AZƏRBAYCAN RESPUBLİKASI

Əlyazması hüququnda

ELEKTRON DÖVLƏT-VƏTƏNDAŞ PLATFORMASINDA MƏLUMATLARIN İNTELLEKTUAL ANALİZİ ÜÇÜN METOD VƏ ALQORİTMLƏRİN İŞLƏNİLMƏSİ

İxtisas: 3338.01 – Sistemli analiz, idarəetmə və informasiyanın işlənməsi

Elm sahəsi: Texnika elmləri

Fəlsəfə doktoru

elmi dərəcəsi almaq üçün təqdim edilmiş

DİSSERTASİYA

İddiaçı: _____ **Günay Yavər qızı İskəndərli**

Elmi rəhbər: _____ AMEA-nın müxbir üzvü, texnika elmləri doktoru **Ramiz Məhəmməd oğlu Alıquliyev**

Bakı – 2021

MÜNDƏRİCAT

GİRİŞ

İşin aktuallığı. İnformasiya cəmiyyətinin əsas elementlərindən olan e-dövlətin yaradılmasında əsas məqsəd İnternet və İnformasiya-Kommunikasiya Texnologiyalarının (İKT) imkanlarından istifadə etməklə, dövlət idarələrinin vətəndaşlara göstərdikləri xidmətlərin səviyyəsini yüksəltmək, dövlət informasiya resurslarına girişi sadələşdirmək, cəmiyyətin bütün təbəqələrinin dövlət idarəçiliyində aktiv iştirakını, o cümlədən idarəetmədə şəffaflığı təmin etmək, nəticədə dövlət idarəçiliyində səmərəliliyi yüksəltməyə və xərcləri minimallaşdırmağa nail olmaqdır. Elektron dövlət (e-dövlət) mühitində səmərəli qərarların qəbulu və dövlət təhlükəsizliyinin təmin olunması üçün bu mühitin tədqiqi mühüm əhəmiyyət kəsb edir.

Qeyd edək ki, e-dövlət mürəkkəb sosiotexnoloji mühitdir və bu mühitdə əsas hədəf hakimiyyət-vətəndaş arasında olan münasibətlərdir. E-dövlət mühitinin müxtəlif seqmentləri üzrə problemlər mövcuddur. Amma bu dissertasiyada hədəfdə olan əsas məsələlər hakimiyyət-vətəndaş münasibətlərində mövcud olan informasiya fəzasının analizindən ibarətdir. E-dövlət mühitində insanların maraq dairəsində olan məsələlərin vaxtında müəyyən olunması dövlət orqanlarına xidmətlərin keyfiyyətini yaxşılaşdırmağa, vətəndaş məmnuniyyətini artırmağa kömək edə bilər. E-dövlət xidmətlərinin əlçatanlığını və effektivliyini artırmaq üçün mütəmadi olaraq istifadəçi yönümlü qiymətləndirmə aparmaq lazımdır.

E-dövlətin əsas funksiyalarından biri də vətəndaşları ehtimal olunan zərər və zorakılıqlardan qorumaqdır. Linders [106, s.453] vətəndaş-dövlət münasibətlərinin təkamülünü araşdıraraq, belə qənaətə gəlmişdir ki, ehtimal olunan cinayətlər haqqında əvvəlcədən məlumat vermək, o cümlədən cəmiyyət üzvləri ilə hüquq-mühafizə orqanları arasındakı münasibətlərin yaxşılaşdırılması baxımından İnternet, xüsusi halda e-dövlət ən effektiv və əlverişli vasitədir. Təcrübə göstərir ki, bu əlverişli mühitdən cinayətkar qruplar da yaxşı "yararlanırlar" və onlar bu imkandan istifadə edərək dövlət və cəmiyyət üçün böyük təhlükə mənbəyinə çevrilirlər. Deməli, dövlətin mühüm vəzifələrindən biri də virtual mühitdə – İnternetdə və e-dövlətdə gizli fəaliyyət göstərən kriminal şəbəkələrin fəaliyyətini aşkarlamaq və analiz

