

**AZƏRBAYCAN MİLLİ ELMLƏR AKADEMİYASI
İNFORMASIYA TEXNOLOGİYALARI İNSTİTUTU**

Əlyazması hüququnda

MƏKRUFƏ ŞƏRİF QIZI HACIRƏHİMOVA

**ELEKTRON DÖVLƏT MÜHİTİNDƏ MƏTN SƏNƏDLƏRİN
İNTELLEKTUAL İDARƏ EDİLMƏSİ ÜÇÜN MODEL VƏ
ALQORİTMLƏRİN İŞLƏNMƏSİ**

3338.01 – Sistemli analiz, idarəetmə və informasiyanın işlənməsi

Texnika üzrə fəlsəfə doktoru elmi dərəcəsi
almaq üçün təqdim olunmuş dissertasiyanın

A V T O R E F E R A T I

BAKİ – 2013

İ ş Azərbaycan Milli Elmlər Akademiyası
İnformasiya Texnologiyaları İnstitutunda yerinə yetirilmişdir

Elmi rəhbər:

AMEA-nın müxbir üzvü, texnika
elmləri doktoru, professor

R.M.ƏLİQULİYEV

Rəsmi opponentlər:

texnika elmləri doktoru, dosent

N.B.AĞAYEV

texnika üzrə fəlsəfə doktoru

N.Ə.HƏSƏNOVA

Aparıcı təşkilat:

Azərbaycan Texniki Universitetinin “İnformasiya texnologiyaları və
proqramlaşdırma” kafedrası

Dissertasiyanın müdafiəsi “29” noyabr 2013-cü il, saat 14⁰⁰–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 “ 28 ” oktyabr 2013-cü il tarixdə paylanmışdır.

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

R.H.ŞIXƏLİYEV

İŞİN ÜMUMİ XARAKTERİSTİKASI

İşin aktuallığı. İnformasiya cəmiyyətinə keçidlə əlaqədar olaraq informasiyanın emalı və ötürülməsi formaları, informasiya daşıyıcıları və onlarda informasiyanın saxlanması, bu informasiyaya hüquqi sənəd statusu verən əsas rekvizitlər əhəmiyyətli dərəcədə dəyişmişdir. İnformasiyanın (*sənədin*) yeni forması – elektron sənəd (*e-sənəd*) meydana çıxmışdır. Bu tip sənədlər biznes, vətəndaşlar və hakimiyyət orqanları arasında informasiya mübadiləsinin əsas formasına, onların idarə edilməsində bir alət rolunu oynayan elektron sənədlərin idarə edilməsi sistemləri (ESİS) isə yaradılmaqda və inkişaf etməkdə olan elektron dövlətin (e-dövlət) ən vacib komponentinə çevrilmişdir. Keçən əsrin 90-cı illərindən mövcud olan ESİS-lər ilkin dövrlərdə təşkilatlarda kargüzarlıq fəaliyyətinin avtomatlaşdırılmasını həyata keçirmək üçün yaradılırdısa, hazırda bu sistemlər dövlət idarələri ilə vətəndaşların (G2C -Government to Citizen) və biznes sektorunun (G2B - Government to Business), eyni zamanda dövlət idarələrinin (G2G - Government to Government) həm şaquli, həm də üfqi istiqamətdə elektron qarşılıqlı əlaqəsini təmin edir. E-dövlətdə informasiya sistemlərinin, xüsusilə ESİS-in tətbiqində məqsəd informasiya texnologiyalarından istifadə etməklə vahid informasiya məkanını formalaşdırmaq, dövlət strukturları tərəfindən göstərilən elektron xidmətlərə çıxış imkanlarını genişləndirmək, təşkilatların tabeçiliyində olan və digər təşkilatlarla əlaqələrin sürətləndirilməsini təmin etməyə və nəzarət mexanizminin sadələşdirilməsinə və s. nail olmaqdır. Bu texnologiyanın tətbiqi eyni zamanda dövlət, əhali və biznes sektor arasında qarşılıqlı əlaqə zamanı bürokratik əngəllərin aradan qaldırılmasında, dövlət idarəçiliyində kargüzarlıq fəaliyyətinin səmərəliliyinin və şəffaflığın təmin edilməsində önəmli rol oynayır.

Hazırda informasiya-kommunikasiya texnologiyalarının sürətli inkişafı ilə külli miqdarda e-sənədlər yaradılır və hər il onların sayı eksponensial artır. IDC (*International Data Corporation*) analitik şirkətin tədqiqatında 2009-cu ildə rəqəmsal informasiyanın 62% artaraq 800000 petabayta çatdığı və strukturlaşdırılmamış informasiyanın (*mətnlərin*) isə 85% təşkil etdiyi göstərilir. 2009-2020-ci illər arasında yaradılan informasiyanın həcmnin 44 dəfə artaraq 2020-ci ildə 35 zeta bayt (*35 trilyon gigabayt*) olacağı proqnozlaşdırılır. İnformasiyanın həddindən artıq çoxalması “informasiya yükü” problemi yaradır. Problem e-dövlət qurumlarında tətbiq olunan ESİS-lər üçün daha xarakterikdir. Belə ki, e-dövlət xidmətlərinin populyarlaşması ilə bu sistemlərdə çoxlu sayda e-

sənədlər (*vətəndaşların müraciətləri, ərizə və şikayətləri, biznes sektordan daxil olan sənədlər, dövlət qurumları arasında xidməti yazışmalar və s.*) yaradılır, emal olunur və saxlanılır və dövr edir. Onların əksəriyyəti isə mətn tiplidir. Emal olunan sənədlərin çox böyük sürətlə artması, strukturlaşdırılmamış sənədlərin əksəriyyət təşkil etməsi, bu tip sənədlərin tez bir zamanda rəhbər və məmurlar tərəfindən oxunması, məzmunu barəsində müəyyən fikir əldə olunmasında və düzgün qərar verilməsində, şikayət və problemlərin çeşidlənməsində ciddi problem yaradır. E-dövlətin əsas funksiyalarını effektiv həyata keçirmək üçün bu sənədlər müxtəlif məqsədlər üzrə analiz edilməlidir. Aydındır ki, bu tip e-sənədləri hazırda kargüzarlıqda tətbiq olunan kifayət qədər funksional-texniki imkanlara malik mövcud ESİS və ya verilənlərin idarə edilməsi sistemləri vasitəsi ilə analiz etmək qeyri-mümkündür. Bu mühitdə dinamik xarakter daşıyan sənədlərin axınıni operativ emal etmək, sistemləşdirmək, dövriyyəsinə düzgün təşkil etmək çox vacibdir. E-dövlət mühitində toplanmış e-sənədlərin məzmununa görə təsnifatlandırılması, klasterləşdirmə, referatlaşdırma, ünvanlara çatdırılma – sənədin marşrutunun avtomatik təyini, intellektual axtarışı, zaman keçdikcə biznes dəyəri, əlyetərliyi və mühafizəsinə olan tələblərin dəyişməsindən asılı olaraq onların yaradıldığı andan son ana (*fiziki məhv edilməsi*) qədər keçdiyi mərhələləri daha səmərəli idarə etmək, bu mərhələlərdə sənədləri etibarlı şəkildə yadda saxlamaq, autentikliyinə təmin etmək və e-dövlətin göstərdiyi elektron xidmətlərə vətəndaşların və biznes-sektorun etimadını formalaşdırmaq da çox aktual məsələdir. Problemlərin aradan qaldırılması ilk növbədə ESİS-in intellektuallaşdırılmasına imkan verən model və alqoritmlərin işlənməsini tələb edir. Göstərilən problemlərin mövcudluğu və qəbul olunmuş bir sıra normativ-hüquqi aktlar dissertasiya işinin aktuallığını deməyə əsas verir.

