

**AZƏRBAYCAN MİLLİ ELMLƏR AKADEMİYASI
İNFORMASIYA TEXNOLOGİYALARI İNSTİTUTU**

Əlyazması hüququnda

RƏNA TOFİQ QIZI QASIMOVA

**MİLLİ DOMEN ADLARININ İNTELLEKTUAL
MONİTORİNQİ SİSTEMİNİN SİNTEZİ ÜÇÜN METOD VƏ
ALQORİTMLƏRİN İŞLƏNMƏSİ**

3338.01 – Sistemli analiz, idarəetmə və informasiyanın işlənməsi

Texnika üzrə fəlsəfə doktoru elmi dərəcəsi
almaq üçün təqdim edilmiş dissertasiyanın

AVTOREFERATI

Bakı – 2014

İş Azərbaycan Milli Elmlər Akademiyası İnformasiya Texnologiyaları
İnstitutunda yerinə yetirilmişdir.

Elmi rəhbər:

Akademik

R.M.ƏLİQULİYEV

Rəsmi opponentlər:

Texnika üzrə elmlər doktoru

R.M. ALIQLİYEV

Texnika üzrə fəlsəfə doktoru

L.H. KƏRİMOVA

Aparıcı təşkilat:

BDU-nun “İnformasiya texnologiyaları və proqramlaşdırma” kafedrası

Dissertasiyanın müdafiəsi “12” dekabr 2014-cü il saat 12⁰⁰-da AMEA
İnformasiya Texnologiyaları İnstitutunun nəzdindəki FD.01.231 Dissertasiya
Şurasının iclasında keçiriləcəkdir.

Ünvan: Az1141, Bakı şəhəri, B.Vahabzadə küçəsi, 9

Dissertasiya işi ilə Azərbaycan Milli Elmlər Akademiyası İnformasiya
Texnologiyaları İnstitutunun kitabxanasında tanış olmaq olar.

Avtoreferat “11” noyabr 2014-cü il tarixdə paylanmışdır.

FD.01.231 Dissertasiya şurasının
elmi katibi, texnika üzrə fəlsəfə doktoru

R.H.ŞIXƏLİYEV

İŞİN ÜMUMİ XARAKTERİSTİKASI

İşin aktuallığı. Bu gün Azərbaycanda informasiya cəmiyyətinin (İC) qurulması dövlət siyasətinin əsas tərkib hissəsinə çevrilmiş və ardıcıl, məqsədyönlü şəkildə həyata keçirilir. Bu siyasətin əsas istiqamətləri İC-nin qurulmasını nəzərdə tutan “Elektron Azərbaycan” dövlət proqramında müəyyən olunmuşdur. Eyni zamanda Azərbaycan Respublikasının Prezidenti İlham Əliyevin 2013-cü ilin Azərbaycan Respublikasında “İnformasiya-kommunikasiya texnologiyaları ili” elan edilməsi haqqında imzaladığı Sərəncamda qarşıya qoyduğu məsələlər (.az domenindən istifadənin genişləndirilməsi, qeydiyyat prosesinin təkmilləşdirilməsi, milli domen adlarının artırılması, Azərbaycan həqiqətlərini virtual məkanda əks etdirən informasiya resurslarının yaradılması, təbliği və yayılması üzrə tədbirlərin görülməsi) normativ hüquqi bazanın, idarəetmənin təkmilləşdirilməsinə, informasiya-kommunikasiya texnologiyaları (İKT) potensialının gücləndirilməsinə və ümumiyyətlə, İC-nin inkişafına yönəldilmişdir.

Son illər İnternetdə yüksək səviyyəli domenlərin kütləvi meydana gəlməsi və çox sürətlə artımı, eyni zamanda korporativ informasiya sistemlərinin, veb saytların, portalların yaradılması domen adları sahəsində bir çox problemlərin yaranmasına səbəb olmuşdur. Təhlillər göstərir ki, domen adlarının ilkin təyinatı İnternet mühitində ünvanlaşdırma atributu, əmtəə nişanlarının, şirkət və digər qurumların adlarının daşıyıcısı olmasına baxmayaraq, bu gün qeyri-sağlam, ədalətsiz rəqabət vasitəsinə çevrilmişdir. Bu məqsədlə aparılan araşdırmalar nəticəsində müəyyən olunmuşdur ki, Azərbaycana mənsub olan bəzi coğrafi adları, tarixi, mədəni və digər dəyərləri özündə əks etdirən milli domen adları müxtəlif ölkələrdə yaşayan xarici vətəndaşlar tərəfindən qeydiyyatdan keçirilmişdir. Bu da nəinki Azərbaycan, eyni zamanda dünya ölkələri üçün də xarakterik məsələ olaraq ciddi əhəmiyyət kəsb edir.

Yuxarıda qeyd olunan problemləri nəzərə alaraq domen adları sisteminin (Domain Name System, DNS) serverlərində yığılan domen adları haqqında informasiya resurslarının elmi təhlilinin aparılması zəruridir. Müxtəlif DNS-serverlərdə yığılan bu məlumatlardan səmərəli istifadə etmək üçün əlverişli axtarış sisteminin və verilənlər xəzinəsinin (VX) işlənməsinə ehtiyac vardır.

Hazırda DNS trafikinin analizi, şəbəkə aktivliyinin yoxlanılması, domen zonalarının təhlili, domenlərin statistikasını, fişinq tədqiqatı, veb-kontentlərin və xaker proqramlarının izlənməsi, tapılması üçün metod və

hesablama vasitələrinin yaradılmasına xüsusi diqqət edilir. Lakin, müasir informasiya texnologiyalarından istifadə etməklə milli domen adlarının qeydiyyat verilənlərini toplayan, onları sistemləşdirən, intellektual emalını təmin edən, domen adlarının qeydiyyat sahiblərinin məqsədlərini avtomatik təyin edən, qərar qəbul edilmədən operativliyi təmin edən, insan əməyini yüngülləşdirən intellektual monitoring sistemləri praktiki olaraq yoxdur. Hazırda yaradılan və tətbiq olunan DNS və veb-analizatorlar, eləcə də digər analitik sistemlər bu tələblərə cavab vermir.

