus E^{-1} is computationally feasible, $C(V_{W'}, V_F, \delta) = 1$, $C(V_{W'}, V_W, \delta) = 1$ and $C(D(V_{W'}, V_F), W_F, \delta) = 1$.

For using the structural-technological PR-redundancy at the control of information authenticity, algorithms are developed based on built-in expert systems, knowledge bases, procedures of logic control and structures of data. Principles of choice of characteristics, criterion and place of realization of decision rules in structure of examined EDIS are offered. Formulated principles and criteria serve as bases at constructing algorithms and program modules for systems to control of information authenticity in EDIS.

The combined approach is offered to using the program and temporary redundancy for realization of tools of automatic control and restoration the data and to functioning of protection tools on the basis of PR-redundancy. For organizational and operative methods to control of information authenticity in EDIS is carried out the classification.

The system of information authenticity control is constructed on the basis of programming object-oriented technology and includes thirteen functional and eighteen applied program modules which are steady to various perturbations and are capable to save sufficient quality of results in all real conditions.

For developed methods and algorithms to control the authenticity of information, the comparative analysis of efficiency are carried out. Rapprochement of experimental values P_H^* to analytical, getting in theoretical sections of dissertation are checked up by Pearson criterion \aleph^2 with guarantee probability 0.9, that confirms good conformity of experimental results to theoretical.

In the fourth chapter of dissertation methods and algorithms are developed to control the information authenticity by principles of use the statistical redundancy on the basis of n-gram language models, models of parsing code and recognition of text elements.

Processes of designing of distribution laws simulated for n-gram statistics serves as the basis for algorithms developed to: revealing of deformed text elements; clustering, search, structuring; getting the frequency characteristics which realized at constructing the system of mistakes control and correction.

It is revealed, that the job of control system constructed on the basis of search, recognition and clustering models is connected to attraction the large volume of memory. Using the mechanisms and systems of parallel calculation based on GPU models and CUDA technology is recommended for elimination such lacks at the control of text spelling.

The estimation of minimized memory is represented by function

$$f(x) = -\frac{nx}{\ln(1 - P^{1/x})}$$

Where n is length of codes and has fixed value;

x is variable, designating number of hash-functions.

Principles are proved, models and algorithms are offered for construction systems of information authenticity control by combine: mechanisms of dictionary, statistical and hash coding; ways, algorithms and procedures of parallel calculations.

The unified hardware-software decision for parallel calculations on video chips NVIDIA are used to constructing software-technical complex (STC) of information authenticity control (IAC). Feature of system on the basis of parallel calculations is that during its performance the task of elimination the superiority of logic flows number above physical devices.

For estimation the efficiency of control algorithms on the basis of parsing modeling formula is offered to use complete probability of allocation and recognition a line of words

$$W \doteq \arg \max_W P(W | A)P(A),$$

where W denotes a researched word as set of text elements $W = w_1, w_2, \dots, w_n$;

$P(W / A)$ is conditional probability that word in line A is represented as W ;

$P(A)$ is a priory probability of occurrence line $A = \{a_1, a_2, \dots, a_n\}$ in the text.

The job of system to control of spelling mistakes on the basis of algorithms with parsing codes are organized on the offered treelike model

The scheme of functioning of system based on parsing recognition consisting of three modules is illustrated in fig. 1.

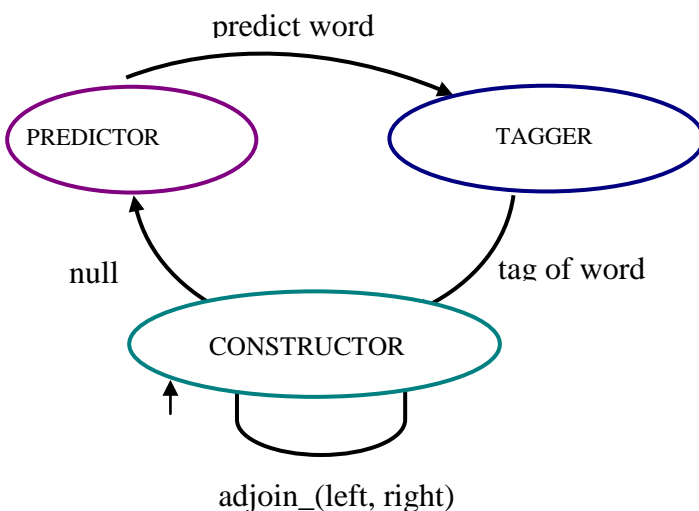


Fig. 1. Interaction of modules at parsing coding system

«Word predictor» predicts the following word w_{k+1} given by k -prefix of word-parsing, then passes control to «Tagger». «Tagger» predicts the POS-attribute t_{k+1} of the next word, given k -prefix of word-parsing and the newly predicted word w_{k+1} . Module passes control to «Constructor». «Constructor» grows the already existing binary branching structure by repeatedly generating transitions until module passes control to «Predictor».

Further in the work the algorithm of recognition on a basis of neural network (NN) with associative memory are developed for control of text elements authenticity. The process of input and control of image signal authenticity on a basis of NN is illustrated in fig. 2.

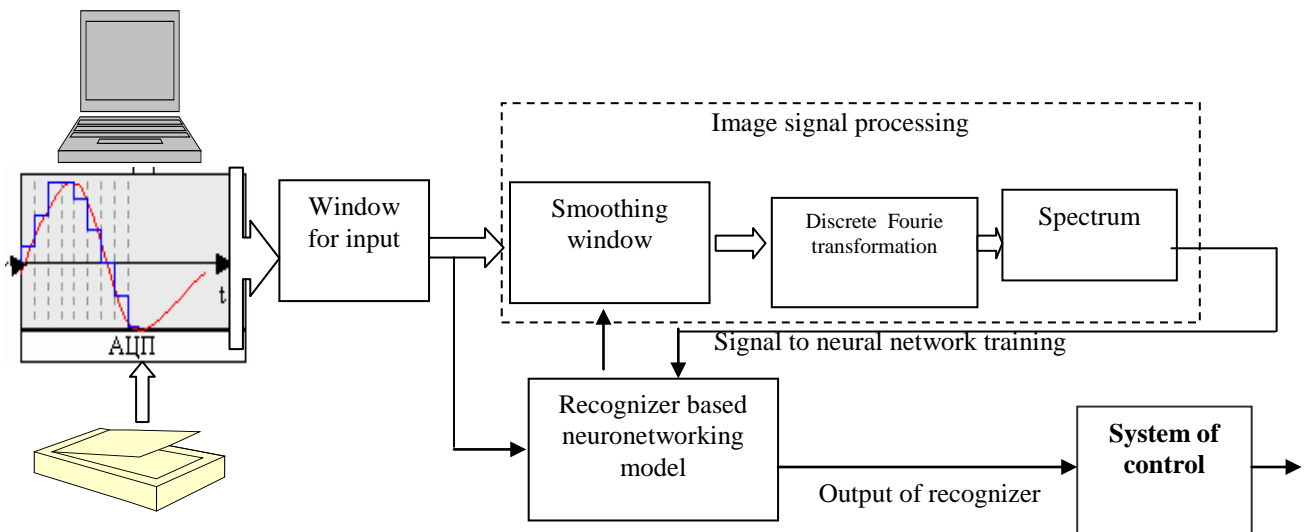


Fig. 2. Control of authenticity of text element signal

The technique is offered to estimate the probability of unauthentic recognition by comparison of recognized text element with special directories of word-forms and frequency dictionaries of monogram, bigram, trigram, k-gram.