etməkdir. Bu mühit tez kommunikasiya yaratmaq və fəaliyyəti operativ koordinasiya etmək baxımından çox geniş imkanlara malikdir. Kriminal şəbəkənin üzvləri ünsiyyət qurmaq üçün veb-saytlardan, e-poçtdan, bloqlardan, onlayn çatdan və s. istifadə edir. Aydındır ki, belə kommunikasiya vasitələrində ötürülən informasiya növləri arasında mətnlər üstünlük təşkil edirlər. Ona görə də, mümkün ola biləcək terror aktlarının qarşısının alınması və dövlətin təhlükəsizliyinin təmin olunması üçün virtual mühitdə, o cümlədən e-dövlətdə dövr edən mətnlərin analizi mühüm əhəmiyyət kəsb edir.

Yuxarıda qeyd olunduğu kimi təhlükəsizlik çox mühüm məsələdir və təhlükəsizliyin təmin olunması üçün müxtəlif yanaşmalar, baxışlar mövcuddur. Təəssüflər olsun ki, e-dövlət mühitində vətəndaşların özünəmənsub olan şərhləri yetərincə analiz olunmayıb. Hal-hazırda biliklərin idarə olunmasında, müxtəlif mənbələrdə toplanmış mətnlərin intellektual analizində text mining ən qabaqcıl və effektiv texnologiyalardan biri hesab olunur [9, s.38]. Dissertasiya işində bunları nəzərə alaraq, sosial şəbəkə və mətnlərin intellektual analizi (text mining) texnologiyalarının köməyilə e-dövlət-vətəndaş münasibətləri araşdırılmış, bu mühitdə səmərəli qərarların qəbulunu dəstəkləyən sistemlər üçün yeni yanaşmalar, metod və alqoritmlər təklif edilmişdir.

Tədqiqat obyekti e-dövlət-vətəndaş münasibətləri, **tədqiqatın predmeti** e-dövlət-vətəndaş platformasında məlumatların intellektual analizi üçün metod və alqoritmlərdir.

İşin məqsədi. Dissertasiya işinin məqsədi e-dövlətin təhlükəsizlik səviyyəsinin yüksəldilməsi və bu mühitdə göstərilən xidmətlərin keyfiyyətinin yüksəldilməsi məqsədilə sosial şəbəkə və mətnlərin intellektual analizi texnologiyalarının köməyilə dövlət-vətəndaş münasibətlərinin analizi üçün yeni yanaşmalar, metod və alqoritmlərin təklif edilməsidir.

Bu məqsədlə dissertasiya işində aşağıdakı **məsələlər** qoyulmuşdur:

- e-dövlətin analizi texnologiyalarının müasir vəziyyətinin analizi və problemlərin müəyyən olunması;

- e-dövlətdə terrorizmi təbliğ edən mətnlərin təsnifatı və filtrasiyası üçün metodların işlənməsi;
- e-dövlətdə fəaliyyət göstərən gizli sosial şəbəkələrin aşkarlanması və analizi üçün metodun işlənməsi;
- e-dövlət xidmətlərindən vətəndaş məmnuniyyətinin avtomatik qiymətləndirilməsi üçün metodun işlənməsi;
- e-dövlətdə vətəndaş şərhlərinin analizi əsasında vətəndaşları maraqlandıran aktual mövzuların müəyyən edilməsi üçün metodun işlənməsi.

Tədqiqat metodları. Dissertasiyada qarşıya qoyulmuş məsələləri həll etmək üçün təbii dilin emalı, data mining, text mining, mövzu modelləşdirmə (topic modeling), qraflar nəzəriyyəsi, ehtimal nəzəriyyəsi, sosial şəbəkə analizi texnologiyaları metodlarından istifadə edilmişdir.