Yuxarıdakıları nəzərə alaraq demək olar ki, domen adlarının düzgün istifadəsi və qorunması mexanizmlərinin yaradılması, DNS-də tamlığın, təhlükəsizliyin və stabilliyin qorunması, domen adı sahiblərinin qruplaşdırılması, identifikasiyası və məqsədlərinin öyrənilməsinə dair faydalı biliklərin aşkarlanması, domen adlarının qeydiyyat verilənlərinin intellektual analizinin aparılması aktual problemlərdəndir. Bu problemlərin həlli DNS serverlərdəki qeydiyyat məlumatlarının intellektual analizinə imkan verən metod və alqoritmlərin işlənməsini tələb edir. Göstərilən problemlərin mövcudluğu və qəbul edilmiş bir sıra normativ aktlar mövzusunun aktual olduğunu deməyə əsas verir və bu istiqamətdə tədqiqatların aparılmasını zəruri edir.

İşin məqsədi: virtual məkanda Azərbaycanın milli və mənəvi dəyərlərinin qorunması, domen bazarının tənzimlənməsi, domen adlarının müxtəlif təhdidlərdən qorunması üçün domen sahiblərinin məqsədlərinin avtomatik aşkarlanması məqsədi ilə qərarların qəbul edilməsini dəstəkləyən intellektual monitoring sisteminin sintezi metodlarının və alqoritmlərinin işlənməsidir.

Dissertasiya işində qarşıya qoyulmuş məqsədlərə çatmaq üçün aşağıdakı məsələlərə baxılmış və həll olunmuşdur:

- İnternet şəbəkəsində domen adlarının monitoringinin aparılması, mövcud problemlər və onların həlli yollarının müəyyənəşdirilməsi;
- milli domen adları intellektual monitoring sisteminin sintezinin aktual elmi-nəzəri problemlərinin araşdırılması;
- milli domen adlarının intellektual monitoringi sisteminin konseptual modelinin işlənməsi;
- domen adlarının qeydiyyat verilənlərinin ilkin emalı üçün alqoritmin işlənməsi;
- domen adlarının qeydiyyat verilənlərinin identifikasiyası üçün metodun işlənməsi;

- domen adları qeydiyyat verilənlərinin intellektual analizi üçün metodun işlənməsi;
- virtual məkanda Azərbaycan Respublikasına aid olan domen adlarının inkişaf dinamikasının qiymətləndirilməsi;
- əldə olunan nəticələrin yoxlanılması məqsədi ilə eksperimentlərin aparılması.

Tədqiqat metodları. Dissertasiya işində qarşıya qoyulmuş məsələni həll etmək üçün Data mining texnologiyasından, qraflar nəzəriyyəsi, qərar qəbuletmə nəzəriyyəsi, verilənlərin intellektual analizi, assosiativ qaydalar, klasterləşmə və təsnifat metodlarından istifadə edilmişdir.

Müdafiyə çıxarılan əsas müddəalar:

- milli domen adlarının intellektual monitorinqi sisteminin konseptual modeli;
- domen adlarının qeydiyyat verilənlərinin ilkin emalı algoritmi;
- milli domen adlarının qeydiyyat verilənlərinin identifikasiyası metodu;
- domen adlarının qeydiyyat verilənlərinin intellektual analizi metodu.

Elmi yeniliklər. Dissertasiya işində alınan elmi yeniliklər aşağıdakılardır:

- milli domen adlarının intellektual monitorinqi sisteminin konseptual modeli işlənmişdir;
- domen adlarının qeydiyyat verilənlərinin ilkin emalı üçün alqoritm işlənmişdir;
- domen adlarının qeydiyyat verilənlərinin identifikasiyası üçün metod işlənmişdir;
- domen adlarının qeydiyyat verilənlərinin klasterləşdirilməsi üçün qeyri-iyerarxik metoddan istifadə edilməsi təklif edilmişdir;
- domen adlarının qeydiyyat verilənlərinin intellektual analizi üçün metod işlənmişdir.

İşin praktiki əhəmiyyəti. Dissertasiya işində alınmış elmi-nəzəri nəticələrdən domen adlarının vəziyyətinin monitorinqində və nəzarətində, domen zonaları üzrə qiymətləndirilmənin aparılmasında, domen adlarının qeydiyyat verilənlərinin klasterləşdirilməsində, domen adları sahiblərinin gizli məqsədlərinin aşkarlanmasında və identifikasiyasında, İnternetdə domen bazarının cari vəziyyətinin analizi və qiymətləndirilməsində, informasiya təhlükəsizliyi məsələlərində tətbiq etmək olar.

İşin nəticələrinin tətbiqi. Dissertasiya işində əldə olunmuş nəzəri və praktiki nəticələr AzScienceNet korporativ şəbəkəsində tətbiq edilmişdir və müvafiq akt alınmışdır.

İşin aprobeiasyası. Dissertasiya işinin əsas elmi-nəzəri və praktiki nəticələri aşağıdakı siyahıda təqdim olunan bir sıra beynəlxalq konfranslarda, elmi seminarlarda məruzə edilmiş və müzakirə olunmuşdur: третья Международная научно-практическая конференция «Теория, методы проектирования, программно-техническая платформа корпоративных информационных систем» (г. Новочеркасск, Россия, 20 май 2005 г.); четвертая Международная научно-практическая конференция «Теория, методы проектирования, программно-техническая платформа корпоративных информационных систем» (г. Новочеркасск, Россия, 26 май 2006 г.); first International Conference “Problems of Cybernetics and Informatics” PCI’-2006 (Bakı şəh., 24-26 oktyabr, Azərbaycan, 2006-cı il); седьмая Международная научно-практическая конференция «Интернет – Образование – Наука – 2010» (г. Винница, Украина, 28 сентября – 3 октября 2010 г.); the third International Conference “Problems of Cybernetics and Informatics” PCI’2010 (Baku, Azerbaijan, 6-8 September, 2010); the fourth International Conference “Problems of Cybernetics and Informatics” PCI’2012 (Baku, Azerbaijan, 12-14 September, 2012); elektron elm problemləri üzrə I Respublika elmi-praktiki konfransı (Bakı ş., Azərbaycan, 15-16 noyabr, 2012-ci il); riyaziyyatın tətbiqi məsələləri və yeni informasiya texnologiyaları üzrə II Respublika elmi konfransı (Sumqayıt ş., Azərbaycan, 27-28 noyabr, 2012-ci il); ümummilli lider Heydər Əliyevin anadan olmasının 90 illiyinə həsr olunmuş “İnformasiya təhlükəsizliyi problemləri” I Respublika elmi-praktiki konfransı, Bakı ş., Azərbaycan, 17-18 May, 2013-cü il).

Elmi nəşrlər. Dissertasiyanın nəticələri üzrə 19 elmi iş çap edilmişdir. Onlardan 9 məqalə nüfuzlu beynəlxalq səviyyəli jurnallarda, 9 məruzə isə beynəlxalq və Respublika konfranslarının materiallarında dərc olunmuş və 1 kitab nəşr edilmişdir.