The probability of incorrect recognition on total volume of directory is determined as

$$p(\tau) = 1 - (1 - p(a)) \prod_{i=1}^{n-1} (1 - ((1 - p(a)) + p(a \equiv a_i)(2p(a) - 1))), \quad p(a) < 10^{-3},$$

where τ is a volume of special directory for checking the authenticity of image recognition;

$p(a)$ is probability of mistakes in text elements during information transfer and processing;

n is a number of recognized text elements.

To estimate the recognition fidelity of system, where as a reference sample is used the authentic text element, the formula is obtained as:

$$p(\tau) = 1 - (1 - p(a)) \prod_{i=1}^{n-1} \prod_{\substack{j < i \\ a_i, a_j \in A}} (p(a_j \equiv a_i)),$$

where $p(a_j \equiv a_i)$ is probability of information occurrence on output of examined system of information control.

For construction the generalized algorithm of system to process Uzbek metatext the scheme is offered to definition the hierarchy of classes which represents a tree with any quantity of branching and on this basis the classification of controllable objects have been carried out. In the realized system, the necessary primary data are stored in KB and they are specified during purchase new knowledge when EDIS works. The real model is formed in KB environment, and connection with server computer is carried out through inquiries of user.

The developed system consists of the following functional parts: database that stores reference and evidence data about object, results of their comparison, conceptual, info logical and physical models of object; static knowledgebase, i.e. expert knowledge, formulas, facts, dependence, tables; dynamic knowledgebase, i.e. combined models of NN represented as reference dynamic processes, in view of partial or complete uncertainty of parameters; the module of logic conclusion based on algorithm to generating the causal network of events in functional and structural model; the module of adaptations coordinating job of DB and KB during logic conclusion depending on a usual situation; the module of explanation, which represents interpretation the process of logic conclusion; the scheduler coordinating process of task solving; solver for finding the effective solutions in the direct, return and mixed statement of tasks.

In the fifth chapter of dissertation the basic approaches and methods are investigated, and models and algorithms are developed to control and correction of spelling mistakes in the texts on Uzbek language by multilevel morphological analysis and using the description of n-gramm word-forms.

As results of analysis of literature the principles, features and lacks were revealed for the most known methods to constructing the systems of control on the basis of morphological analysis: splitting of the text into words, comparison to dictionary, approximated extraction of words roots, «merge» of text with dictionary, normalization of words.

It is determined, that considered most effective systems are based on principles of use the two-level models of word-forms analysis, and they are recommended as the base mechanism at construction of systems to control and correction of Uzbek spelling.

Hence, at construction of systems to control errors in texts on Uzbek language greatest theoretical and practical importance is represented by realisation of multilevel model of the morphological analysis on the basis of synthesis principles and possibilities of two-level models and ways of word forms analysis.

Developed in the work the multilevel graphic scheme to construct the Uzbek word-forms represents a sequence of joining the affixes to words roots.

The techniques and algorithms have been designed to formal description and analysis of Uzbek word-forms and their phonetic changes. Interrelations of Uzbek affixes are described as multilevel table allowed clearly to parse the given word from the end, by recursive approach to spread out word on levels and to be

convinced that spelling is correctness. Developed algorithms and structure of system to control and correction of spelling are based on obtained theoretical rules.

To estimate probabilities of mistakes occurrence in information the technique are devised on the basis of n-gramm model.

The average probability of mistake is calculated by averaging the conditional probabilities of a mistake on all ensemble of the messages:

$$P = \sum_{\alpha_i} P_{\alpha_i} \sum_{\alpha_j (\alpha_i \neq \alpha_j)} P_{\alpha_i \alpha_j},$$

where $P_{\alpha_i \alpha_j}$ is probability of correct reception the α_i -th message.

In connection with fact that definition of n-gramm mistakes probabilities bring to processing the statistical data with large volume and labour-consuming calculations, we carry out special modeling and automation of calculations of n-gramm mistakes probabilities statistics.

The offered technique determines the value of probability $p(\alpha)$ of occurrence n-gramms $\alpha = \alpha_1 \dots \alpha_n$, equal to the attitude $C(\alpha)$ to total number of all met in aggregate n-gram copies:

$$p(\alpha) = \frac{C(\alpha)}{\sum_{\alpha^*} C(\alpha^*)},$$

where α^* is the copy of n-gram from «general collection of all sequences», differ from α .

On the basis of conditional probability $p(\alpha_i = \alpha_j^* | \alpha_i = \alpha_i^* \forall i \neq j)$ of mistakes α_j^* the complete probability is got as:

$$p(\alpha_1^* \dots \alpha_n^*) = p(\alpha_i = \alpha_j^* | \alpha_i = \alpha_i^* \forall i \neq j) p(\alpha_i = \alpha_j^* \forall i \neq j).$$

For calculation the conditional probability of erroneous symbols in a string with view of all previous symbols the rule is offered as:

$$p(\alpha_1 \dots \alpha_n) = \left(\prod_{k=2}^n p(\alpha_k | \alpha_1 \dots \alpha_{k-1}) \right) p(\alpha_1).$$

Using Markov circuit of (n-1)-th order we got simple n-gram model of mistakes, which allowed to estimate statistical parameters of examined n-gram.

The modified variant of the system to spelling control are developed on the basis of framework Sphinx-4 technology, which allows to create modular structure of system, includes separate components devoted to solution of specific tasks and adaptation by support of context free n-gram. The shell of system and ways of description of system kernel are developed, thus it is supposed that the system functions at support of all component and program modules of framework.

The architecture of developed system on the base framework Sphinx-4 is illustrated on fig. 3, which includes three primary modules: the FrontEnd, the Decoder, the Linguist, and also auxiliary modules the Configuration and the Utilities.