Müdafiəyə çıxarılan əsas müddəalar:

- e-dövlətdə terrorizmlə əlaqəli mətnlətin aşkarlanması üçün hibrid təsnifatlandırma metodu;
- e-dövlətdə terrorizmi təbliğ edən mətnlərin filtrasiyası üçün sentiment analiz texnologiyasına və Bayes klassifikatoruna əsaslanan metod;
- e-dövlətdə fəaliyyət göstərən gizli sosial şəbəkələrin aşkarlanması və analizi üçün sentiment analiz texnologiyasına əsaslanan metod;
- e-dövlət xidmətlərindən vətəndaş məmnuniyyətinin avtomatik qiymətləndirilməsi üçün metod;
- e-dövlətdə vətəndaşları (o cümlədən regionları) maraqlandıran aktual mövzuların müəyyən edilməsi üçün metod.

Elmi yeniliklər. Dissertasiya işi çərçivəsində elmi yeniliyə malik aşağıdakı əsas nəticələr əldə edilmişdir:

- e-dövlətdə terrorizmlə əlaqəli mətnlərin aşkarlanması üçün hibrid təsnifatlandırma metodu işlənmişdir;
- e-dövlətdə terrorizmlə əlaqəli mətnlərin filtrasiyası üçün sentiment analiz texnologiyasına və Bayes klassifikatoruna əsaslanan metod işlənmişdir;

- e-dövlətdə gizli sosial şəbəkələrin aşkarlanması və analizi üçün sentiment analiz texnologiyasına əsaslanan metod işlənmişdir;
- e-dövlət xidmətlərindən vətəndaş məmnuniyyətinin avtomatik qiymətləndirilməsi üçün metod işlənmişdir;
- e-dövlətdə vətəndaşları (o cümlədən regionları) maraqlandıran aktual mövzuların müəyyən edilməsi üçün klasterləşmə və mövzu modelləşdirmə texnologiyalarına əsaslanan metod işlənmişdir.

İşin praktiki əhəmiyyəti. Əldə edilmiş elmi-nəzəri və praktiki nəticələr onlayn mühitlərdə fəaliyyət göstərən müxtəlif təbiətli sosial şəbəkələrin aşkarlanması və analizində, terrorizmlə əlaqəli fəaliyyətin aşkarlanması, terrorizmlə bağlı mətnlərin filtrasiyasında, e-xidmətlərin keyfiyyətinin yüksəldilməsində, vətəndaşların, regionların əsas maraq dairələrinin müəyyən olunmasında, istifadəçi şərhlərinin analizi əsasında aktual mövzuların çıxarılmasında və s. istifadə oluna bilər.

İşin aprobasiyası. Əsas elmi-nəzəri və praktiki nəticələr aşağıda adı çəkilən konfranslarda məruzə olunmuşdur:

- "İnformasiya təhlükəsizliyinin multidissiplinar problemləri" II respublika elmi-praktiki konfransı, Bakı, 14 may 2015-ci il;
- "Big data: imkanları, multidissiplinar problemləri və perspektivləri" I respublika elmi-praktiki konfransı, Bakı, 25 fevral 2016-cı il;
- 10th International Conference on Application of Information and Communication Technologies – AICT-2016, Baku, 12-14 October 2016;
- "İnformasiya təhlükəsizliyinin aktual multidissiplinar problemləri" IV respublika elmi-praktiki konfransı, Bakı, 14 dekabr 2018-ci il;
- 2nd International Symposium on Applied Sciences and Engineering, Turkey, 7-9 April 2021.

Elmi nəşrlər. Dissertasiya mövzusu üzrə 14 elmi iş çap olunmuşdur. Onlardan 6 məqalə resenziya olunan jurnallarda, 7 tezis konfrans materiallarında və 1 ekspress-informasiya nəşr edilmişdir. Bu elmi işlərdən 3 məqalə Web of Science bazasında indeksləşən jurnallarda çap edilmişdir.

Dissertasiya işinin yerinə yetirildiyi təşkilatın adı: Azərbaycan Milli Elmlər Akademiyası İnformasiya Texnologiyaları İnstitutu.

İşin strukturu və həcmi. Dissertasiya işi giriş, 4 fəsil, nəticə, 179 adda ədəbiyyat siyahısı və bir əlavədən ibarətdir. İşin ümumi həcmi – 135, əsas mətn – 111 səhifədə şərh edilmişdir.