Dissertasiyanın strukturu və həcmi. Dissertasiya giriş, dörd fəsil, əsas nəticələr, 209 adda ədəbiyyat siyahısı və əlavələrdən ibarətdir. İşin əsas hissəsi 119 səhifədə şərh olunmuş, 22 şəkil, 3 cədvəldən ibarətdir.

İŞİN MƏZMUNU

Girişdə dissertasiya işinin aktuallığı əsaslandırılmış, tədqiqat işinin məqsədi formalaşdırılmış, məqsədə çatmaq üçün həll olunacaq məsələlər müəyyən edilmiş, müdafiəyə çıxarılan əsas müddəalar açıqlanmış, alınmış nəticələrin elmi yeniliyi və praktiki əhəmiyyəti göstərilmişdir.

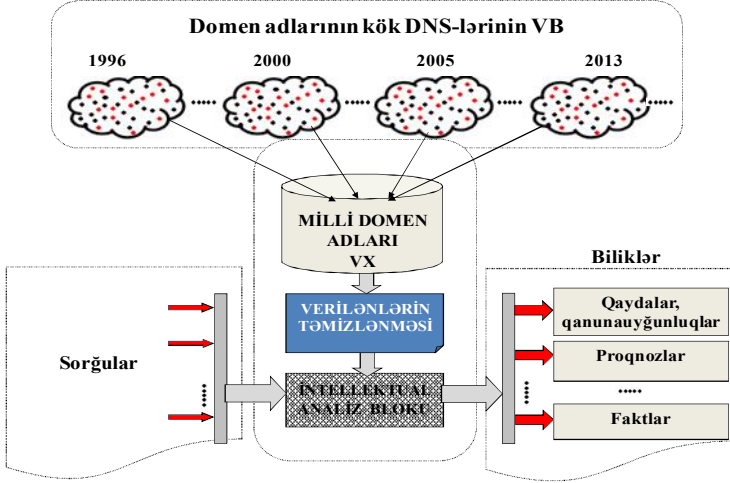
Birinci fəsil İnternet şəbəkəsində milli domen adlarının monitorinq problemlərinin araşdırılmasına həsr olunmuşdur. Bu məqsədlə ilk növbədə domen adları sisteminin əsas xüsusiyyətləri, müasir vəziyyəti, domen adlarının istifadəsi, idarəçiliyi və qeydiyyatı prosesindəki problemlər, domen adları ilə bağlı mübahisə halları araşdırılmışdır. Daha sonra domen zonalarının monitorinqi və müqayisəli təhlili aparılmış, domen adlarının qeydiyyatı qaydaları, milli domenlərin xüsusiyyətləri geniş tədqiq olunmuş, domen adının qeydiyyatı və idarə edilməsi ilə əlaqədar yaranan mübahisələrin həlli yolları göstərilmişdir.

Fəslin sonunda domen adlarının müxtəlif məqsədlərlə (siyasi, biznes, informasiya müharibəsi, qazanc əldə etmək və s.) qeydiyyatdan keçirilməsini nəzərə alaraq, milli domen adları ilə bağlı intellektual monitorinq sisteminin yaradılması zamanı meydana çıxan aktual elmi-nəzəri problemlər analiz olunmuş və həlli yolları göstərilmişdir. Dissertasiya işininin aktuallığını əsaslandıran əsas elmi-nəzəri problemlər müəyyənləşdirilmiş, bu istiqamətdə metod və alqoritmlərin işlənməsi tövsiyə olunmuşdur.

İkinci fəsildə milli domen adlarının qeydiyyat verilənlərinin identifikasiyası üçün model və alqoritmlər işlənmişdir. İlk növbədə domen adlarının qeydiyyat verilənlərinə tətbiqi zamanı sistemin işinin səmərəliliyini artıran, insan əməyini yüngülləşdirən intellektual analiz blokundan təşkil olunmuş milli domen adlarının konseptual modeli təklif edilmişdir (şəkl. 1.). Təklif olunan modeldə bu məsələlər klasterləşdirmə, proqnozlaşdırma metodlarına, qərarlar ağacı, assosiativ qaydaların axtarışı və s. alqoritmlərinə gətirilməklə həll edilir.

Bu fəsildə baxılan ən mühüm məsələlərdən biri də VX-nin yaradılması zamanı qeydiyyat verilənlərinin ilkin emalı (təmizlənməsi) və identifikasiyasıdır. Xəzinənin qurulması zamanı VX-də domen adlarının qeydiyyat verilənlərinin əsas problemi qeydiyyat verilənlərində səhvlərin, boşluqların, uyğunsuzluqların, dublikatların və s. olması, həmçinin verilənlərin təmizlənməsi və keyfiyyətinin yaxşılaşdırılması üçün redaktə məsafəsinin seçilməsi və həll alqoritmidir. Sadalanan problemləri nəzərə alaraq domen adlarının qeydiyyat verilənlərinin identifikasiyası üçün qərarların qəbul edilməsi metodu işlənmişdir.

Təklif olunan metod əsasən iki hissədən ibarətdir. Birinci hissədə qeydiyyat verilənlərinə qeyri-səlis müqayisə alqoritm tətbiq olunaraq mexaniki səhvlərə malik və ya təhrif olunmuş verilənlər təmizlənir, ikinci hissədə isə qərarların qəbulu alqoritm müqayisə olunan verilənləri identifikasiya edir.



Şək. 1. Milli domen adlarının intellektual monitorinqi sisteminin konseptual modeli

Qeyd edək ki, müxtəlif DNS-in VB-lərinin analizi göstərdi ki, domen adlarının qeydiyyat verilənlərinin eyni sahələrində yazıların məzmunu formasına görə müxtəlif (qeyri-identik) ifadə oluna bilər. Daxil edilən “registrar”, “person”, “address”, “organization”, “admin-o” və s. kimi qeydiyyat verilənlərinin keyfiyyətinin yüksəldilməsi məqsədilə səhvlərin aradan qaldırılması, yazılarda təkrarların olmaması tələb olunur. Məsələn, ölkələrin (şəhərlərin) adında səhv daxiletmələr var – Kanata (Kanada), Russian (Rusia), Frankfrut Am Main (Frankfurt Am Main) və s.; qeydiyyat verilənlərindən olan təşkilatın identifikatoru (registrar) “MONIKER ONLINE SERVICES, INC” – “MONIKER, INC”, “MONIKER”, “MONIKERS ONLINE SERVICES” qiymətlərinə malik ola bilər. Bu səhvlərin müxtəlif olmalarına baxmayaraq, aydındır ki, bütün bu adlar bir domen adı sahibini (person) ifadə edir.