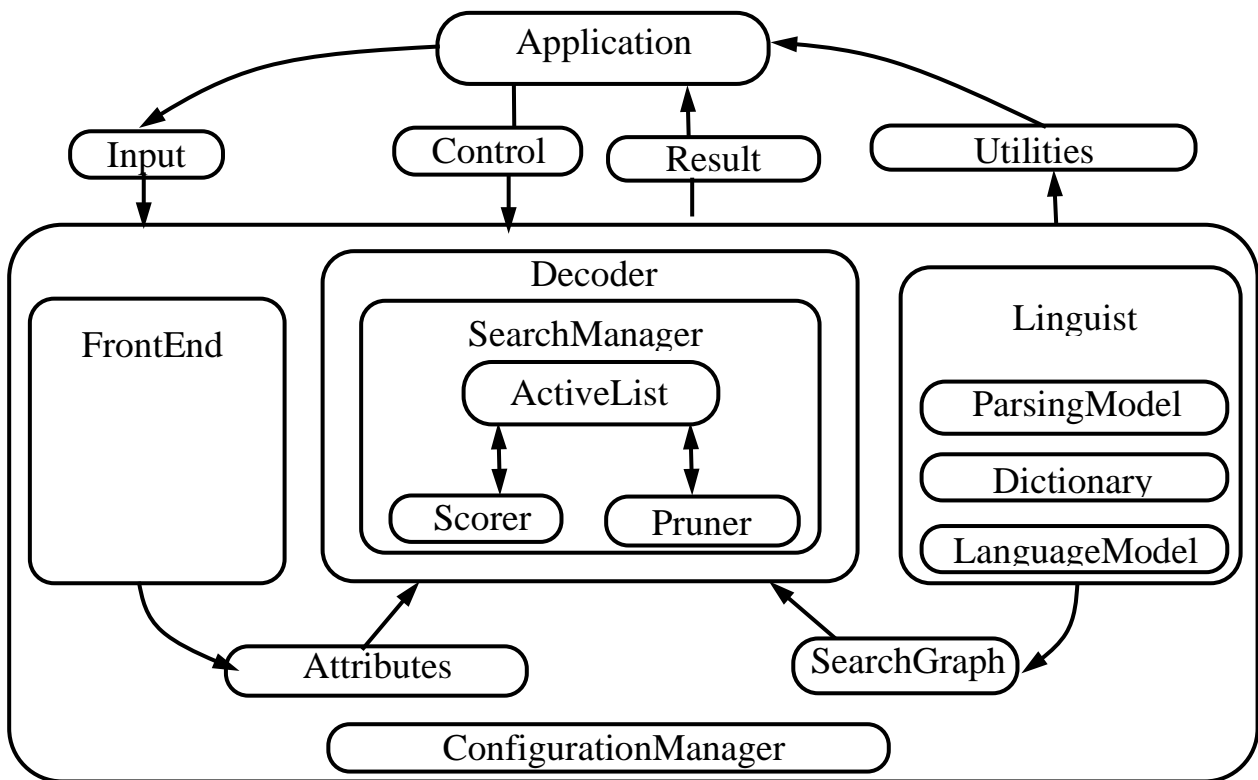


Fig. 3. Architecture of system to spelling control on the base Sphinx-4

Scheme illustrate relatives between modules and connection of modules with the basic application. In dissertation we described the functions of basic and auxiliary blocks of system: the FrontEnd, Decoder and Linguist and their interaction in framework Sphinx-4.

The function of the FrontEnd is taking one or more text elements (symbols, letters, words) on input and parameterizes them in a sequence of attributes.

The Linguist is intended to translation the standard language model with information about n-gram parsing structure from one or more sets of the Parsing-model in the SearchGraph.

The SearchManager from the Decoder uses attributes received from the FrontEnd and the SearchGraph, generated in the Linguist to performing actual decoding and generation of results.

Carried out tests show that the constructed variant of system improves a parameter “mistake on word” at recognition and speed of system measured in a mode of real time twice, in comparison with the system to control of spelling on the basis of morphological analysis. Let's note, that the work of system was organized in dual-processor mode with processors CPU 1015MHZ ULTRASPARC RIII WITH 2G RAM. Due to such use of modules which are included in the shell of Sphinx-4, adapted algorithms and stochastic models of language there is an opportunity to support a wide variety of tasks connected to the analysis of text, generation of text in speech, speech in text and other tasks of textual information processing.

For organization of convenient order of processing flows and data loadings in systems to control of information authenticity on the basis of n-gram models

we developed ways of performance the statistical information used in n-gramm Grammatik and treelike model of its description, which allows to reduce sizes of processable files by elimination of redundancy in format structure.

In the work the treelike scheme illustrates manipulation with Grammatik elements for reduction of redundancy.

The constructed format of n-gram Grammatik description allows to support the various programming languages and to carry out networking processing of information because the description of n-gramm models is carried out on XML.

In the sixth chapter of dissertation details are stated about development of classes, methods and software modules of systems to control the information authenticity on the basis of use PR-redundancy and realization them in structure of data transfer packages of systems to networking information processing and EDIS.

The complex of software tools are developed for system of electronic document circulation intended to produce, editing, storage and transfer the elements of enterprises office-work, created in integrated environment 1C 8.2. Feature of system realization is that in editors connected to input, set, transformation of information, the subsystem named CONTROL functions to control of data authenticity and it consist from the complex of program modules developed in Borland IDE Delphi.

Developed complex of programs are software which becomes attached to the basic subsystems and functional tasks of used information system.

It is necessary to note, that the subsystem of information authenticity control can be used locally as service support for any computer system to processing the text information and it may be included to structure of functional subsystems of EDIS or to structure of protocols in telecommunication systems for struggle with mistakes at Applied and Representation levels of OSI.

In examined EDIS the follow main objects are identified for the control and they carry out: full-scale search of documents by information from registration card; journalize the directory of staff positions; writing the directory of standard phrases and verbal revolutions; regular updating of the directory of organizations; authorization of employee access rights to system in binding to their functional roles, i.e. by sections of documents, card files of registration, businesses, thematic rubricates and other criteria, that allows effectively to use a method of object-oriented programming, DB and KB.