İşin məqsədi: Dövlət qurumlarında idarəetmənin səmərəliliyini artırmaq məqsədilə müasir informasiya texnologiyalarından istifadə etməklə e-dövlət mühitində dövr edən sənədlərin idarə olunması sistemlərində mətn tipli sənədlərin intellektual emalını təmin etmək üçün model və alqoritmlərin işlənməsidir.

Dissertasiya işində qarşıya qoyulmuş məqsədə çatmaq üçün aşağıdakı elmi-nəzəri və praktiki məsələlər qoyulmuş və həll edilmişdir:

- e-sənədlərin idarə edilməsi texnologiyalarının müasir vəziyyətinin analizi, mövcud elmi-nəzəri problemlər və onların həlli yollarının müəyyənləşdirilməsi;

- e-dövlətdə sənədlərin idarə edilməsi sisteminin arxitektura prinsiplərinin işlənməsi;
- e-dövlətdə sənədlərin intellektual idarə edilməsi sisteminin konseptual modelinin işlənməsi;
- e-dövlət mühitində e-sənədlərə etimad modelinin işlənməsi;
- mətn tipli e-sənədlərin referatlaşdırılması üçün yeni model və alqoritmlərin işlənməsi;
- əldə olunan nəticələrin yoxlanılması məqsədi ilə eksperimentlərin aparılması.

Tədqiqat üsulları: Qoyulmuş məsələlərin həlli üçün verilənlər bazası nəzəriyyəsi, informasiya təhlükəsizliyi, text mining, riyazi proqramlaşdırma üsulları, təkamül alqoritmləri istifadə olunmuşdur.

Müdafiəyə çıxarılan əsas müddəalar:

- elektron sənədlərin idarə edilməsi texnologiyalarının müasir vəziyyəti: problemlər və həlli yollarının təhlili;
- e-dövlətin vahid ESİS-nin arxitektura prinsipləri;
- e-dövlətdə elektron sənədlərin intellektual idarə edilməsi sisteminin konseptual modeli ;
- e-dövlət mühitində e-sənədlərin təhlükəsizliyinə etimad modeli;
- ESİS-də mətn sənədlərin referatlaşdırılması üçün optimallaşma modelləri;
- optimallaşma məsələlərinin həlli üçün alqoritmlər.

Elmi yeniliklər: Dissertasiya işində alınmış aşağıdakı əsas nəticələr elmi yenilikləri özündə əks etdirir:

- e-dövlətdə ESİS-nin arxitektura prinsipləri işlənməmişdir;
- e-dövlətdə ESİS-nin intellektuallaşdırılması üçün konseptual model təklif edilmişdir;
- e-dövlət mühitində e-sənədlərə etimad modeli və onun fəaliyyətinə aid tövsiyələr təklif edilmişdir;
- mətn tipli sənədlərin referatlaşdırılması üçün riyazi proqramlaşdırma məsələləri kimi formalizə olunan modellər təklif edilmişdir;
- optimallaşma məsələlərinin həlli üçün diferensial təkamül və “sürü” intellektinə əsaslanan alqoritmlərin modifikasiya olunmuş binar variantları işlənməmişdir.

İşin praktiki əhəmiyyəti: Dissertasiya işində alınmış elmi-nəzəri və praktiki nəticələrin e-dövlət çərçivəsində böyük informasiya axınının intellektual idarə olunması, vətəndaşların müraciətlərinin

cavablandırılmasında, baxılması üçün müvafiq qurumlara və ya struktur bölmələrə göndərilməsində, məmurlar tərəfindən daha uyğun və səmərəli qərarların qəbul edilməsində, sənədlərin qısa məzmununun əldə olunmasında, məzmunu görə avtomatik təsnifatlandırılmasında və axtarışında geniş tətbiqinin mümkünlüyü ilə müəyyən olunur.

İşin nəticələrinin realizasiyası və tətbiqi. Dissertasiya işinin əsas elmi-nəzəri nəticələri AMEA fizika-riyaziyyat və texnika elmləri bölməsi tərəfindən mühüm nəticə kimi 2011-ci ilin müvafiq illik hesabatına daxil edilmişdir. Dissertasiya işində təklif olunmuş model və alqoritmlər əsasında korporativ ESİS layihələndirilmiş, sistemin proqram təminatı işlənilib hazırlanmış və AzScienceNet şəbəkəsində tətbiq edilmişdir və müvafiq akt alınmışdır.

İşin aprobasiyası: Dissertasiya işinin əsas elmi-nəzəri və praktiki nəticələri aşağıdakı siyahıda təqdim olunan bir sıra beynəlxalq və respublika səviyyəli konfranslarda məruzə və müzakirə olunmuşdur: международная научно-техническая конференция «Информационные системы и технологии» (Новосибирск, Россия, 22-25 апрель 2003); 9-ая международная научная конференция «Теория и техника передачи, приема и обработки информации» (Тяпсе, Украина, 7-10 октябрь 2003); 10-ая международная научная конференция «Теория и техника передачи, приема и обработки информации» (Тяпсе, Украина, 28 сентября-1 октября 2004); “İnformasiyalasdırma, Kibernetika və İnformasiya Texnologiyalarının müasir problemləri” 2-ci respublika konfransı (Bakı, Azərbaycan, 26-28 oktyabr 2004); first International Conference «Problems of Cybernetics and Informatics» – PCI’2006 (Baku, Azerbaijan, 24-26 October, 2006); third International Conference «Problems of Cybernetics and Informatics» – PCI’2010 (Baku, Azerbaijan, 6-8 September, 2010); «E-elm problemləri» üzrə I respublika elmi-praktiki konfransı (Bakı, Azərbaycan, 15-16 noyabr 2012).

Elmi nəşrlər: Dissertasiya mövzusu üzrə 16 elmi iş çap olunmuşdur. Onlardan 9 məqalə resenziya olunan jurnallarda və 7 məruzə isə beynəlxalq və respublika konfranslarının materiallarında dərc olunmuşdur.

İşin strukturu və həcmi: Dissertasiya işi giriş, dörd fəsil, nəticə, 198 adda ədəbiyyat siyahısı və əlavədən ibarətdir. İşin əsas məzmunu 126 səhifə, 24 şəkil və 16 cədvəldən ibarətdir.

İŞİN MƏZMUNU

Girişdə dissertasiya işinin aktuallığını şərtləndirən əsas amillər müəyyən edilmiş, sistemin intellektuallaşdırılması istiqamətində tədqiqat

işinin məqsədi formalaşdırılmış, məqsədə çatmaq üçün məsələlər müəyyən edilmiş, müdafiəyə çıxarılan əsas müddəalar açıqlanmış, alınmış nəticələrin elmi yeniliyi və praktiki əhəmiyyəti göstərilmişdir.

Birinci fəsil e-sənədlərin idarə olunması texnologiyalarının müasir vəziyyətinin araşdırılmasına həsr olunmuşdur. Bu məqsədlə ilk növbədə e-sənədlərin idarə olunması texnologiyaları üzrə əsas anlayışlar, baza funksiyaları, idarəetmədə faydası, inkişaf mərhələləri və perspektivləri, mövcud sistemlər və onların təsnifatı, beynəlxalq və inkişaf etmiş ölkələrin standartları, ESİS sənayesi sahəsində aparılan texnoloji həllər və elmi tədqiqatlar, bu sistemlərin yaradılmasında müxtəlif obyekt yönümlü, veb əsasında yanaşmalar, konseptual, arxitektura, semantik, təhlükəsizlik, saxlama, informasiya və s. modelləri araşdırılmışdır.