Bunları nəzərə alaraq bu fəsilə Damerau-Levenşteyn məsafəsindən – iki sətirin simvollarının fərqi ölçüsündən istifadə edilir, bu isə bir sətirin digərinə çevrilməsi üçün lazım olan daxiletmə, silmə, əvəzetmə və transpozisiya (iki qonşu simvolların yerdəyişməsi) əməliyyatlarının minimal sayı kimi müəyyən edilir ($d_{DL}(GCAT,CGAT) = 1$), yəni

$$D_{ij} = \min(X_d + 1, X_s + 1, X + C_{\text{əvəzetmə}}, X_y + C_{\text{transpozisiya}}), \quad (1)$$

$$C_{\text{əvəzetmə}} = \begin{cases} 1, & \text{əgər } S_1[i] \neq S_2[j] \\ \infty & \text{əks halda, } 0 \end{cases} \quad (2)$$

$$C_{\text{transpozisiya}} = \begin{cases} 1, & \text{əgər } S_1[i] \neq S_2[j-1] \text{ və } S_1[i-1] = S_2[j] \\ \infty & \text{əks halda, } 0 \end{cases} \quad (3)$$

Burada, D_{ij} - S_1 sətirindəki birinci i simvolu ilə və S_2 sətirindəki birinci j simvolunun prefiksləri arasındakı məsafədir. Bu məsafəni domen adlarının qeydiyyat verilənlərinə tətbiq edərkən onun uyğunlaşdırılmış variantı təklif edilmişdir.

Qeydiyyat verilənlərinin identifikasiyası zamanı açar sahələri kimi fiziki şəxsin adı (person), fiziki şəxsin əlaqə ünvanı (address) domen administratorunun identifikatoru (admin-o), administrativ əlaqə üçün təşkilatın identifikatoru (admin-c), təşkilatın adı (organization) götürülür. Bu sahələrin komponentləri (tərkibi) ixtiyarı qaydada yerləşə bilirlər, bu da belə informasiyanın avtomatik emalı üçün mürəkkəb məsələ olur. Belə informasiyaya malik iki sətiri müqayisə etmək üçün əvvəlcə hər bir sahəni tərkib hissələrinə ayırmaq, sonra isə yalnız eyni mənə yüklü sətirləri, məsələn, fiziki şəxsin ünvanında şəhərlərin adını şəhərin adı ilə, küçələrin adını küçənin adı ilə, müqayisə etmək lazımdır. Sahələrin komponentlərinin (tərkiblərinin) müqayisəsi üçün şablonlardan istifadə edilir. Məsələn, ünvanın şablonu ünvanı təşkil edən informasiyaların kombinasiyasından ibarətdir:

[mənzil], [bina], [ev], [küçə], [şəhər], [rayon], [indeks], [ölkə]

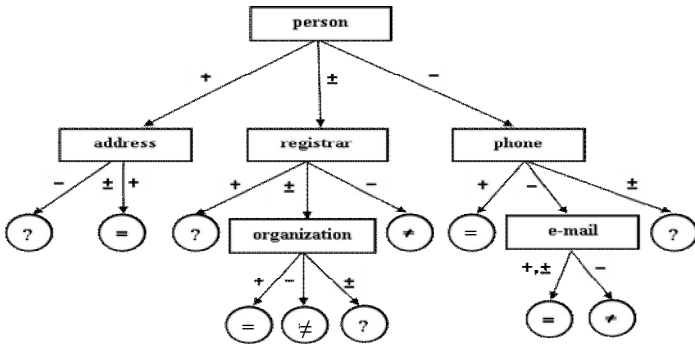
Məsələn, azerbaijan.info domeni üçün ünvanın şablonu belə qiymət alır:

, , 4676, Admiralty Way, Marina del Rey, California, 90292-660, US.

Sahələrin komponentlərinin müqayisəsi üçün şablonlardan istifadə etməklə qeyri-səlis müqayisə alqoritmi domen adlarındakı sahələrin müqayisəsinin dəqiqliyini və keyfiyyətini artırmağa imkan verir. Qeyd edək ki, qeyri-səlis müqayisə alqoritmi VX-nin qurulması zamanı domen adlarındakı rekvizitləri müqayisə edərək səhv və ziddiyyətli verilənlərin təmizlənməsi, eyni zamanda böyük həcmə malik qeydiyyat verilənlərinin oxşarlığa görə müqayisəsi problemini həll edir.

Təklif olunmuş qeydiyyat verilənlərinin identifikasiyası metodu isə təhrif olunmuş verilənlərin yazıların identifikasiyasına təsirini aradan qaldırmağa imkan verir ki, bu da doğruluğu əhəmiyyətli dərəcədə artırır və VX-nin doldurulması vaxtını azaldır. Domen adlarının qeydiyyat verilənlərindəki yazıların identifikasiyası məsələsinin həllində alqoritmik yanaşma kimi qərarlar ağacının tətbiqi təklif olunmuşdur. Məsələnin bu şəkildə qoyuluşu yenidir. İdentifikasiya məsələsinin həlli üçün təklif olunan yazıların identifikasiyası metodu qiymətləndirmə üçün istifadə olunan meyarların yaradılmasında çətinliklər yaratmır. Eksperimentlər zamanı alınan nəticələr təklif edilmiş metodun effektivliyini göstərir. Obyektlərin dəqiq identifikasiyası üçün obyektləri birqiymətli identifikasiya edən sahələrdən istifadə edilir.

Domen adlarının qeydiyyat verilənlərinin identifikasiyası üçün ağacın nümunəsi şək. 2-də göstərilmişdir. Domen adlarının qeydiyyat verilənlərinin identifikasiyası üçün təklif edilmiş qərarların qəbulu alqoritminin proqram reallaşdırılması həm DR Racket və həm də nəticələrin müqayisəsi üçün Python dilində yerinə yetirilmişdir. Keçirilən eksperimentlər eyni müsbət nəticələr verilmişdir.



Şək. 2. Domen adlarının qeydiyyat verilənlərinin identifikasiyası üçün qərarların qəbulu ağacı

Üçüncü fəsildə domen adlarının qeydiyyat verilənlərinin intellektual analizi üçün metod təklif edilir. İlk növbədə domen adlarının qeydiyyat verilənlərindən gizli biliklərin aşkarlanması məqsədi ilə mövcud olan intellektual analiz metodları və alqoritmləri müqayisəli analiz edilmiş, bu

metodların çox ümumi xarakter daşdığı və konkretləşdirilməsinə ehtiyac olduğu aşkarlanmışdır.