The software shell of systems to control and correction of spelling mistakes are developed in environment Sphinx-4 and on the basis of technology CUDA NVIDIA, for which the modules realizing following functions are common: module F1 – takes as input a word and analyzed it through the virtual table, module F2 – realizes a method of binary search in dictionary; module F3 – creates an array to store and load the dictionary, module F4 – realizes fast sorting; module F5 – provides replacement of objects, module F6 – checks borders of words; module F7 – performs binary search in the loaded dictionary; module F8 – checks spelling mistakes; modules F9-F15 – realize procedures on transliteration of text in systems of spelling check; modules F16-F20 – realize parsing methods for input text and carry out procedures of parsing modeling; module F21 – allows users to

work with document and edit various font properties, add, edit and delete the text and save it to a file; module F22 – MDI application realizing a word-processor for editing and formatting the Uzbek text; module F23 – provides correct closing of all documents, if whole application is closed.

For performance the functions of correction of spelling mistakes, modules based on the adaptive Huffman coder are realized. Model and algorithm of coding and compression by rule of Ziv-Lempel are used for convertible transformation. On the basis of statistical coding model of Lempel-Ziv-Welch the modified module are realized and principle of its application consists in use of dictionary, storing a lines of text and containing numbered cells. The module of control are developed on the basis of simple, self-adaptive and adaptive models of arithmetic coding.

The basic classes are devised, methods and general structure are offered for organization the functioning of system of Uzbek spelling control, certificates on registration the software are got. Uzbek spell-checker has ability to check spelling in both views (Cyrillic and Latin script) and, if necessary, to translate from one representation to another.

Experimental verification and analysis of efficiency of developed algorithms and systems to control of text information authenticity were carried out by the data on functioning systems of activity monitoring at Samarkand State University and EDIS in Joint Venture «Tasty-Fruit». The results are obtained under conditions close to practical where: volume of entered information is equal to 10^6 marks (about 300 pages of electronic text); value of mistakes probability is $P = 3,7 \cdot 10^{-3}$; the word in document is assumed equal on the average to 8 decimal marks; the whole line of document is protected by application of algorithms based on use of PR-redundancy; during calculation mistakes of processing caused by failures of electronic devices are neglected.

Table represents results of the comparative analysis of efficiency for developed algorithms by various criteria. In columns 6 and 7 we give the values of coefficient for reduction the probability of undetected mistakes errors (coefficient of advantage for information authenticity), respectively $K_H = P/P_H$ is theoretical and $K_H^* = P/P_H^*$ is experimental.

In columns 8 and 9 we give the values of coefficients $K_{TIA} = T_P/T_{IA}$, $K_{TA} = T_P/T_A$ meaning the labor input, in columns 10 and 11 values of coefficients $K_{CIA} = C_P/C_{IA}$, $K_{CA} = C_P/C_A$ for cost of control. For existing conditions of data processing on objects we considered, that the labour content of input the information by keyboard is equal 400 hours, cost of input is equal 60 thousand sum, labour content of manual control is 350 hours, cost of manual control is 45 thousand sum, labour input of scanning and recognition is 150 hours and cost of recognition is 25 thousand sum.

It is established, that using of algorithm with linear summation provides: the minimal probability of undetected mistakes $P_H \approx 3 \cdot 10^{-5}$; labour content – 0,5 hours; cost of control – 1 thousand sum.

Table

The comparative analysis of efficiency of algorithms in systems of information authenticity control

№	Algorithms to control, detection and correction of mistakes	Values of efficiency criterion							
		Probability P_H		Coefficient $K_H = P/P_H$		Labour content of control - K_T ,		Cost of control - K_C ,	
		Theor.- P_H	Exper. P_H^*	Theor K_H	Exper. K_H^*	K_{TIA}	K_{TA}	K_{CIA}	K_{CA}
Using the artificial redundancy									
1	Linear summation	10^{-5}	$2 \cdot 10^{-5}$	100	200	1,8	1,7	2	2
2	Modular summation	$10^{-5}-10^{-6}$	$5 \cdot 10^{-6}$	1000	5000	2	2	2	2
Using the natural redundancy									
3	n-gram models	$10^{-5}-10^{-6}$	$4 \cdot 10^{-6}$	1000	4000	2	2	3	3
4	Morphological model	10^{-6}	$2 \cdot 10^{-6}$	1000	2000	3	3	3	3
Using the structural-technological redundancy									
5	Expert systems and knowledge bases	$10^{-5}-10^{-6}$	$1.5 \cdot 10^{-6}$	1000	1500	1,8	1,7	2	2
6	Logical control criterion	10^{-5}	$2 \cdot 10^{-5}$	100	200	2	2	2	2
Using the statistical redundancy									
7	On the basis of Lempel-Ziv-Welch method and Burrows-Weeler	$2 \cdot 10^{-5}-10^{-6}$	$6 \cdot 10^{-6}$	2000	6000	3	3	3	3
8	On the basis of arithmetical coding	$3 \cdot 10^{-6}-10^{-7}$	10^{-7}	3000	10000	3	3	3	3

Using the algorithm with plane summation provides: the minimal probability of the undetected mistakes – $P_H \approx 4 \cdot 10^{-6}$; labour content of control – 1,2 hours; cost of control – 1,5-2 thousand sum.

Using the algorithm with modular summation provides: the minimal probability of undetected mistakes - $P_H \approx 6 \cdot 10^{-5}$; labour content of control - 0,8 hours; cost of control - 1,2 thousand sum.

It is determined, that developed program methods, in comparison with traditional methods, reduce labour content and cost of control in 2-3 times.

The attitudes of labour content and cost of manual mode of information system to labour content and cost of semi-automatic and automatic modes testify values of obtained coefficients decrease almost in 2-3 times.

Rapprochement of experimental values P_H^* to analytical is checked up by Pierson criterion \aleph^2 with guarantee probability 0,9. This confirms that experimental results match theoretical.

Technique of calculating the economic efficiency and efficiency of implementation are developed for methods, algorithms and software as computer systems of automation of office-work (EDCS) and adapted data transfer, processing and analysis in infocommunication networks. It is determined that value of estimated economic efficiency of computer system for adapted data transfer, processing and analysis in infocommunication networks in BE «SamTelecom» and in JV «Lazzat-Meva» by results of data processing represented in monthly reports makes 5,1 million sums.

CONCLUSION

1. Developed in the dissertation constructive methods, models, software and algorithms complexes to control of information authenticity by principles of using of PR-redundancy on the basis of concepts of system analysis, control and information processing allow to increase efficiency and productivity of EDIS.

2. The estimation of current state of the theory and practice of code, hardware and program methods of control of authenticity information transfer had shown insufficiency of existing types of redundancy for providing qualitative functioning of EDIS. Principles of use of PR-redundancy at electronic documents for working out applied methods, software and algorithmic complexes to providing of information transfer authenticity have allowed to design toolkit of development of existing technologies.