Daha sonra son illərdə bu sistemlərin intellektuallaşdırılmasında əsas istiqamətlərdən biri kimi strukturlaşdırılmamış – mətn tipli sənədlərin intellektual analizində tətbiq olunan – *text mining* texnologiyası və onun həll etdiyi mətnlərin məzmununa görə avtomatik təsnifatlandırma, klasterləşdirmə və referatlaşdırma metod və alqoritmləri geniş tədqiq olunmuşdur.

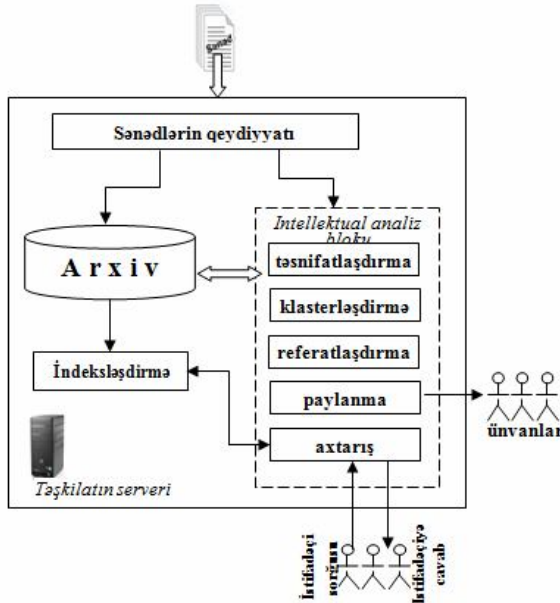
Fəslin sonunda isə dövlət qurumlarında ESİS-lərin tətbiqi zamanı mövcud problemlərə baxılmış və həlli yolları göstərilmişdir. Dissertasiya işinin aktuallığını əsaslandıran əsas elmi-nəzəri problemlər dəqiqləşdirilərək, konkret model və alqoritmlərin işlənməsi tövsiyə olunmuşdur.

İkinci fəsil e-dövlət mühitində ESİS-lərin sintezi üçün modellər işlənməmişdir. İlk növbədə ayrı-ayrı hakimiyyət orqanlarının (mərkəzi, regional və yerli özünüidarəetmə) korporativ informasiya fəzalarından təşkil olunmuş e-dövlətin vahid ESİS-nin arxitektura prinsipləri işlənməmişdir. Dövlət hakimiyyəti orqanlarında elektron sənəd dövriyyəsinin tətbiqi zamanı, baza funksional imkanlarla yanaşı, bu sistemlərdə insan əməyini yüngülləşdirən, sistemin işinin səmərəliliyini artıran intellektual analiz blokundan təşkil olunmuş konseptual model təklif edilmişdir (şəkl.1).

Təklif olunan modeldə “informasiya yükü” problemi təsnifatlandırma, klasterləşmə və referatlaşdırma ilə azaldılmaqla və həll kimi birgə istifadəsi nəzərdə tutulmuşdur. Təsnifatlandırma və klasterləşdirmə məzmunca yaxın sənədləri qruplaşdırmağa imkan verir, referatlaşdırma isə sənəd və ya sənəd çoxluğundan ən vacib mətn vahidlərini çıxarmaqla, onu istifadəçiyə qısaldılmış formada və anlaşıqlı şəkildə təqdim etmək, lazımi informasiyanın filtirlənməsi və seçilməsinə kömək edir. Böyük həcmdə informasiyalarla işləməkdə daha effektiv görünən avtomatik mətn referatlaşdırılması müraciətlərə cavabların hazırlanması və ya baxılması üçün müvafiq quruma

yönəldilməsi, operativ qərarların qəbul edilməsində məmurlara dəstək ola biləcək intellektual ESİS üçün çox vacib komponentdir. Mətn referatlaşması, istifadəçiyə böyük həcmli sənəddəki informasiyanı dərk etməyə imkan verən, daha uyğun informasiyanı göstərməklə informasiya axtarışını sadələşdirməyə və axtarış müddətinin azaldılmasına kömək edir.

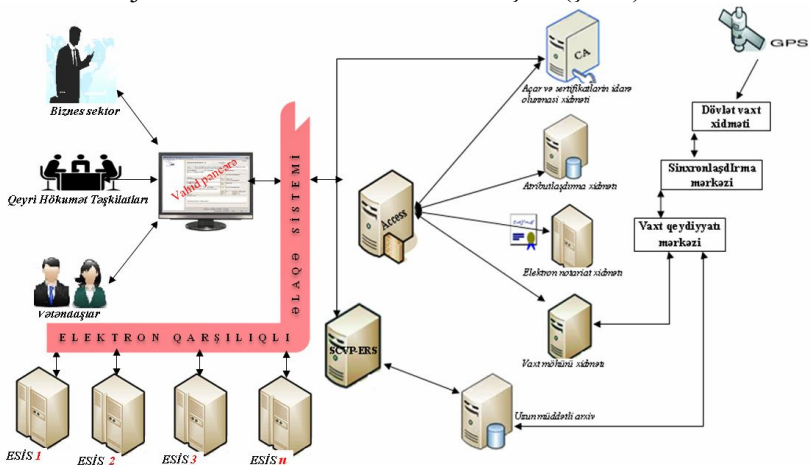
Modeldə məzmunu görə təsnifatlandırma və referatlaşmadan ESİS-də sənədlərin avtomatik olaraq ünvanlara çatdırılması məsələsinin həllində də istifadəsi mümkündür. Bu, həm istifadəçilərin vaxtına qənaət edir, həm də axtarış sistemlərinin xidmət keyfiyyətini daha da yaxşılaşdırır.



Şək.1. ESİS-in intellektuallaşdırılması sxemi

Bu fəsilə baxılan ən mühüm məsələlərdən biri də e-dövlət mühitində yaradılan, dövr edən və toplanan e-sənədlərin bütün həyat dövründə autentiqliyinin təmin olunmasıdır. İnformasiya təhlükəsizliyi xidmətləri həyata keçirilərkən elektron imza (e-imza) və digər kriptografik texnologiyalar istifadə olunur. E-imza açıq açarlar infrastrukturu (PKI – Public Key Infrastructure) vasitəsi ilə həyata keçirilir və sertifikat mərkəzləri (CA – Certificate Authority) tərəfindən verilən açıq açarlar sertifikatına olan etimada əsaslanır. Lakin, e-sənədlərin bütün həyat dövründə autentiqliyini təmin etmək üçün adi e-imzanın istifadəsi kifayət deyildir. E-imzanın

etibarlılığı sertifikatın etibarlılıq müddəti qurtardıqda yaxud sertifikat geri çağırıldıqda itir. Bu, mükəmməl ESİS yaratmağa imkan vermir. Bu məqsədlə Beynəlxalq Telekomunikasiya İttifaqının (International Telecommunication Union – ITU) tövsiyələri və Avropa Telekomunikasiya Standartları İnstitutunun (European Telecommunications Standards Institute – ETSI) standartı əsasında yaradılmış təkmilləşdirilmiş e-imzanın tətbiqi ilə açar və sertifikatların idarə olunması, identifikasiya, autentifikasiya və daxilolmalara nəzarət, vaxt möhürü (*time stamping*), elektron notarius, elektron arxiv, atributlaşdırma kimi əsas xidmətlərdən ibarət e-sənədlərə etimad infrastrukturunun funksional-texnoloji arxitektura modeli təklif edilmişdir (şəkl.2).