Domen adlarının qeydiyyat verilənlərinin intellektual analizi məsələlərinin həll edilməsi üçün problemlər müəyyənləşdirilmişdir. Bunlar serverlərdə böyük informasiya massivlərinin toplanması; domen adlarının qeydiyyat verilənlərinin ədədi və kateqoriyalı atributlardan ibarət olması; qeydiyyat verilənlərinin oxşar xarakteristikalara görə müəyyən edilməsi və ya qruplaşdırılması; metrikanın seçilməsi; domen adlarının qeydiyyat verilənlərindən faydalı, lazım olan biliklərin əldə olunmasıdır. Bu metodda ilk olaraq qeydiyyat verilənlərinə qeyri-iyerarxialı klasterləşdirmə alqoritmi tətbiq edilərək oxşar obyektlər qruplaşdırılır. Daha sonra isə assosiativ qaydaların axtarışı alqoritmindən istifadə edərək hər bir klaster üçün qaydalar generasiya olunaraq domen adı sahiblərinin məqsədlərini aşarlamaq mümkün olur.

Domen adlarının qeydiyyat verilənlərinin klasterləşdirilməsi məsələsi CLOPE alqoritmi vasitəsi ilə həll edilmişdir. Alqoritm avtomatik olaraq klasterlərin sayını seçir, həm də bu yeganə vahid parametrlə – itələmə əmsalı ilə tənzimlənir. Alqoritmin əsasında klaster histoqramı parametrinin artırılması ilə klasterlərdə tranzaksiyaların yaxınlığını artıran $Profit(C)$ – in (dəyər funksiyasının) maksimallaşdırılması ideyası dayanır.

Tutaq ki, verilən $D = \{t_1, t_2, \dots, t_n\}$ tranzaksiyalar bazasında, hər bir tranzaksiya $\{i_1, \dots, i_m\}$ obyektlər yığıdır. $\{C_1, \dots, C_k\}$ klasterlər çoxluğu $\{t_1, \dots, t_n\}$ çoxluqlarının bölünməsidir ki, burada $C_1 \cup \dots \cup C_k = \{t_1, \dots, t_n\}$ və $C_i \neq \emptyset \wedge C_i \cap C_j = \emptyset, 1 \leq i, j \leq k$. Dəyər funksiyasının hesablanmasını aşağıdakı kimi formalizə etmək olar: verilən $D = \{t_1, t_2, \dots, t_n\}$ tranzaksiyalar bazası və r itələmə əmsalı üçün elə $C = \{C_1, C_2, \dots, C_k\}$ bölünməsinə tapmaq lazımdır ki,

$$Profit(C, r) \rightarrow \max, \quad (4)$$

burada, r klaster daxilində oxşarlıq səviyyəsini tənzimləyir – r nə qədər böyükdürsə, klasterlərin yekun sayı o qədər çoxdur. Dəyər funksiyası aşağıdakı düsturla hesablanır:

$$Profit(C, r) = \frac{\sum_{i=1}^k \frac{S(C_i)}{W(C_i)^r} \times |C_i|}{\sum_{i=1}^k |C_i|}, \quad (5)$$

burada, $S(C_i)$ – i obyektlərinin C_i klasterinə daxil olmalarının yekun sayıdır, $W(C_i)$ – C_i unikal obyektlərin sayıdır, $|C_i|$ – i klasterində obyektlərin sayıdır, k – klasterlərin sayıdır, r – 1-dən böyük müsbət həqiqi ədəddir. Eksperimentin nəticələri göstərdi ki, optimal bölgü $r = 1,6$ qiymətində alınır. CLOPE alqoritmindən istifadə etməklə aparılan eksperimentlərin nəticəsi nümunədə verilmişdir (şək. 3.).

Cluster	Private Per	Domain	Country	Year	IP	Registrar	Contact	Status	Location
Cluster1	Private Per	azerbajian.ru	RUCENTER	2000	2011	Russian	+7.095930 - di@bugtrac -	active	San Francis
	Private Per	baku.ru	RUCENTER	2000	2011	Russian	+7.095930 - di@bugtrac -	active	San Francis
	Private Per	baku.su	RUCENTER	2005	2010	Russian	+7.095123 +7.095123 taragir@ge -	active	Houston, U
	Marina V Zy	baku.ru	RUCENTER	2005	2011	Russian	+7.903793 - useshahny -	active	Moscow, R
	Eugene V P	baku.su	RUCENTER	2010	2011	Russian	+7.963344 - etot_domer -	inactive	Moscow, R
	Private Per	karabakh.su	RUCENTER	2005	2010	Russian	+7.095234 +7.095234 taragir@ge -	active	Moscow, R
	Private Per	shuhsa.ru	RUCENTER	2005	2011	Russian	+7.327262 +7.327262 hostmaster -	active	Kazakstan
	Private Per	lacin.ru	RUCENTER	2009	2011	Russian	+7.495507 - MUSHFIG@ -	inactive	-
Cluster2	Minskamni	azerbajian.com	ENOM, INC	1997	2017	Malaysia	347482144 1.656873 nicolov@en -	active	London, Un
	Suren Grigc	karabakh.net	ENOM, INC	2001	2012	Armenia	+374.7143 +374.7171 ryrikian@bk -	active	Yerevan, A
Cluster3	M.A. Stenz	azerbajian.net	GODADDY	1999	2017	US	- - - -	inactive	San Antonic
	Role	azerbajian.org	Moniker Prr	2000	2011	US	+1.954984 +1.954969 support@im sale	passive	United Stat
	-	azerbajian.tv	ENOM, INC	2003	2011	-	- - - -	active	Germany
	-	baku.com	NETWORK	1996	2012	US	- - - -	inactive	Herndon, U
	M.A. Stenz	baku.org	GODADDY	2000	2011	US	+1.808878 - contact@st -	inactive	-
	Moniker Prr	baku.info	Moniker Prr	2001	2012	US	+1.954984 +1.954969 BAKU.INFO -	active	Houston, U
	Moniker Prr	baku.biz	Moniker Prr	2005	2011	US	+1.954984 +1.954969 BAKU.BIZ@ -	active	Houston, U
	-	baku.tv	ENOM, INC	2004	2011	-	- - - -	passive	Houston, U
	-	baku.ws	Wild West C	2004	2011	US	+1.480624 - dns@jmax -	active	Erfurt, Gerr
	-	baku.net	NETWORK	1998	2011	US	+1.310929 - question@ sale	passive	Frankfurt A
	Registratr	baki.org	Domains by	2005	2011	US	+1.480624 +1.480624 BAKU.ORG sale	passive	Frankfurt A
	-	baki.ws	GODADDY	2009	2011	US	+1.480505 - dns@jmax -	passive	San Diego,
	Vaheek Hac	karabakh.com	GODADDY	1999	2011	US	- - - -	passive	Scottsdale,
	-	karabakh.tv	GODADDY	2007	2011	Malta	+356.888 +356.888 loiborloo@ -	passive	Scottsdale,
	-	nakhchivan.net	DYNADOT	2007	2011	US	+1.240620 - sleepyauct sale	passive	Frankfurt A
	Kelsey Won	sharur.com	GODADDY	2010	2012	US	- - - -	passive	Scottsdale,
	Domains by	shuhsa.net	GODADDY	2009	2011	US	- - - -	passive	Scottsdale,
	Moniker Prr	ganca.com	MONIKER C	2005	2011	US	+1.954984 +1.954969 GANCA.CO -	passive	Secaucus, t
	-	azer.ws	Wild West C	2003	2011	US	+1.480624 +1.480624 dns@jmax -	passive	Pittsburgh,
	Anthony	mingejevic.com	ONLINEINC	2005	2010	US	+1.707362 - anthony@ -	inactive	-
	James Kenn	lacin.org	GODADDY	2009	2011	US	+1.310230 - jkennedy@ -	passive	Scottsdale,
	Registratr	lacin.info	GODADDY	2010	2011	US	+1.480624 +1.480624 LACHIN.INFO -	passive	Scottsdale,
	Kelsey Won	sharur.com	GODADDY	2010	2011	US	- - - -	passive	Scottsdale,
	aghdam.com	com	TERRANET	2009	2011	US	+1.800683 +1.240292 admin@don sale	passive	United Stat