Dissertasiyanın işarə ilə ümumi həcmi: 203050 işarə.

Dissertasiyanın struktur bölmələrinin ayrılıqda həcmi:

Titul səhifəsi – 436 işarə.

Mündəricat – 1820 işarə.

Giriş – 7804 işarə.

Dissertasiyanın əsas məzmunu (fəsil, paraqraf, bənd) – 150824 işarə.

Nəticə – 983 işarə.

İstifadə edilmiş ədəbiyyat siyahısı – 28972 işarə.

Əlavələr – 5420 işarə.

İddiaçı AMEA-nın həqiqi üzvü, texnika elmləri doktoru, professor Rasim Əliquliyevə və elmi rəhbər, AMEA-nın müxbir üzvü, texnika elmləri doktoru Ramiz Alıquliyevə qiymətli məsləhətlərinə, dissertasiya işinin yerinə yetirilməsində göstərdikləri daimi diqqətə və hərtərəfli dəstəyə görə dərin minnətdarlığını bildirir.

I FƏSİL. E-DÖVLƏTİN ANALİZİ TEXNOLOGİYALARI: TEXT MINING VƏ SOSİAL ŞƏBƏKƏLƏR

Yaşadığımız əsr informasiya və yüksək texnologiyalar əsridir. İnkişaf etmiş texnologiya məhsulları həyatımızın ayrılmaz hissəsinə çevrilmişdir. İnformasiya texnologiyalarının (İT) inkişafı günümüzdə müxtəlif təşkilatların, şirkətlərin, dövlət qurumlarının fəaliyyətinə, iş prinsipinə, səmərəliliyinə təsir edən əsas amillərdən biridir. İT-in cəmiyyətin idarə olunması proseslərinə cəlb olunması e-dövlət anlayışının yaranması ilə nəticələnmişdir. E-dövlət – informasiya və kommunikasiya texnologiyalarının geniş tətbiqi ilə dövlət orqanlarının fəaliyyətinin səmərəliliyinin - operativliyinin yüksəldilməsini və insanların e-xidmətlərə çıxışının asanlaşdırılmasını nəzərdə tutur. E-dövlətin yaradılmasında əsas məqsəd vətəndaşlara göstərilən xidmətlərin sadələşdirilməsinə, biznes, sənaye müəssisələri ilə əlaqələrin təkmilləşdirilməsinə, daha səmərəli dövlət idarəetməsinə, idarəetmə xərclərinin minimallaşdırılmasına və şəffaflığın artırılmasına nail olmaqdır [40,124].

E-dövlət anlayışı ilk dəfə 1990-cı illərin ortalarında dövlət idarəçiliyində aparılan islahatlara İT-in tətbiqi nəticəsində yaranmışdır. Ümumilikdə qeyd etmək lazımdır ki, e-dövlət anlayışı bu dövrdə ortaya çıxsa da, kompüterin idarəetmədə tətbiqi ilk kompüterin tarixinə qədər gedib çıxır [135]. Əvvəllər İT-nin idarəetmədə tətbiqi dedikdə dövlət strukturlarının kompüterləşməsi nəzərdə tutulurdusa, hazırda istifadə edilən e-dövlət anlayışı bilavasitə dövlətin əhaliyə göstərdiyi e-xidmətləri nəzərdə tutur. Əhalinin kommunal ödəmələrini, sənədləşmə prosedurlarını, qeydiyyatlarını, seçkilərdə iştirakını, ərizə və müraciətlərini elektron platforma üzərindən həyata keçirmələrini e-xidmətlərə nümunə göstərmək olar.