Şəkl.2. E-dövlətdə sənədlərə etimad infrastrukturunun arxitektura modeli

Üçüncü fəsil e-dövlətdə sənədlər axınının intellektual analizində ən perspektivli istiqamətlərdən sayılan mətn sənədlərin avtomatik referatlaşdırılması məsələsinə həsr olunmuşdur. Ümumiyyətlə, mətn referatlaşdırmasının məqsədi mətndən elə cümlələr çoxluğunu seçməkdir ki, o sənədin əsas məzmununu özündə əks etdirsin. Başqa sözlə elə referat yaratmaq lazımdır ki, sənəd kolleksiyası ilə referat arasındakı yaxınlıq maksimum olsun. Referatlaşdırmanın əsas problemi: ölçü; informativ cümlələrin və tematik bölmələrin aşkarlanması; izafilik – referatda eyni məna daşıyan cümlələrin təkrarlanmaması; yaxınlıq ölçüsünün seçilməsi və həll algoritmidir. Referatda seçilmiş cümlələrin hər biri fərdi şəkildə vacib olmalıdır, məzmunca bir-birini təkrarlamamalıdır və referatın uzunluğu nəzərə alınmalıdır. Ən yaxşı cümlələrin seçim proseduru ilə müqayisədə ən

yaxşı referatın seçilməsi qlobal optimallaşma problemidir. Problemin həlli ilə əlaqədar olaraq avtomatik mətn referatlaşdırma üçün maksimum əhatəli və izafiliyi idarə edən optimallaşma modelləri təklif edilmiş və referatlaşdırma məsələsinin həlli üçün təkamül alqoritmləri işlənmişdir.

Tutaq ki, D sənəd kolleksiyası verilmiş və o, cümlələr çoxluğu kimi təsvir edilmişdir, $D = \{s_1, s_2, \dots, s_n\}$. Burada s_i D -dəki i -ci cümləni göstərir, n isə sənəd çoxluğundakı cümlələrin sayıdır. Mətn vahidləri (*sənəd*, *cümlə*) arasındakı yaxınlığı hesablamaq üçün *kosinus yaxınlıq ölçüsündən* istifadə edirik. Mətn analizində tez-tez istifadə olunan kosinus metrikası iki vektor arasındakı bucağın kosinusunu hesablayır. İki cümlə, $s_i = \{w_{i1}, w_{i2}, \dots, w_{im}\}$ və $s_j = \{w_{j1}, w_{j2}, \dots, w_{jm}\}$ arasındakı kosinus yaxınlığı (1) düsturu ilə hesablanır:

$$\text{sim}(s_i, s_j) = \cos(s_i, s_j) = \frac{\sum_{k=1}^m w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^m w_{ik}^2 \cdot \sum_{k=1}^m w_{jk}^2}}, \quad i, j = 1, \dots, n \quad (1)$$

burada w_{ik} – s_i cümləsində t_k termininin çəkisidir və *tf-isf* (*term frequency – inverse sentence frequency*) sxemini istifadə etməklə hesablanır:

$w_{ij} = f_{ij} \times \log(n / n_j)$, burada $f_{ij} = m_{ij} / m - t_j$ sözünün s_i cümləsində işlənmə tezliyi; m_{ij} – t_j sözünün s_i cümləsində rast gəlmə sayı; n – cümlələrin ümumi sayı; n_j – t_j sözünün rast gəldiyi cümlələrin sayıdır, $i = \overline{1, n}, j = \overline{1, m}$.

Fərz edək ki, D çoxluğunda hər bir cümlə referata daxil edilmək şansına malikdir, onda referatlaşdırma məsələsi aşağıdakı xətti proqramlaşdırma məsələsi kimi formalizə oluna bilər:

$$f(X) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sim}(O, s_i) \cdot \text{sim}(O, s_j) x_{ij} - \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sim}(s_i, s_j) x_{ij} \rightarrow \max, \quad (2)$$

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n [\text{len}(s_i) + \text{len}(s_j)] x_{ij} \leq L, \quad (3)$$

$$x_{ij} \in \{0, 1\}, \quad \forall i, j \quad (4)$$

burada s_i və s_j cümlələr cütü referata daxil edilmək məqsədilə seçilmişsə, x_{ij} dəyişənləri 1, əks halda isə 0 -a bərabərdir, L referatın və $\text{len}(s_i)$ isə s_i

cümləsinin uzunluğunu bildirir. (2)-də $O - D = \{s_1, s_2, \dots, s_n\}$ kolleksiyasının mərkəzini bildirir və onun k -cı koordinatı o_k aşağıdakı düsturla hesablanır:

$$o_k = \frac{1}{n} \sum_{i=1}^n w_{ik} . \quad (5)$$

(2)-də $\text{sim}(O, s_i)$ həddi s_i cümləsinin sənədin məzmun əhatəliliyini müəyyən edir. (2)-(4) tamədərli xətti proqramlaşdırma (TXP) məsələsidir. Burada (2) məqsəd funksiyası və (3) məhdudiyyəti (4) dəyişənlərinə nəzərən xəttidirlər. (2)-dəki birinci hədd mənbənin əsas məzmununun referat tərəfindən əhatə ediləcəyini, ikinci hədd isə referatda minimum izafiliyi təmin edir. Seçim prosesi referatdakı cümlələrin uzunluğu cəminin uzunluq həddinə (3) çatmasına qədər davam etdirilir.

(2)-(4)-də təklif olunan çoxkriteriyalı TXP məsələsini həll etmək üçün budaq və sərhəd metoduna əsaslanan hazır paketlərdən istifadə olunmuşdur.

Fərz edək ki, hər bir cümlə referat üçün seçilmək şansına malikdir. Onda maksimum əhatəliliyi və minimum izafiliyi təmin edən referatlaşdırma modelini aşağıdakı kimi formalizə etmək olar:

$$f = \sum_{i=1}^{n-1} \sum_{j=i+1}^n [\text{sim}(\vec{D}, \vec{s}_i) + \text{sim}(\vec{D}, \vec{s}_j) - \text{sim}(\vec{s}_i, \vec{s}_j)] x_{ij} \rightarrow \max, \quad (6)$$

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n [\text{len}(s_i) + \text{len}(s_j)] x_{ij} \leq L, \quad (7)$$

$$x_{ij} \in \{0,1\}, \quad \forall i, j, \quad (8)$$

burada $\vec{s}_i$ s_i cümləsinin əlamətlər vektorudur.

(6)-(8) TXP məsələsidir. (6)-dəki birinci hədd mənbənin əsas məzmununun referat tərəfindən əhatə ediləcəyinə zəmanət verir. Bu funksiya həmçinin zəmanət verir ki, referatda eyni informasiyanı daşıyan təkrar mətn hissələri olmayacaqdır. Yəni referatdakı izafilik minimum olacaqdır. (7) referatın uzunluğunu məhdudlaşdıran şərtidir.