Şək. 3. Domen adlarının tranzaksiya verilənlərinin klasterlərə bölünməsi

Qeydiyyat verilənlərinin klasterləşməsi məsələsinin həlli Azərbaycanca aid domen adlarının vəziyyəti, domen adı sahibləri arasındakı münasibətləri, İnternətdə domen bazalarının cari vəziyyəti, domen adı zəbt edənlərin aşkarlanması və s. haqqında informasiya verə bilər. Qeyd edək ki, klasterləşmə verilənlərdə mövcud qanunauyğunluqları təsvir etmək üçün

rahat olsa da bizim halda qeydiyyat verilənlərində mümkün ola bilən gizli biliklər haqqında lazımi informasiyanı vermir. Baxılan tədqiqat işində domen adlarının qeydiyyat verilənlərindən avtomatik olaraq yeni qanunauyğunluqların aşkarlanması üçün Data mining metodlarından olan assosiativ qaydaların (Association Rules) axtarışı alqoritminin tətbiqi təklif olunmuşdur.

Tutaq ki, $I = \{i_1, i_2, \dots, i_n\}$ – elementlər çoxluğu (element adlanan binar atributlar). İstənilən I altçoxluğunu elementlər toplusu və ya sadəcə toplu adlandırırlar (itemset). Topluda elementlərin sayı onun uzunluğu adlanır. Tutaq ki, D – tranzaksiyalar çoxluğu, burada hər bir tranzaksiya $T, T \subseteq I$ – binar vektoru əks etdirir, burada ki, $T.i_k = 1$, əgər element $i_k, k = \overline{1, n}$ tranzaksiyada mövcuddursa, yoxsa $T.i_k = 0$ ($T.i_k - k$ -ci atributun qiymətidir). Tutaq ki, X I -dəki elementlər toplusudur. Deyəcəyik ki, $T, T \subseteq I$ tranzaksiyası X -dən ibarətdir, yəni $X \subseteq T$, əgər $\forall i_k \in X$ üçün, $T.i_k = 1$.

Assosiativ qayda $X \rightarrow Y, [c, s]$ şəkilli implikasiya adlanır, burada $X \subset I, Y \subset I, X \cap Y = \emptyset$, c (*confidence*) – qaydanın doğruluğu, s (*support, significance*) – qaydanın dəstəyidir. Qaydanın dəstəyi – D tranzaksiyasından s faizdir, $X \cup Y$ -dən ibarətdir, yəni $s(X \rightarrow Y) = s(X \cup Y)$; qaydanın doğruluğu göstərir ki, X -dən ibarət D tranzaksiyasından c faiz, həmçinin Y -dən ibarətdir, yəni $c(X \rightarrow Y) = s(X \cup Y) / s(X)$. Əgər $c(X \rightarrow Y) \geq \text{minimum_}c$ və $s(X \rightarrow Y) \geq \text{minimum_}s$ olarsa, onda $X \rightarrow Y$ qaydası tezlik adlanır. Burada, minimal doğruluq və qaydanın minimal dəstəyi adlandırılan $\text{minimum_}c$ və $\text{minimum_}s$ qabaqcadan verilmiş sərhəd qiymətləridir. Assosiativ qaydaların axtarışı alqoritmləri doğruluq və dəstəklənmə əmsalları əvvəlcədən verilmiş sərhəd qiymətlərindən böyük olan qaydaların əldə olunması üçün təyin olunmuşdur.

Assosiativ qaydaların əldə olunma alqoritminin ideyası belədir. Alqoritmin birinci addımında bir elementli tez-tez rast gəlinən toplular tapılır. Bunun üçün bütün verilənlər toplusu içərisində onların dəstəklənmə əmsalını, yəni bazada rastgəlmə sayı hesablanır. Növbəti addım iki hissədən ibarət olacaqdır: potensial olaraq tez-tez rast gəlinən elementlər toplusunun (onları alternativ həllər adlandırırlar) generasiyası və alternativ həllər üçün dəstəklənmə əmsalının hesablanması. Alternativ həllərin

generasiyası funksiyasında k -elementli topluları əldə etmək üçün əvvəlki addımda müəyyən olunan və tez-tez rast gəlinən $(k-1)$ -elementli toplulardan istifadə edilir. Alternativ həllərin generasiyası alqoritmi iki addımdan ibarətdir:

1. *Birləşmə*. Hər bir C_k alternativ həlli tez-tez rast gəlinən $(k-1)$ ölçülü topluya digər $(k-1)$ elementli topludan bir elementin əlavə edilməsi yolu ilə formalaşdırılır.