3. Methods of: linear, plane and modular summation; codings by rules of Haffman, Lempel-Ziv-Velch, Barrouse-Willer, arithmetic coding, statistical recognition, logic control form methodical bases of use of PR-redundancy for expansion possibilities of algorithms and software complexes to providing of the information authenticity during drawing up, transfer and processing of electronic documents texts.

4. Efficiency of developed algorithms is shown and solutions are got for tasks of control the information authenticity on the basis of criterion of mistakes undetecting probabilities. It is established, that they find up to 92 % of all kinds of

mistakes, capable to correct single, double and adjacent transpositional mistakes, in comparison with existing methods reduce labour content and cost of control in 2-3 times if probability of mistakes is accepted as $P \approx 4 \cdot 10^{-3}$, and also raise the information authenticity up to three orders.

5. To solving tasks of control and correction of spelling mistakes in Uzbek texts methods, algorithms and systems are developed included morphological and n-gram structured models. The developed technique of getting frequency characteristics of n-gram on the basis of distortions probability parameters statistics are applied during systematization hash-codes for parsing coding.

6. Interpolation and extrapolation methods of construction the logic and arithmetic function of statistical recognition are used for working out algorithms to control of text elements images authenticity. Methods are developed to control authenticity of alarm characteristics of text elements images in neuronetworking system to information processing which includes parts of automatic recognition and control of images signals. Methods and algorithms are realized in the structure system to control of information authenticity for cases when information in EDIS is represented as metatext on the basis of belonging attributes and classification of metatext on fuzzy semantic hypernet.

7. Methods and algorithms of the control of the information authenticity, based on methods of dictionary, statistical and hash-codings provide effective applying of the hardware-software environment of parallel computations NVIDIA with use of standard libraries of numerical analysis, optimized data exchange between CPU and GPU for optimization.

8. It is determined that at realization of the system to control of spelling on the basis of developed ways of description and identification of the software shell, treelike representation of n-gramm grammar and architecture of framework Sfinks-4 focused on various language models used PR-redundancy, the number of undetected errors and cost of realization considerably decreases, and labour content in comparison with the spelling control system on the basis of morphological analysis decreases twice.

9. Developed simulating algorithms, complexes of software and systems to control of information authenticity on the basis of using PR-redundancy have found practical application in systems of: automated organization of educational environment in high schools; adapted data transfer, processing and analysis in infocommunication networks; EDIS of enterprises.

10. The developed software complexes to control the information authenticity in structure of EDIS and computer system of adapted transfer, handling and data analysis are implemented in real working conditions in the Samarkand branch «UzTelecom» of the State committee or communication, for informatization and telecommunication technologies of the Republic of Uzbekistan and in Joint Venture «Tasty-Fuit». Appropriate certificates confirm cost of economical efficiency of dissertation results.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I бўлим (I часть; I part)

1. Бобомуродов О.Ж., Ахатов А.Р. Обработка информации в системах приобретения знаний. - Ташкент: «ФАН», 2009. - 136 с.
2. Ахатов А.Р. Логический алгоритм обнаружения и исправления ошибок в машинописных текстах на узбекском (латиница, кириллица), русском, английском языках // Журнал «Проблемы информатики и энергетики», Изд-во «Фан» АН РУз. - Ташкент, 2003. - №2. - с. 66-71.
3. Ахатов А.Р. Основы повышения качества переработки текстовой информации // Журнал «Вестник ТашГТУ». – Ташкент, 2003. - №4. - с. 151-155.
4. Ахатов А.Р., Джуманов О.И. Позиционное кодирование визуальной информации об объектах, представляемых для распознавания // НТЖ «Химическая технология. Контроль и управление». - Ташкент, 2006. -№ 5. - с. 48-52.
5. Ахатов А.Р. Оценка информативности признаков для построения алгоритмов распознавания текстовой информации // НТЖ «Химическая технология. Контроль и управление». - Ташкент, 2007. - № 4. - с. 88-93.
6. Ахатов А.Р. Построение интерполяционного алгоритма распознавания текстов по неполной информации // Журнал «Вестник ТУИТ». - Ташкент, 2007. - №2, - с. 98-102.
7. Ахатов А.Р., Жуманов И.И. Алгоритм контроля качества текстов в системах электронного документооборота // Журнал «Вестник ТУИТ». – Ташкент, 2007. - №2. - с. 64-68.
8. Ахатов А.Р. Оценка эффективности программных методов контроля достоверности информации на основе избыточных кодов // Журнал «Алока дунеси» – Ташкент, 2007. - №2, - с. 41-46.
9. Ахатов А.Р. Методы и алгоритмы обеспечения достоверности текстовой информации на основе статистической избыточности // Журнал «Вестник ТашГТУ». - Ташкент, 2007. - №2. - с.41-44.
10. Ахатов А.Р. Разработка экстраполяционных алгоритмов распознавания образов для контроля и коррекции ошибок // НТЖ «Химическая технология. Контроль и управление». - Ташкент, 2007. - №5. - с.41-45.
11. Джуманов О.И., Ахатов А.Р. Разработка правил обучения нейросистем для модификации алгоритмов распознавания изображений элементов текста // Журнал «Проблемы информатики и энергетики», Изд-во «Фан» АН РУз. - Ташкент, 2007. - №4. - с. 28-31.
12. Джуманов О.И., Ахатов А.Р. Построение нейросетевых алгоритмов обработки информации для распознавания изображений элементов текста // Журнал «Проблемы информатики и энергетики», Изд-во «Фан» АН РУз. - Ташкент, 2007. - №5-6. - с. 71-74.
13. Ахатов А.Р., Курбанов М.М. Контроль достоверности заголовков IP-пакетов в системах автоматизированного управления // НТЖ «Химическая технология. Контроль и управление». - Ташкент, 2008. - № 2. - с. 41-44.
14. Ахатов А.Р. Программные методы контроля достоверности в структуре пакетов передачи данных // Журнал «Естественные и технические науки», ISSN 1684-2626. – Москва, 2008. - №3 (35). - с. 310-318
15. Жуманов И.И., Ахатов А.Р. Методика оценки условных вероятностей ошибок в текстах на основе n-граммной языковой модели // Журнал «Инфокоммуникации: Сети-Технологии-Решения», Узбекское агентство связи и информатизации. - Ташкент, 2008. - №2 (6) - с. 34-39.