Yaxınlıq ölçüsü mətn referatlaşdırmasında mühüm rol oynayır. Kosinus və NGD-yə (Normalized Google Distance – Normallaşdırılmış Google Məsafəsi) əsaslanan ölçünün referatın nəticəsinə necə təsir etdiyini öyrənmək üçün son məqsəd funksiyası aşağıdakı kimi təklif edilmişdir:

$$f_{\alpha} = \alpha \cdot f_{\cos} + (1-\alpha) \cdot f_{NGD} \rightarrow \max, \quad (9)$$

burada

$$f_{\cos} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n [sim_{\cos}(\vec{D}, \vec{s}_i) + sim_{\cos}(\vec{D}, \vec{s}_j) - sim_{\cos}(\vec{s}_i, \vec{s}_j)] x_{ij} , \quad (10)$$

$$f_{\text{NGD}} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n [sim_{\text{NGD}}(\vec{D}, \vec{s}_i) + sim_{\text{NGD}}(\vec{D}, \vec{s}_j) - sim_{\text{NGD}}(\vec{s}_i, \vec{s}_j)] x_{ij} . \quad (11)$$

α çəki parametri $[0,1]$ intervalında qiymətlər alır. Əgər $\alpha = 0$, onda $f_{\alpha} = f_{\text{NGD}}$ -yə bərabərdir; əgər $\alpha = 1$, onda $f_{\alpha} = f_{\cos}$ -a bərabərdir; və əgər $\alpha = 0.5$, onda kosinus və NGD əsasında ölçünün eyni dərəcədə vacib olduğu fərz edilir. $\vec{s}_i$ və $\vec{s}_j$ vektorları arasındakı kosinus yaxınlığı (1) düsturu ilə hesablanır. İki cümlə arasındakı NGD əsasında yaxınlığı hesablamaq üçün onların (*cümlələrin*) hər biri terminlər çoxluğu kimi təsvir olunur, s_i və s_j cümlələri arasındakı yaxınlıq aşağıdakı kimi hesablanır:

$$sim_{\text{NGD}}(s_i, s_j) = \frac{\sum_{t_k \in s_i} \sum_{t_l \in s_j} sim_{\text{NGD}}(t_k, t_l)}{|s_i| \cdot |s_j|} , \quad (12)$$

burada $sim_{\text{NGD}}(t_k, t_l) = \exp(-\text{NGD}(t_k, t_l))$, (13)

t_k və t_l terminləri arasında NGD əsasında yaxınlıq ölçüsüdür. (13)-də $\text{NGD}(t_k, t_l)$ - t_k və t_l terminləri arasındakı Normallaşdırılmış Google Məsafəsidir və aşağıdakı kimi hesablanır:

$$\text{NGD}(t_k, t_l) = \frac{\max\{\log(f_k), \log(f_l)\} - \log(f_{kl})}{\log n - \min\{\log(f_k), \log(f_l)\}} , \quad (14)$$

burada f_k - t_k termininin, f_{kl} - t_k və t_l terminlərinin hər ikisinin rast gəlinədiyi cümlələrin, n isə sənəd kolleksiyasındakı cümlələrin sayını bildirir.

Optimallaşma məsələsinin həlli üçün “sürü” intellektinə əsaslanan optimallaşma alqoritminin yeni binar versiyası işlənmişdir.

Maksimum əhatəliliyi və minimum izafiliyi təmin edən növbəti mətn referatlaşdırma modelində referatlaşdırma məsələsi kvadratik Bul proqramlaşdırma (*Quadratic Boolean Programming*) məsələsi kimi formalizə olunmuşdur.

Əgər sənəddə bir neçə mövzudan bəhs edilsə, onda bütün mövzuları əhatə edə biləcək referatın yaradılması çətin olur. Çətinlik ilk növbədə sənəddə bəhs edilən mövzuların aşkarlanması və hər mövzunu əhatə edə biləcək informativ cümlələrin seçilməsi ilə bağlıdır. Tək-sənəddən fərqli olaraq, çox sayda sənəd referatlaşdırmasında izafilik, müxtəlif sənədlərdə

cümlələrin eyni informasiyanı ifadə etdiyindən, xüsusilə vacibdir. Əvvəlcə iki məqsəd funksiyası təyin edilmişdir:

1) $f_{\text{cov}}(s_i)$: referatda iştirak edən s_i cümləsinin məzmun əhatəliliyi:

$$f_{\text{cov}}(s_i) = \text{sim}(s_i, O), \quad i = 1, \dots, n, \quad (15)$$

burada $O = \{s_1, s_2, \dots, s_n\}$ sənəd kolleksiyasının mərkəzidir və onun əsas məzmununu əks etdirir. (15) düsturu sənədin O mərkəzinə yaxınlığını ölçməklə, s_i cümləsinin vacibliyini qiymətləndirir. $O = [o_1, o_2, \dots, o_m]$ mərkəzinin k -cı koordinatı (4) düsturu ilə hesablanır. $f_{\text{cov}}(s_i)$ -in daha yüksək qiyməti, s_i cümləsinin daha yüksək məzmun əhatəliliyinə uyğundur.

2) $f_{\text{red}}(s_i, s_j)$: s_i və s_j , cümlələri arasındakı izafilik:

$$f_{\text{red}}(s_i, s_j) = 1 - \text{sim}(s_i, s_j), \quad i \neq j = 1, \dots, n. \quad (16)$$

$f_{\text{red}}(s_i, s_j)$ məqsəd funksiyasının daha böyük qiyməti s_i və s_j , cümlələrinin məzmunca fərqli olmasına uyğundur, yəni o, referatda minimum izafiliyi (*müxtəliflik*) təmin edir. İzafiliyi minimallaşdırarkən, yüksək keyfiyyətli referat verilən sənəd toplusunun məzmununu maksimum əhatə etməlidir, onda referatlaşdırma məsələsi aşağıdakı optimallaşma məsələsi kimi formalizə oluna bilər:

$$f(X) = w \cdot f_{\text{cov}}(X) + (1-w) \cdot f_{\text{red}}(X) = w \cdot \sum_{i=1}^{n-1} \text{sim}(s_i, O) x_i + \quad (17)$$

$$+ (1-w) \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n (1 - \text{sim}(s_i, s_j)) x_i x_j \rightarrow \max,$$

$$\sum_{i=1}^n l_i x_i \leq L, \quad (18)$$

$$x_i \in \{0, 1\}, \quad \forall i. \quad (19)$$

burada x_i binar dəyişəndir, belə ki, s_i seçilmişdirsə $x_i = 1$; əks halda $x_i = 0$.

L - referatda verilmiş uzunluq həddidir, l_i isə s_i cümləsinin uzunluğunu göstərir. Referat və cümlənin uzunluğu sözlərin sayı və ya baytlarla ölçülür.

(17)-(19) optimallaşma modeli referatın məzmun əhatəliliyi və müxtəlifliyini balanslaşdırır. Bu modeldə iki məqsəd funksiyasını, (15) və (16) düsturlarını (17) skalyar düsturunda birləşdirmək üçün w parametri istifadə olunur. $w = f_{\text{cov}}(\cdot)$ və $f_{\text{red}}(\cdot)$ funksiyalarının nisbətini müəyyən edən çəki parametridir. $w = 1$ olduqda bu model müxtəliflikdən asılı olmayaraq

sənədin məzmununa maksimum relevant olan cümlələrə üstünlük verir. $w = 0$ olduqda, model məzmun əhatəliliyindən asılı olmayaraq digərlərindən fərqlənən cümlələrə üstünlük verir. w parametrinin qiymətini dəyişməklə və (17)-(19) kvadratik Bul proqramlaşdırma məsələlərini ardıcıl həll etməklə, maksimum əhatəli ($w = 1$) və minimum izafiliklə ($w = 0$), effektiv referatı tapmaq olar. Əgər $w = 0.5$ -dirsə $f_{\text{cov}}(\cdot)$ və $f_{\text{red}}(\cdot)$ funksiyalarının eyni dərəcədə vacib olduğu fərz edilir. Eksperimentdə $w = 0.75$ götürülmüşdür.

s_i və s_j , cümlələri arasındakı yaxınlıq aşağıdakı düsturla təyin olunmuşdur:

$$\text{sim}(s_i, s_j) = \sum_{l=1}^m \left[\left(\sum_{k=1}^m \omega_{ik} p_{kl} \right) \left(\sum_{k=1}^m \omega_{jk} p_{kl} \right) \right], \quad (20)$$

burada p_{kl} k -cı və l -ci sözlər arasındakı yaxınlıqdır və $p_{kl} = \exp(-NGD_{kl})$ düsturu ilə ölçülür. Burada NGD_{kl} k -cı və l -ci sözlər arasındakı Normallaşdırılmış Google Məsafəsidir və (14) düsturu ilə hesablanır.