2. *İzafi qaydaların ləğvi*. Əgər heç olmasa onun $(k-1)$ altçoxluqlarından biri tez-tez rast gəlinən deyildirsə, anti-monotonluq xassəsi əsasında C_k -dakı bütün toplular ləğv edilir.

Alqoritm Magnum Opus v.5.4.1. proqram paketi əsasında reallaşdırılmışdır. Magnum Opus istifadədə sadədir və proqramın işinin nəticəsi assosiasiativ qaydalardır. Aşkar edilmiş qanunauyğunluqlar analitiklərə, menecerlərə və rəhbər işçilərə müxtəlif verilənlərə istədikləri formada baxmağa və qərar qəbul etməyə imkan verir.

Dördüncü fəsilə təklif olunmuş model və metodlar eksperimental yoxlanılmışdır və AzScienceNet korporativ şəbəkəsində sınaqdan keçirilmişdir. Eksperimentdə VX-də 6 ay ərzində toplanmış qeydiyyat verilənlərindən istifadə olunmuşdur. Aparılan eksperimentlər Microsoft SQL Server və Magnum Opus v.5.4.1. proqram paketi əsasında reallaşdırılmışdır. Virtual məkanda Azərbaycan Respublikasına aid olan domen adlarının inkişaf dinamikasının qiymətləndirilməsi üçün metodika işlənmişdir. Eksperimentlərin nəticələri göstərir ki, təklif olunan model və metodlar aşağıdakı məsələlərin həllində istifadə oluna bilər:

- Domen adlarının qeydiyyat verilənlərinin intellektual emalında;
- Domen adı sahiblərinin identifikasiyasında və məqsədlərinin aşkarlanmasında;
- Domen adların və domen sahiblərinin nəzarətində;
- Domen sahəsində proqnozların, operativ və səmərəli, analitik qərarların verilməsində;
- Domen sahəsində real vəziyyətin qiymətləndirilməsində;
- Milli domen adların inkişaf dinamikasının izlənməsində;
- İKT və onun tətbiqi ilə məşğul olan qurumlarda qərarların qəbul edilməsi üçün lazım olan biliklərin üzə çıxarılmasında;
- Regionda qlobal şəbəkənin təkmilləşdirilməsində və s.

Alınan nəticələr qərarların qəbul edilməsinə dəstək sistemlərinin xüsusi proqram kompleksinin işləyib hazırlanmasının məqsədəuyğun olduğunu göstərmişdir.

DİSSERTASIYA İŞİNİN ƏSAS NƏTİCƏLƏRİ

Dissertasiya işi üzrə aparılmış tədqiqatlar zamanı qoyulmuş məsələlər həll olunmuş və aşağıdakı elmi-nəzəri və təcrübi nəticələr əldə olunmuşdur:

1. İnternet şəbəkəsində domen adlarının müasir vəziyyəti araşdırılmış, domen zonalarının müqayisəli təhlili aparılmış və bu sahədə mövcud olan problemlər müəyyənləşdirilmişdir;
2. Domen adlarının qeydiyyat verilənlərinə tətbiqi zamanı sistemin işinin səmərəliyini artıran, insan əməyini yüngülləşdirən intellektual analiz blokundan təşkil olunmuş milli domen adlarının intellektual monitorinqi sisteminin konseptual modeli işlənmişdir;
3. Verilənlər xəzinənin qurulması zamanı domen adlarının qeydiyyat verilənlərinin ilkin emalı üçün alqoritm işlənmişdir;
4. Qeydiyyat verilənlərinin təmizlənməsi məsələsində domen adlarının qeydiyyat verilənlərinin identifikasiyası üçün metod işlənmişdir;
5. Domen adlarının qeydiyyat verilənlərindən faydalı, lazım olan biliklərin əldə olunması məqsədi ilə qeydiyyat verilənlərinin intellektual analizi üçün metod işlənmişdir.

Dissertasiya işinin əsas nəticələri aşağıdakı elmi işlərdə dərc edilmişdir:

1. Алгулиев Р. М., Касумова Р. Т., Алекперова И. Я. Система многомерного анализа корпоративных данных / Третья Международная научно-практическая конференция «**Теория, методы проектирования, программно-техническая платформа корпоративных информационных систем**», г. Новочеркасск, Россия, 20 май 2005 г., с. 45-47.
2. Əliquliyev R.M., Qasımova R.T., Ələkbərova İ.Y. Qərarların qəbul edilməsini dəstəkləyən müasir konsepsiyalar haqqında // **AMEA Xəbərləri. Fizika-riyaziyyat və texnika elmləri seriyası**, 2005, №2 cild XXV, s. 70-74.
3. Алгулиев Р.М., Касумова Р.Т., Алекперова И.Я. Выполнение сложных запросов в многомерных аналитических системах Международная научно-практическая конференция / «**Теория, методы проектирования, программно-техническая платформа корпоративных информационных систем**», г. Новочеркасск, Россия, 26 май 2006 г., с. 58–59.
4. Əliquliyev R.M., Qasımova R.T., Ələkbərova İ.Y. Çoxölçülü analitik sistemlərdə mürəkkəb sorğuların realizasiyası məsələlərinin analizi

- / Proceedings of the First International Conference “Problems of Cybernetics and Informatics”**, Bakı, Azərbaycan, 24-26 oktyabr, 2006, s. 71-72.
5. Qasimova R.T. About development of knowledge base of national domain name system / **Proceedings of the Third International Conference “Problems of Cybernetics and Informatics”**, Baku, Azerbaijan, 6-8 September, 2010, pp. 171-174.
 6. Qasimova R.T. Milli domen adları ilə bağlı biliklər bazasının yaradılmasının konseptual əsasları haqqında // **Bakı Universitetinin Xəbərləri. Fizika-Riyaziyyat Elmləri Seriyası**, 2010, №4, s. 95-102.
 7. Касумова Р.Т. Об одном подходе к кластеризации регистрационных данных доменных имен / **Седьмая Международная научно-практическая конференция «Интернет – Образование – Наука – 2010»**, г. Винница, Украина, 28 сентября – 3 октября 2010 г., с. 457-458.
 8. Gasimova R.T. On intellectual analysis of domain name registration data // **Информационные технологии и компьютерная инженерия**, 2010, №3 (19) с. 57-61. <http://itce.vntu.edu.ua>
 9. Касумова Р.Т. Сравнительный анализ географических доменов верхнего уровня сети Интернет // **Информационные технологии**, 2011, №7, с. 18-23. <http://www.novtex.ru>
 10. Əliquliyev R.M., Qasimova R.T. Milli domen adları intellektual analiz sisteminin yaradılması // **İnformasiya Texnologiyaları Problemləri**, 2011, №1, s. 29-36. <http://www.jpit.az>
 11. Rasim Alguliyev, Rena Gasimova. On a approach for intellectual analysis of registration data of domain names // **International Journal of Ubiquitous Computing and Internationalization**, April 2011, Vol. 3, No. 1, pp. 27-30. <http://www.itirc.org>
 12. Qasimova R.T. Virtual Azərbaycanın inkişaf dinamikasının qiymətləndirilmə metodikası // **İnformasiya Cəmiyyəti Problemləri**, 2012, №1, s. 52-58. <http://www.jpis.az>
 13. Rena Gasimova, About the Identification of Categorical Registration Data of Domain Names in the Data Warehouse / **The fourth International Conference “Problems of Cybernetics and Informatics”**, Baku, Azerbaijan, 12-14 September, 2012, pp. 132-134.
 14. Qasimova R.T. İnternet mühitdə domen adları, onların seçilməsi, qeydiyyatı və problemləri / **Elektron elm problemləri üzrə I Respublika elmi-praktiki konfransı**, Bakı, Azərbaycan, 15-16 noyabr, 2012, s.75-76.