16. Жуманов И.И., Ахатов А.Р., Курбанов М.М., Каршиев З.А. Определение показателей системы контроля достоверности информации заголовков IP-пакетов // Журнал «Алока дунеси». – Ташкент, 2008. - № 5-6. - с. 40-47.
17. Ахатов А.Р. Повышение достоверности информации систем электронного документооборота на прикладных уровнях телекоммуникационных сетей // Журнал «Техника и технология», ISSN 1811-3532. – Москва, 2008. - №4. - с. 25-32.
18. Жуманов И.И., Ахатов А.Р. Программный контроль достоверности информации в IP-пакетах за счет избыточности различной природы // НТЖ «Химическая технология. Контроль и управление». - Ташкент, 2008. - № 4. - с. 52-57.
19. Жуманов И.И., Ахатов А.Р. Оценка качества работы системы орфографических ошибок на основе n-граммной языковой модели // НТЖ «Химическая технология. Контроль и управление». - Ташкент, 2008. - №5. - с. 21-26.
20. Ахатов А.Р. Разработка моделей и построение алгоритма оценки качества контроля элементов текста на основе n-грамм // Журнал «Вестник ТУИТ». - Ташкент, 2008. - №2. - с. 64-70.
21. Жуманов И.И., Ахатов А.Р., Тишликов С. Алгоритмы контроля достоверности зашифрованной текстовой информации по цифровым и числовым модулям // НТЖ «Химическая технология. Контроль и управление». - Ташкент, 2009. - №1. - с. 35-41.
22. Ахатов А.Р. Алгоритмы программной системы контроля текстовой информации на основе n-граммной языковой модели // Журнал «Актуальные проблемы современной науки», ISSN 1811-3532. - Москва. 2009. - №3. - с. 156-161.
23. Жуманов И.И., Ахатов А.Р. Оценка эффективности программного комплекса контроля достоверности текстовой информации систем электронного документооборота // НТЖ «Химическая технология. Контроль и управление» - ТГТУ, Ташкент, 2009 - № 2. - с. 46-52.
24. Ахатов А.Р. Программные методы контроля достоверности информации в структуре пакетов передачи данных систем электронного документооборота // Журнал «Вестник СибГУТИ» СибГУТИ, ISSN 1998-6920. - Новосибирск, 2008. - №2. - с.3-20.
25. Ахатов А.Р., Жуманов И.И., Тишликов С.А. Эффективность контроля качества передачи и обработки текстовой информации на основе n-граммной модели выявления ошибок // Журнал «Вестник ТУИТ». - Ташкент, 2009. - №1. - с. 71-77.
26. Ахатов А.Р. Модель оценки показателей эффективности методов контроля достоверности в пакетах передачи данных // Доклады Академии наук Республики Узбекистан, «Фан». – Ташкент, 2009. - №3-4. - с. 30-34.
27. Ахатов А.Р., Жуманов И.И. Оценка потенциальных возможностей методов контроля достоверности информации в системах электронного документооборота // Журнал «Проблемы информатики и энергетики», Изд. «Фан» АН РУз. – Ташкент, 2009. - №2 - с.55-62.
28. Жуманов И.И., Ахатов А.Р. Построение интеллектуальной системы контроля достоверности элементов текста на основе алгоритмов распознавания нейронными сетями // Журнал «Вестник ТУИТ». - Ташкент, 2009. - №2 - с. 67-73.
29. Жуманов И.И., Ахатов А.Р. Программные методы повышения достоверности информации на уровне сетевых служб многоуровневой модели OSI // Журнал «Проблемы информатики и энергетики», Изд. «Фан» АН РУз. - Ташкент, 2009. - №3- с. 65-73.
30. Ахатов А.Р., Жуманов И.И. Моделирование компонентов системы контроля передачи текстовых сообщений на основе статистической избыточности // Инфокоммуникации: Сети-Технологии-Решения, Узбекское агентство связи и информатизации. - Ташкент, 2009. - №4, с. 33-37.
31. Жуманов И.И., Ахатов А.Р. Алгоритмы контроля достоверности передачи текстовой информации на основе механизмов статистического кодирования // Журнал «Вестник ТУИТ», Ташкентский университет информационных технологий. – Ташкент, 2009. - №4. - с. 61-67.

32. Жуманов И.И., Ахатов А.Р. Оптимизация контроля передачи и обработки информации на базе технологии параллельных вычислений CUDA // НТЖ «Химическая технология. Контроль и управление». - Ташкент, 2009. - № 5. - с. 33-39.
33. Ахатов А.Р., Тишликов С.А. Парсинговое моделирование компонентов и оценка качества системы контроля и коррекции орфографии // Журнал «Актуальные проблемы современной науки», ISSN 1680-2721. – Москва, 2009. - №6 (50) - с. 180-185.
34. Ахатов А.Р., Тишликов С.А. Моделирование процессов контроля орфографии на основе парсингового представления n-граммной языковой модели // Журнал «Техника и технология», ISSN 1811-3532. - Москва, 2009. - №6. - с.43-48.
35. Ахатов А.Р. Методы оценки рационального объема избыточности для повышения достоверности пакетов передачи данных // Доклады Академии наук Республики Узбекистан, «Фан». - Ташкент, 2009. - № 6. - с.29-34.
36. Жуманов И.И., Ахатов А.Р. Алгоритмы контроля достоверности информации на основе нейросетевой системы распознавания элементов текста // НТЖ «Химическая технология. Контроль и управление» - Ташкент, 2009. - № 6. - с. 29-37.
37. Камиллов М.М., Ахатов А.Р. Система контроля достоверности текстовой информации на основе n-граммных парсинговых моделей // Журнал «Проблемы информатики», Сибирское отделение РАН. - Новосибирск, 2010. - № 1(5). - с. 42-52.
38. Жуманов И.И., Ахатов А.Р. Оценки достоверности передачи изображений элементов текста в телекоммуникационных сетях // НТЖ «Химическая технология. Контроль и управление», ТГТУ. - Ташкент, 2010. - № 2. - с. 23-30.
39. Жуманов И.И., Ахатов А.Р. Алгоритмы контроля достоверности информации на основе методов словарного кодирования // Журнал «Вестник ТУИТ». - Ташкент, 2010. - №1. - с.22-27.
40. Ахатов А.Р., Жуманов И.И. Система контроля и коррекции ошибок в текстах на основе графематического анализа и самообучения // Журнал «Вестник ТУИТ», Ташкент, 2010. - №1 - с. 65-72.
41. Ахатов А.Р., Жуманов И.И. Контроль и коррекция орфографических ошибок на основе n-граммной статистики // Журнал «Проблемы информатики и энергетики», Изд. «Фан» АН РУз. - Ташкент, 2010. - №2. - с. 74-81.
42. Жуманов И.И., Ахатов А.Р. Алгоритмы контроля достоверности изображений элементов текста в структуре пакетов передачи данных// НТЖ «Химические технологии. Контроль и управление». - Ташкент, 2010. - № 3. - с.39-46.
43. Жуманов И.И., Ахатов А.Р. Контроль достоверности пакетов передачи данных методами модульного суммирования // НТЖ «Химическая технология. Контроль и управление». – Ташкент, 2010. – № 6 (36),. - с. 27-33.
44. Ахатов А.Р. Оценки контроля достоверности информации в структуре пакетов передачи данных // Доклады Академии наук Республики Узбекистан, «Фан». - Ташкент. - № 4, 2010. - с.20-23.
45. Ахатов А.Р., Жуманов И.И. Определение условных вероятностей ошибок в текстах на основе n-граммной языковой модели // Журнал «Проблемы информатики и энергетики», Изд. «Фан» АН РУз. – Ташкент, 2010. - №3. - с. 68-74.
46. Ахатов А.Р., Жуманов И.И. Способы моделирования основных компонентов программной системы контроля и коррекции орфографических ошибок на основе Фреймворк Сфинкс-4// Журнал «Вестник ТУИТ». - Ташкент. - 2010, №4. - с. 14-20.
47. Ахатов А.Р. Контроль достоверности элементов текста в пакетах передачи данных на основе модульного суммирования // Доклады Академии наук Республики Узбекистан, «Фан». - Ташкент, 2011. - № 1, - с.23-25.
48. Жуманов И.И., Ахатов А.Р. Интеллектуальный контроль достоверности текстовой информации на основе учета свойств и распределений данных // НТЖ «Химическая технология. Контроль и управление» - Ташкент, 2011. - № 1 (37), - с.41-47.