(17)-(19) optimallaşma məsələsi binar differensial təkamül alqoritmi vasitəsi ilə həll edilmişdir.

Dördüncü fəsilə təklif olunmuş model və alqoritmlər eksperimental yoxlanılmışdır. AzScienceNet şəbəkəsində korporativ ESİS layihələndirilmiş, sistemin proqram təminatı işlənib hazırlanmışdır və sınaqdan keçirilmişdir.

Təklif olunmuş mətn referatlaşdırma modellərinin effektivliyini yoxlamaq üçün testlər aparılmışdır. Test üçün DUC (Document Understanding Conference) sənədlər toplusundan istifadə olunmuşdur (DUC2002, DUC2004, DUC2005, DUC2006, DUC2007). Təklif edilmiş modellər əsasında yaradılmış referatlar DUC tərəfindən qəbul edilmiş rəsmi qiymətləndirmə sistemi – ROUGE-1.5.5 (Recall-Oriented Understudy for Gisting Evaluation) paketi istidadə edilməklə qiymətləndirilmişdir. Avtomatik yaradılmış referatla etalon referatlar arasında yaxınlıq dərəcəsi ROUGE ölçüsü ilə qiymətləndirilmişdir:

$$\text{ROUGE} - N = \frac{\sum_{S \in \text{Summ}_{\text{ref}}} \sum_{N\text{-gram} \in S} \text{Count}_{\text{match}}(N\text{-gram})}{\sum_{S \in \text{Summ}_{\text{ref}}} \sum_{N\text{-gram} \in S} \text{Count}(N\text{-gram})},$$

burada $N - N$ -qramın uzunluğunu göstərir; $\text{Count}_{\text{match}}(N\text{-gram})$ – avtomatik yaradılmış və etalon referatlarda üst-üstə düşən N -qramların maksimal sayı; $\text{Count}(N\text{-gram})$ – etalon referatlardakı N -qramların sayıdır.

Təcrübələr zamanı təklif olunmuş birinci model (**ILPS**) doqquz modellə müqayisə edilmişdir. Təklif olunmuş model digər modellərin orta ROUGE qiymətləri ilə müqayisədə daha yaxşı nəticələr göstərmişdir. Təcrübənin nəticələri cədvəl 1-də verilmişdir.

Cədvəl 1

Modellərin ROUGE qiymətləri

Modellər	ROUGE-1	ROUGE-2
MCKP	0.30908	0.08042
MICW	0.32714	0.08609
Centroid	0.36728	0.07379
Reinforcement	0.37082	0.08351
LexPageRank	0.37842	0.08572
MDS	0.37934	0.08934
DUCbest	0.38224	0.09216
BSTM	0.39065	0.09010
ILPS	0.39121	0.09113
WCS	0.39825	0.09641

Cədvəl 1-dən göründüyü kimi ILPS modeli ROUGE-1 və ROUGE-2 qiymətlərinin hər ikisində yalnız WCS, ROUGE-2 qiymətlərində isə ancaq DUCbest-dən zəif nəticə göstərir. Onun nəticələri BSTM modelinin nəticələrinə yaxındır, digərlərindən isə yaxşı nəticə göstərir. Digər modellərin ROUGE qiymətləri ilə müqayisədə təklif olunmuş model mühüm yaxşılaşma əldə edir və onun nəticələri cədvəl 2-də verilmişdir.

Cədvəl 2

ILPS modelinin yaxşılaşması

Modellər	Yaxşılaşma (%)			
	ROUGE-1		ROUGE-2	
DUCbest	0.00897	(2.35)	−0.00103	(−1.12)
MICW	0.06407	(19.58)	0.00704	(8.17)
MCKP	0.08213	(26.57)	0.01271	(15.80)
Reinforcement	0.02039	(5.50)	0.00962	(11.52)
BSTM	0.00056	(0.14)	0.00303	(3.36)
LexPageRank	0.01279	(3.38)	0.00741	(8.64)
Centroid	0.02393	(6.52)	0.01934	(26.21)
WCS	−0.00704	(−1.77)	−0.02262	(−3.40)
MDS	0.01187	(3.13)	0.00379	(4.24)

Təklif edilmiş **MCMR** modelini qiymətləndirmək üçün DUC2005 və DUC2007 sənədlər toplusunda bir neçə modellə müqayisə edilmişdir.

Cədvəl 3-də DUC2005 sənədlər toplusunda aparılmış müqayisənin nəticəsi verilmişdir.

Cədvəl 3

DUC2005 sənədlər toplusunda modellərin ROUGE qiymətləri

Modellər	ROUGE-2	ROUGE-SU4
MCMR(B&B)	0.0790	0.1392
MCMR(PSO)	0.0754	0.1360
TranSumm	0.0755	0.1366
QEA	0.0749	0.1333
Content-term	0.0718	0.1338
Biased LexRank	0.0753	0.1363
Qs-MRC	0.0779	0.1366
TMR+TF	0.0715	0.1304

Digər referatlaşdırma sistemləri ilə təklif edilmiş modelin ROUGE qiymətlərinin müqayisəsindəki nisbi yaxşılaşma cədvəl 4-də təqdim edilmişdir. Məsələn, Qs-MRC modeli ilə müqayisə göstərir ki, MCMR(B&B) modeli ROUGE-2 və ROUGE-SU4 göstəricilərini 1.41% və 1.90%-ə qədər yaxşılaşdırır. Cədvəl 4-də altından xətt çəkilmiş qiymətlər onların MCMR (B&B)-in qiymətlərindən əhəmiyyətli dərəcədə fərqli olduğunu bildirir.

Cədvəl 4

DUC2005 sənədlər toplusunda MCMR(B&B)-in digər modellərlə müqayisəsi

Modellər	Nisbi yaxşılaşma (%)	
	ROUGE-2	ROUGE-SU4
MCMR(B&B)	x	x
MCMR(PSO)	4.77	2.35
TranSumm	4.64	1.90
QEA	<u>5.47</u>	<u>4.43</u>
Content-term	<u>10.03</u>	<u>4.04</u>
Biased LexRank	4.91	2.13
Qs-MRC	1.41	1.90
TMR+TF	<u>10.49</u>	<u>6.75</u>

(9)-da α çəki parametrinin təsirini tədqiq etmək üçün onun qiyməti $[0;1]$ intervalında dəyişdirilmişdir.