15. Qasımova R.T. Domen adlarının qeydiyyatı qaydaları və onların müqayisəli təhlili / **Riyaziyyatın tətbiqi məsələləri və yeni informasiya texnologiyaları üzrə II Respublika elmi konfransı**, Sumqayıt, Azərbaycan, 27-28 noyabr, 2012, s. 196-198.
16. Qasımova R.T. Milli domen və kontentlərin xarici serverlərdə yerləşdirilmə səbəbləri və nəticələri / **Ümummilli lider Heydər Əliyevin anadan olmasının 90 illiyinə həsr olunmuş “İnformasiya təhlükəsizliyi problemləri” I Respublika elmi-praktiki konfransı**, Bakı, Azərbaycan, 17-18 May, 2013, s. 52-55.
17. Qasımova R.T. Milli domen və kontentlərin xarici serverlərdə yerləşdirilməsinin monitorinqi problemləri // **İnformasiya Cəmiyyəti Problemləri, Elmi-praktiki jurnal**, 2013, №1 (7), 49-56. <http://www.jpis.az>
18. Rasim Alguliyev, Rena Gasimova Identification of Categorical Registration Data of Domain Names in Data Warehouse Construction Task // **International Journal Intelligent Control and Automation**, 2013, Vol. 4 No. 2, pp. 227-234 (**Google İmpact Factor 0,28**). <http://www.scirp.org/journal/ica>.
19. Qasımova R.T. İnternetdə domen problemləri və onların həlli yolları, Bakı: **“İnformasiya texnologiyaları” nəşriyyatı**, 2012, 164s.

Həmmüəlliflərlə dərc olunmuş işlərdə iddiaçının şəxsi rolu

- (1), (2), (3), (4) – böyük həcmli informasiya resurslarının analizini həyata keçirən intellektual analitik sistemlərin ümumi strukturu işlənmişdir;
- (5), (6) – milli domen adları ilə bağlı biliklər bazasının konsepsiyası işlənmişdir;
- (7), (8) – domen adlarının qeydiyyat verilənlərinin intellektual analizi üçün metod təklif edilmişdir;
- (9) – İnternet şəbəkəsində yüksək səviyyəli coğrafi domen zonalarının monitorinqi və müqayisəli təhlili aparılmışdır;
- (10) – milli domen adları intellektual analiz sisteminin yaradılması üçün konseptual model təklif edilmişdir;
- (12) – Azərbaycanın yüksək səviyyəli domenlərinin inkişaf dinamikasının qiymətləndirilməsi üçün metodika işlənmişdir;
- (13), (18) – domen adlarının qeydiyyat verilənlərinin identifikasiyası üçün qərarların qəbul edilməsi metodu təklif edilmişdir.

РЕНА ТОФИК КЫЗЫ КАСУМОВА

РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ ДЛЯ СИНТЕЗА ИНТЕЛЛЕКТУАЛЬНОЙ МОНИТОРИНГОВОЙ СИСТЕМЫ НАЦИОНАЛЬНЫХ ДОМЕННЫХ ИМЕН

АННОТАЦИЯ

Диссертационная работа посвящена разработке методов и алгоритмов для синтеза интеллектуальной мониторинговой системы поддержки принятия решений, связанных с доменными именами. В рамках диссертационной работы получены следующие основные результаты, имеющие научную новизну:

- Разработана концептуальная модель интеллектуальной мониторинговой системы национальных доменных имен;
- Разработан алгоритм для очистки регистрационных данных доменных имен;
- Разработан метод идентификации регистрационных данных доменных имен с целью создания хранилища данных;
- Предложен неиерархический метод для кластеризации регистрационных данных доменных имен;
- Разработан метод интеллектуального анализа регистрационных данных доменных имен.

RENA TOFIQ QASIMOVA

**DEVELOPMENT OF METHODS AND ALGORITHMS
FOR SYNTHESIS OF INTELLECTUAL MONITORING
SYSTEMS FOR NATIONAL DOMAIN NAMES**

SUMMARY

Dissertation work is devoted to the development of methods and algorithms for synthesis of intellectual monitoring systems for support in decision-making regarding domain names. Following results that have scientific significance were revealed:

- Conceptual model of intellectual monitoring system for national domain names is created;
- Algorithm for cleansing registration data of Domain Names is developed;
- Method for identification of a domain name registration information in order to create a data warehouse is modeled;
- Method for identification of registration domain names in order to create a data warehouse is developed;
- Non-hierarchical clustering method for Domain Name Registration Data is proposed;
- Method for data mining for Domain Name registration data is developed.

Çapa hazırlanmış 07.11.2014. Tirajı 100 nüsxə.
AMEA İnformasiya Texnologiyaları İnstitutunun
“İnformasiya Texnologiyaları” nəşriyyatı

**НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК АЗЕРБАЙДЖАНА
ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

На правах рукописи

РЕНА ТОФИК КЫЗЫ КАСУМОВА

**РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ ДЛЯ
СИНТЕЗА ИНТЕЛЛЕКТУАЛЬНОЙ МОНИТОРИНГОВОЙ
СИСТЕМЫ НАЦИОНАЛЬНЫХ ДОМЕННЫХ ИМЕН**

3338.01 – Системный анализ, управление и обработка информации

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени
доктора философии по технике

БАКУ – 2014