49. Akhatov A.R. Methods and algorithms of the spelling control and correction on the basis of fuzzy qualifier with MIMO-structure// Open Access Research Journal “Applied Technologies and Innovations”, Prague Development Center. – Prague, 2011. – Volume 5, November 2011. – pp.27-37

50. Akhatov A.R. Methods for controlling the authenticity of textual information transfer on the basis of statistical and structural redundancy//International Journal of Automation and Computing, Chinese Academy of Sciences, China. - Volume 9, Number 5 (2012). - p.p.518-529.

51. Ахатов А.Р. Обнаружение и исправление ошибок оператора в АСУ //«СамДУ ахборотномаси», СамДУ. - Самарканд, 2002. - №1. – с. 26-30.

52. Ахатов А.Р., Тишликов С.А., Сабирова Ф. Оценка надежности систем контроля передачи и обработки текстовой информации для обнаружения N-граммных ошибок // «Илмий тадқиқотлар ахборотномаси» – Самарканд: СамДУ, 2009. - №3 (55). - 5-9 б.б.

53. Жуманов И.И., Ахатов А.Р. Интеллектуализация процессов контроля достоверности элементов текста на основе механизма распознавания изображений // «Илмий тадқиқотлар ахборотномаси».– Самарканд: СамДУ, 2010. - №1 (59). - 14-19 б.б.

54. Жуманов И.И., Ахатов А.Р. Построение программной системы обработки текстов на узбекском языке на основе сегментационного анализа // «Илмий тадқиқотлар ахборотномаси». – Самарканд: СамДУ, 2010. - №3 (61) – 22-27 б.

55. Зарипова Г.И., Ахатов А.Р. Контроль и коррекция ошибок передачи текстовых сообщений на основе нейронечеткой семантической гиперсети // «Илмий тадқиқотлар ахборотномаси». – Самарканд: СамДУ, 2010. - №5 (63) – 20-27 б.

56. Ахатов А.Р. Зарипова Г.И. Интерактивный контроль орфографии на основе семантической гиперсети обработки элементов текста //«Илмий тадқиқотлар ахборотномаси». – Самарканд: СамДУ, 2011. - №1 (65) – 36-43 б.

II бўлим (II часть; II part)

57. Akhatov A.R., Jumanov I.I. Improvement of text information processing quality in documents processing systems // 2nd IEEE/IFIP International Conference In Central Asia On Internet ICI-2006, <http://ieeexplore.ieee.org/>. - Tashkent, 2006. – p.p 1-5.

58. Jumanov I.I., Tursinxanov N.M., Akhatov A.R. Methods and algorithms of input information protection in electronic document processing systems// 2nd IEEE/IFIP International Conference ICI-2006, <http://ieeexplore.ieee.org/>. - Tashkent, 2006. – p.p 1-5.

59. Жуманов И.И., Ахатов А.Р. Повышение надежности программного обеспечения проблемных задач в системах электронного документооборота // Республиканская научная конференция «Современное состояние и пути развития информационных технологий», НТЦ «СИТ» - Ташкент, 2006. - с. 164-167.

60. Ахатов А.Р. Построение систем контроля ошибок в текстах на узбекском языке на основе многоуровневой модели. // Сборник трудов Международной научно-практической конференции «Инновация-2007», ТГТУ. - Ташкент, 2007. - стр. 343-344.

61. Jumanov I.I., A.R.Akhatov, Djumanov O.I. An Effective Quality Control of Textual Information on the Basis of Statistical Redundancy in Distributed Mobile IT Systems and e-Applications // 3-d International Conference in Central Asia on Internet, ISBN: 1-4244-1007-X. - Tashkent, 2007. - IEEE Catalog Number: 07EX1695C, Library of Congress: 2007920881.

62. Жуманов И.И., Ахатов А.Р. Определение показателей эффективности передачи заголовков IP-пакетов для контроля достоверности // Республиканская научная конференция “Современные проблемы математики, механики и информационных технологий”, Национальный университет Узбекистана. – Ташкент, 2008. – с. 89-91.