Cədvəl 5-də α çəki parametrinin f_α modelinin ROUGE-2 və ROUGE-SU4 qiymətlərinə təsiri göstərilmişdir

Cədvəl 5

α parametrinin f_α modelinin ROUGE-2 və ROUGE-SU4 qiymətlərinə təsiri

α parametri	ROUGE-2		ROUGE-SU4	
	DUC2005	DUC2007	DUC2005	DUC2007
0.00	0.0781	0.1185	0.1366	0.1716
0.05	0.0781	0.1186	0.1367	0.1721
0.10	0.0782	0.1189	0.1368	0.1726
0.15	0.0783	0.1194	0.1368	0.1728
0.20	0.0783	0.1201	0.1370	0.1733
0.25	0.0784	0.1208	0.1372	0.1737
0.30	0.0784	0.1209	0.1373	0.1740
0.35	0.0784	0.1211	0.1375	0.1742
0.40	0.0785	0.1213	0.1378	0.1745
0.45	0.0786	0.1217	0.1382	0.1747
0.50	0.0787	0.1216	0.1385	0.1750
0.55	0.0788	0.1221	0.1388	0.1753
0.60	0.0789	0.1218	0.1388	0.1749
0.65	0.0790	0.1216	0.1392	0.1746
0.70	0.0788	0.1215	0.1389	0.1742
0.75	0.0787	0.1213	0.1386	0.1738
0.80	0.0787	0.1211	0.1381	0.1736
0.85	0.0786	0.1209	0.1378	0.1731
0.90	0.0784	0.1209	0.1375	0.1727
0.95	0.0783	0.1208	0.1372	0.1722
1.00	0.0783	0.1207	0.1371	0.1721

$\alpha = 0$ qiyməti f_{NGD} (10) və $\alpha = 1$ f_{cos} (11) funksiyasını təqdim edir. Buradan görürük ki, α 0 və 1 arasında dəyişərkən, DUC2005 sənədlər toplusunun optimal qiyməti $\alpha = 0.65$ və DUC2007 sənədlər toplusunun optimal qiyməti isə $\alpha = 0.55$ -dir.

Təklif olunmuş **MCLR** modelinin səmərəliliyini qiymətləndirmək üçün bir neçə eksperiment aparılmışdır: onun göstəricisi DUC sənədlər toplusunda doqquz ən yaxşı referatlaşdırma modellərinin nəticələri ilə müqayisə edilmiş; w çəki parametrinin ROUGE qiymətlərinə təsiri öyrənilmişdir.

Cədvəl 6-da DUC2006 sənədlər toplusunu istifadə etməklə müxtəlif modellərin ROUGE qiymətləri göstərilmişdir. Seçilmiş yazılarla verilmiş nəticələr təcrübələr toplusundakı ən yaxşı nəticələri, mötərizədəki rəqəmlər isə verilənlər toplusundakı hər bir modelin rəqəmini göstərir.

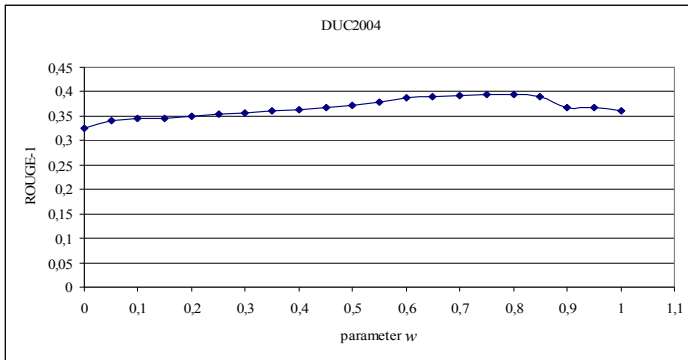
Cədvəl 6

Modellərin DUC2006 toplusunda ROUGE göstəricilərinin müqayisəsi

Modellər	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-SU
DUCbest	0.3796 (4)	0.0754 (4)	0.3476 (4)	0.1321 (4)
MCLR	0.3975 (1)	0.0850 (2)	0.3674 (2)	0.1385 (2)
Random	0.3175 (11)	0.0489 (11)	0.2938 (11)	0.1008 (11)
Centroid	0.3639 (7)	0.0726 (7)	0.3425 (7)	0.1262 (7)
LexRank	0.3666 (6)	0.0733 (6)	0.3442 (6)	0.1288 (6)
LSA	0.3308 (9)	0.0502 (10)	0.3051 (9)	0.1023 (10)
NMF	0.3237 (10)	0.0550 (9)	0.3006 (10)	0.1061 (9)
KM	0.3637 (8)	0.0618 (8)	0.3411 (8)	0.1250 (8)
FGB	0.3713 (5)	0.0748 (5)	0.3450 (5)	0.1308 (5)
BSTM	0.3898 (3)	0.0848 (3)	0.3525 (3)	0.1365 (3)
SNMF+SLSS	0.3955 (2)	0.0855 (1)	0.3680 (1)	0.1398 (1)

Cədvəl 6-dan göründüyü kimi MCLR modeli ilə müqayisədə SNMF+SLSS və DUCbest modellərinin yaxşılaşması çox da böyük deyildir. Məsələn, MCLR modeli ilə müqayisə olunmuş SNMF+SLSS modeli ROUGE-2, ROUGE-L və ROUGE-SU göstəricilərini uyğun olaraq 0,59%, 0,16% və 0,94%-ə qədər yaxşılaşdırır.

(15) və (16) məqsəd funksiyalarının (17) məqsəd funksiyasının referat göstəricisinə nisbi təsirini müəyyən etmək üçün şəkil 3-də DUC2004 sənədlər toplusunda modelin ROUGE-1 göstəricisinin w parametrindən asılılığı göstərilmişdir. Göründüyü kimi $w=0$ ilə müqayisədə $w>0$ olduqda daha yaxşı referat göstəricisi müşahidə olunur. Həmçinin müşahidə edirik ki, $w=0.75$ ən yaxşı göstəricidir.



Şək.3. DUC2004 sənədlər toplusunda MCLR modelinin ROUGE-1 göstəricisinin w parametrindən asılılığı

DISSERTASIYA İŞİNİN ƏSAS NƏTİCƏLƏRİ

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

- E-sənədlərin idarə edilməsi texnologiyalarının müasir vəziyyəti araşdırılmış, e-dövlət mühitində ESİS-lərin ən aktual elmi-nəzəri problemləri müəyyən edilmişdir;
- E-dövlət mühitində ESİS-in arxitektura prinsipləri işlənmişdir;
- E-dövlət mühitində ESİS-in intellektuallaşdırılması üçün konseptual model təklif edilmişdir;
- E-dövlətdə mühitində sənədlərin təhlükəsizliyinə etimad modeli və onun fəaliyyətinə aid tövsiyələr təklif edilmişdir;
- Mətn sənədlərin avtomatik referatlaşdırılması üçün optimallaşma modelləri təklif edilmişdir;
- Optimallaşma məsələlərinin həlli üçün diferensial təkamül və “sürü” intellektinə əsaslanan optimallaşma alqoritmlərinin modifikasiya olunmuş binar variantları işlənmişdir;
- Dissertasiya işində təklif olunmuş model və alqoritmlər əsasında korporativ ESİS layihələndirilmiş, sistemin proqram təminatı işlənib hazırlanmış və AzScienceNet şəbəkəsində tətbiq edilmişdir.
- **Dissertasiya işinin əsas nəticələri aşağıdakı elmi işlərdə dərc edilmişdir:**

1. Фаталиев Т.Х., Гаджирагимова М.Ш. Электронный документо-оборот в корпоративной сети / **Материалы международной научно-технической конференции «Информационные системы и технологии»**, Новосибирск, Россия, 22-25 апрель, 2003, с.78.
2. Алгулиев Р.М., Фаталиев Т.Х., Гаджирагимова М.Ш. К созданию корпоративной распределенной архивной системы // **AMEA-nın xəbərləri, Fiz.-tex. elmləri seriyası**, 2003, №3, s.143-147.
3. Алгулиев Р.М., Фаталиев Т.Х., Гаджирагимова М.Ш. Об одной корпоративной электронной архивной системе / **Тезисы докладов 9-ой международной научной конференции «Теория и техника передачи, приема и обработки информации»**, Тьупсе, Украина, 7-10 октября, 2003, с.290-291.