E-dövlət inkişaf mərhələlərinin modellərinin qurulması yolu ilə öyrənilir. E-dövlətin inkişafını qiymətləndirmək üçün tədqiqatçılar tərəfindən müxtəlif modellər təklif olunmuşdur. E-dövlətin təkmil modelləri e-dövlət portalının kamilliyini müəyyən edən mərhələlər çoxluğundan (sadədən mürəkkəbə doğru) ibarətdir. Bu modellərin yaradılmasında əsas məqsəd e-dövlət portalının keyfiyyətinin artırılmasından ibarətdir. Cədvəl 1.1.1-də e-dövlətin təkmil modelləri xronoloji ardıcıllıqla verilmişdir:

Cədvəl 1.1.1

E-dövlətin təkmil modelləri

Model	Mərhələ 1	Mərhələ 2	Mərhələ 3	Mərhələ 4	Mərhələ 5	Mərhələ 6
Deloitte və Touche [54], 2000	İnformasiyanın nəşr olunması	Rəsmi ikitərəfli əlaqə	Çoxməqsədli portallar	Portal fərdiləşdirmə	Ümumi xidmətlərin klasterləşdirilməsi	Tam inteqrasiya və korporativ sazişlər
Layne və Lee [101], 2001	Kataloqlaşdırma	Tranzaksiya	Şaquli inteqrasiya	Üfüqi inteqrasiya	—	—
Hiller və Belanger [69], 2001	İnformasiya	İkitərəfli əlaqə	Tranzaksiya	İnteqrasiya	İştirak	—
Moon [117], 2002	Sadə informasiyanın yayılması	İkitərəfli əlaqə	Xidmətlər və maliyyə əməliyyatları	İnteqrasiya	Siyasi həyatda iştirak	—
Windley [170], 2002	Sadə veb sayt	Onlayn hökümət	İnteqrasiya edilmiş hökumət	Transformasiya edimiş hökümət	—	—
Siau və Long [140], 2005	Onlayn mövcudluq	Qarşılıqlı əlaqə	Tranzaksiya	Transformasiya	E-demokratiya	—
Shahkooh [136], 2008	Onlayn Mövcudluq	Qarşılıqlı əlaqə	Qarşılıqlı əlaqə	Tam inteqrallaşmış e-dövlət	E-demokratiya	—
Almazan və GilGarcia [26], 2008	Mövcudluq	İnformasiya	Qarşılıqlı təsir	Tranzaksiya	Interqasiya	Siyasi həyatda iştirak
Kim və Grant [91], 2010	Onlayn mövcudluq	Qarşılıqlı əlaqə	Tranzaksiya	İnteqrasiya	Davamlı inkişaf	—

Cədvəl 1.1.1-in ardı

Alhomod və s. [20], 2012	Onlayn mövcudluq	Vətəndaş və hökümət arasında qarşılıqlı əlaqə	İnternet üzərində tam tranzaksiya	Xidmətlə-rin inteqrasi-yası	—	—
United Nations, [160], 2012	İnkişaf etməkdə olan informasiya xidmətləri	Genişləndiril-miş informasiya xidmətləri	Tranzaksiya xidmətləri	Əlaqəli xidmətlər	—	—
Lee və Kwak [102], 2012	İlkin şərtlər (şərait)	Məlumatların şəffaflığı	Açıq iştirak	Açıq əməkdaş-lıq	Hər yerdə mövcud olan iştirak	—