63. Jumanov I.I., Akhatov A.R., Karshiyev Z.A., Kurbanov M.M. Definition of Transfer Efficiency Parameters of IP-packages Headings for the Reliability Control // ICEIC 2008, the 9th International Conference on Electronics, Information and Communication, organized by TUIT and IEE of Korea. - Tashkent, 2008. – p.p. 115-119

64. Жуманов И.И., Ахатов А.Р. Создание мониторинга корпоративной информационно-коммуникационной сети вузов // X международная научно-методическая конференция ВУЗов и факультетов телекоммуникаций. - Ростов-на-Дону, 2008. - с. 78-81.
65. Жуманов И.И., Ахатов А.Р. Повышение достоверности информации в пакетах передачи данных на основе программных методов // Республиканская научная конференция «Современное состояние и пути развития информационных технологий». - Ташкент: НИИ «Алгоритм-инжиниринг» АН РУз, 2008. - с.71-75
66. Ахатов А.Р. Алгоритм распознавания элементов текста на основе n-граммной структурированной модели // Сборник научных статей международной научно-практической конференции «Инновация-2008», - Ташкент: ТГТУ, 2008. - с. 338-339.
67. Akhatov A.R. Technique for definition N-gram mistakes estimations in the automated systems of information processings // The Fifth World Conference on Intelligent Systems for Industrial Automation, Uzbekistan, ISBN 3-933609-27-5, b-Quadrat Verlag-86916. - Tashkent, 2008. – p.p.146-153.
68. Ахатов А.Р. Узатиладиган матнли маълумотлар таркибидаги хатоларни парсинг элементлари асосида таниш ва тахрир қилиш дастури // Инновацион ғоялар, технологиялар ва лойиҳалар II республика ярмаркаси. - Тошкент, 2009. - 239 б.
69. Akhatov A.R., Jumanov I.I. Methods and Algorithms of the Control and Correction of Errors in Texts on the Basis of Statistical Redundancy of the Transferred Information // The Third Congress of the World Mathematical Society of Turkic Countries, Volume 2. - Almaty, 2009. - p.199.
70. Жуманов И.И., Ахатов А.Р. Оценка эффективности контроля и коррекции орфографии в компьютерных системах обучения естественным языкам // Труды международной конференции «Актуальные проблемы прикладной математики и информационных технологий – Аль-Хорезми 2009». - Ташкент, 2009. - с. 252-255.
71. Ахатов А.Р., Бобомуродов О.Ж. Формирование стратегии обучения в интеллектуальной обучающей системе на основе обработки модели данных и знаний // Всероссийская конференция с международным участием «Знания-Онтологии-Теория (ЗОНТ-09)». - Новосибирск: Институт Математики СО РАН, 2009. - с. 129-134.
72. Ахатов А.Р. Система контроля и коррекции орфографии узбекского языка на основе структурированных парсинговых моделей // Сборник научных статей международной научной конференции «Инновация 2009». - Ташкент, 2009. - с. 308-310.
73. Jumanov I.I., Akhatov A.R. Fuzzy semantic hypernet for information authenticity controlling in electronic document circulation systems// “Application of Information and Communication Technologies – AICT 2010”. 12-14 October 2010. Section 2, IEEE. – Tashkent, 2010. – P.21-25.
74. Jumanov I.I., Akhatov A.R. The control of information transfer reliability in intellectual control systems on the basis of statistical redundancy // Sixth World Conference on Intellectual Systems for Industrial Automation, Uzbekistan, TSTU. - Tashkent, 2010. – p. 70-75.
75. Jumanov I.I., Akhatov A.R. Estimation of reliability for the system of mistakes dynamic control at transfer and processing of the text information// Abstracts of Plenary and Invited Lectures of International School and Conference on Foliations, Dynamical Systems, Singularity Theory and Perverse Sheaves, 6-21 October 2009, SamSU, Uzbekistan. – Samarkand, 2010. – 80-85 p.p.
76. Zaripova G.I., Akhatov A.R. Checking of Authenticity of Information Transfer by Using the Statistical Redundancy from Transmitted Text Elements Images // The Seventh World Conference on Intelligent Systems for Industrial Automation, 25-27 November, 2012. – Tashkent, Uzbekistan, 2012. – P.178-184
77. Жуманов И.И., Ахатов А.Р., Мингбоев Н.С., Гужов Е.Е., Жуйлов В.В. Комплекс программ коррекции орфографических ошибок в машинописных текстах на узбекском языке// Государственное патентное ведомство РУз. Свидетельство №DGU 00492, 24.12.2001 г.

78. Жуманов И.И., Ахатов А.Р., Мингбаев Н.С. Программа автоматического обнаружения и исправления ошибок в передаваемых текстах на узбекском (латиница, кириллица), английском и русском языках// Государственное патентное ведомство РУз. Свидетельство №DGU 00488, 24.12.2001 г.

79. Жуманов И.И., Ахатов А.Р., Турсынханов Н.М., Усаров С.А. Комплекс программ для автоматизации делопроизводства на предприятиях// Государственное патентное ведомство РУз. Свидетельство №DGU 00497, 24.12.2001 г.

80. Жуманов И.И., Ахатов А.Р., Мингбаев Н.С. Матнли ва ракамли маълумотларни узатиш ишончлигини чизикли йигиндилар усули билан назорат килувчи дастур // Государственное патентное ведомство РУз. Свидетельство №DGU 00589, 02.12.2002 г.

81. Жуманов И.И., Турсынханов Н.М., Ахатов А.Р. Программа формирования толкового словаря служебных слов на узбекском, русском и английском языках // Государственное патентное ведомство РУз. Свидетельство №DGU 00699, 31.10.2003 г.

82. Жуманов И.И., Турсынханов Н.М., Ахатов А.Р. Программа транслитерации текстовой информации по желанию пользователя// Государственное патентное ведомство РУз. Свидетельство №DGU 00698, 31.10. 2003 г.

83. Жуманов И.И., Ахатов А.Р., Турсынханов Н., Джураев М.К. Матнлардаги хатоларни Хаффман код асосида назорат қилиш дастурлар мажмуаси // Государственное патентное ведомство РУз. Свидетельство №DGU 00955, 31.05.2005 г.

84. Жуманов И.И., Ахатов А.Р., Курбанов М.М. Узатиладиган матнли маълумотлар таркибидаги хатоларни парсинг элементлари асосида таниш ва тахрир қилиш дастури // Государственное патентное ведомство РУз. Свидетельство № DGU 01618, 21.10. 2008 г.

85. Жуманов И.И., Ахатов А.Р., Курбанов М.М. Кўп босқичли морфологик таҳлил модели бўйича ўзбек тилидаги матнларни таниш, назорат ва имловий хатоларни тахрир қилиш программаси тизими // Государственное патентное ведомство РУз. Свидетельство № DGU 01603, 10.10. 2008 г.

86. Жуманов И.И., Ахатов А.Р., С.А.Тишликов. Ўзбек тилидаги иатнлар таркибидаги имло хатоларни сўзларнинг хеш-кодлари жадвалларини қўллаш асосида назорат қилиш дастури // Государственное патентное ведомство РУз. Свидетельство, № DGU 01840, 07.10.2009 г.

87. Жуманов И.И., Ахатов А.Р., С.А.Тишликов. Ўзбек тилидаги матнлар таркибидаги имло хатоларни морфемлар таҳлили ва синтези асосида назорат ва тахрир қилиш дастурий тизим интерфейси // Государственное патентное ведомство РУз. Свидетельство, № DGU 01842, 07.10.2009 г.