4. Алгулиев Р.М., Фаталиев Т.Х., Гаджирагимова М.Ш. Интеграция систем документооборота и архива / **Тезисы докладов 10-ой международной научной конференции «Теория и техника передачи, приема и обработки информации»**, Туапсе, Украина, 28 сентября-1 октября, 2004, с.8-9.
5. Əliquliyev R.M., Fətəliyev T.X., Hacırahimova M.Ş. Korporativ şəbəkə mühitində elektron sənədlərin dövriyyəsinin mühafizəsi / **«İnformasiyalaşdırma, Kibernetika və İnformasiya Texnologiyalarının müasir problemləri» 2-ci respublika elmi konfransının əsərləri**, Bakı, Azərbaycan, 26-28 oktyabr, 2004, I cild, s.113-115.
6. Алгулиев Р.М., Фаталиев Т.Х., Гаджирагимова М.Ш., Интеграция корпоративных систем документооборота и архива // **AMEA-nın xəbərləri, Fiz.-tex. elmləri seriyası**, 2005, №2, s.12-17.
7. Алгулиев Р.М., Гаджирагимова М.Ш. Некоторые аспекты организации и реализации мультиагентной системы поиска информации в распределенной информационной среде / **Proceedings of the First International Conference «Problems of Cybernetics and Informatics»**, Baku, Azerbaijan, 24-26 October, 2006, vol.II, pp.31-34.
8. Həjirahimova M.S. On intellectualization of management of electronic documents / **Proceedings of the Third International Conference «Problems of Cybernetics and Informatics»**, Baku, Azerbaijan, 6-8 September, 2010, v.I, pp.167-170.
9. Hacırahimova M.Ş. Elektron dövlət mühitində sənəd dövriyyəsi sistemlərinin aktual problemləri və həlli yolları // **İnformasiya cəmiyyəti problemləri**, 2010, №2, s.21-29. <http://www.jpis.az/>
10. R.Alguliev, R.Aliguliyev, M.Hajirahimova Multi-document summarization model based on integer linear programming // **Intelligent Control and Automation**, 2010, v.1, №2, pp.105-111. <http://www.scirp.org/journal/ica/>
11. Гаджирагимова М.Ш. Архитектурные принципы построения системы электронного документооборота в рамках электронного государства // **Телекоммуникации**, 2011, №8, с.43-48. <http://www.nait.ru>
12. Alguliev R.M., Aliguliyev R.M., Hajirahimova M.S., Mehdiyev C.A. MCMR: maximum coverage and minimum redundant text summarization model // **Expert Systems with Application (Impact**

- Factor: 2.203**), 2011, v.38, №12, pp.14514-14522. <http://www.journals.elsevier.com/expert systems with applications/>
13. Имамвердиев Я.Н., Гаджирагимова М.Ш. Архитектура инфраструктуры доверия электронным документам в среде электронного государства // **Телекоммуникации**, 2011, №11, с.18-26. <http://www.nait.ru>
 14. Hacırahimova M.Ş. E-elm mühitində sənəd dövriyyəsi texnologiyaları / **«E-elm problemləri» üzrə I respublika elmi-praktiki konfransın materialları**, Bakı, Azərbaycan, 15-16 noyabr, 2012, s. 37-38.
 15. Alguliev R.M., Aliguliyev R.M., Hajirahimova M.S. GenDocSum + MCLR: Generic document summarization based on maximum coverage and less redundancy // **Expert Systems with Application (Impact Factor: 1.854)**, 2012, v.39, №16, pp.12460-12473. <http://www.journals.elsevier.com/expert systems with applications/>
 16. Hacırahimova M.Ş. Elektron sənəd dövriyyəsi sistemlərində standartlaşdırma məsələləri: beynəlxalq təcrübə // **İnformasiya cəmiyyəti problemləri**, 2012, №1, s.59-68. <http://www.jpis.az/>

Həmmüəlliflərlə dərc olunmuş işlərdə iddiaçının şəxsi rolu

- (2) – paylanmış arxiv sisteminin konsepsiyası işlənmişdir;
- (6) – korporativ ESIS və arxiv sisteminin integrasiyası işlənmişdir;
- (10), (12), (15) – mətn tipli e-sənədlərin avtomatik referatlaşdırılması üçün optimallaşma modelləri təklif edilmiş və optimallaşma məsələsinin həlli üçün təkamül alqoritmləri işlənmişdir;
- (11) – e-dövlət mühitində e-sənədlərə etimad modeli təklif edilmişdir.

МАКРУФА ШАРИФ КЫЗЫ ГАДЖИРАГИМОВА

РАЗРАБОТКА МОДЕЛЕЙ И АЛГОРИТМОВ ДЛЯ ИНТЕЛЛЕКТУАЛЬНОГО УПРАВЛЕНИЯ ТЕКСТОВЫХ ДОКУМЕНТОВ В СРЕДЕ ЭЛЕКТРОННОГО ГОСУДАРСТВА

АННОТАЦИЯ

Диссертационная работа посвящена разработке моделей и алгоритмов для синтеза систем электронного документооборота и интеллектуального анализа текстовых электронных документов в среде электронного государства. В рамках диссертационной работы получены следующие основные результаты, обладающие научной новизной:

- ✓ Разработаны архитектурные принципы систем электронного документооборота в среде электронного государства;
- ✓ Разработана концептуальная модель для интеллектуализации систем электронного документооборота;
- ✓ Предложена модель функционально-технологической архитектуры инфраструктуры доверия документам в э-государстве и даны рекомендации по ее реализации;
- ✓ Предложены оптимизационные модели для автоматического реферирования текстовых документов;
- ✓ Разработаны эволюционные алгоритмы (binary particle swarm optimization, binary differential evolution) для решения оптимизационных задач.

MAKRUFA SHARIF HAJIRAHIMOVA

**DEVELOPMENT OF MODELS AND ALGORITHMS FOR
INTELLIGENT MANAGEMENT OF TEXT DOCUMENTS IN
ELECTRONIC GOVERNMENT ENVIRONMENT**

SUMMARY

The thesis is devoted to the development of models and algorithms for synthesis of electronic document management systems and intellectual analysis of electronic text documents in the electronic government environment. According to the specified purpose in the thesis the following results are received:

- ✓ Architectural principles of electronic document management systems in the electronic government environment are developed;
- ✓ A conceptual model for the intellectualization of electronic document management systems is developed;
- ✓ A model for functional and technological architecture of trust infrastructure to the electronic documents in electronic government is suggested and recommendations for its implementation are given;
- ✓ Optimization models for automatic summarization of text documents are proposed;
- ✓ Evolutionary algorithms (binary particle swarm optimization, binary differential evolution) for solving the optimization problems are developed.

Çapa imzalanıb 25.10.2013. Tirajı 100 nüsxə.
AMEA İnformasiya Texnologiyaları İnstitutunun
“İnformasiya texnologiyaları” nəşriyyatı

**НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК АЗЕРБАЙДЖАНА
ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

На правах рукописи

МАКРУФА ШАРИФ КЫЗЫ ГАДЖИРАГИМОВА

**РАЗРАБОТКА МОДЕЛЕЙ И АЛГОРИТМОВ ДЛЯ
ИНТЕЛЛЕКТУАЛЬНОГО УПРАВЛЕНИЯ ТЕКСТОВЫХ
ДОКУМЕНТОВ В СРЕДЕ ЭЛЕКТРОННОГО ГОСУДАРСТВА**

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 – 2013