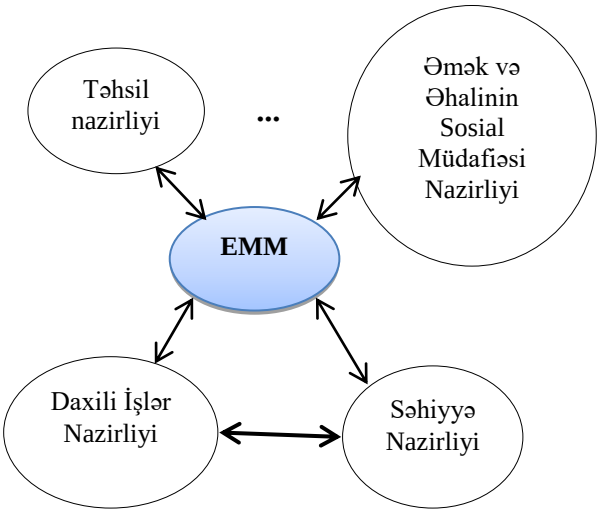
Bu modellərin təkmilləşdirilməsi üçün e-dövlət portalı yaxşı analiz olunmalı, vətəndaşların fikri öyrənilməlidir. Məlum olduğu kimi, e-dövlətin konseptual modeli tam orqrafdır. E-dövlətin ilkin konseptual modelində üç qovşaq var: dövlət-dövlət (G2G), dövlət-biznes (G2B) və dövlət-vətəndaş (G2C). Onların arasında tam qraf yaradılır: vətəndaş-vətəndaş (C-C), biznes-biznes (B-B), dövlət-dövlət (G-G). Bu şəbəkələri analiz etməklə, dövlət xidmətlərinin keyfiyyətini yüksəltmək, onun təhlükəsizliyini təmin etmək olar. Həmçinin vətəndaşların arzu və təkliflərini analiz etməklə, qərar qəbuletmə prosesində vətəndaşların rolunu təmin etmək olar. Elektron mühitdə vətəndaşlar öz fikirlərini mətn şəklində ifadə etdiyindən, bu fikirlərin analizində hal-hazırda ən qabaqcıl texnologiyalardan biri olan text mining-dən istifadə oluna bilər. Baxılan dissertasiya işində dövlət-vətəndaş münasibətlərində sosial şəbəkələr vasitəsilə internet üzərindən müxtəlif hakimiyyət orqanlarının saytları üzərində olan forumlarda, müxtəlif mənbələrdə, sosial şəbəkələrdə müxtəlif servislər vasitəsilə gələn rəylər toplanılır, süni intellekt texnologiyaları vasitəsilə analiz olunur.

Göründüyü kimi text mining və sosial şəbəkə analizi texnologiyaları vasitəsilə e-dövlət sistemini daha da təkmilləşdirmək mümkündür. Problemin aktuallığını nəzərə alaraq, fəsil 1-də e-dövlət sahəsində bu texnologiyaların rolu və tətbiqi araşdırılmışdır.

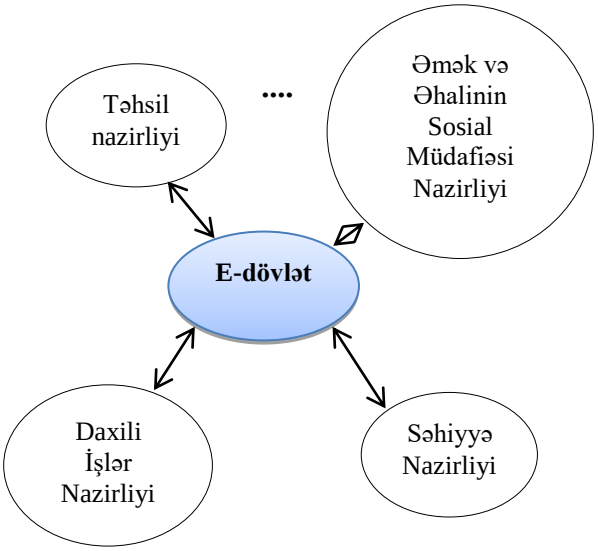
1.1. E-dövlətin analizi sahəsində problemlərin müasir vəziyyəti

E-dövlətin əsas məqsədi vətəndaşlara göstərilən xidmətlərin keyfiyyətinin artırılmasından ibarətdir. Buna görə də, e-dövlətə qoyulan tələblər əsasən vətəndaşların ehtiyaclarından irəli gəlməlidir. Ümumi e-dövlət tələbləri dövlət informasiya və biliklər sisteminin mübadiləsini və keyfiyyətini əhatə edir. Lakin səriştəli İT kadrların və dövlət qurumlarının sayını nəzərə alsaq, insan resursları çatışmazlığının hələ də mövcud olduğunu görə bilərik. Autsorsinq konsepsiyası bu cür problemlərin azaldılmasında tətbiq edilir.

Qeyd edək ki, inkişaf etmiş ölkələr ilə inkişaf etməkdə olan ölkələrdə e-dövlətin inkişafına yanaşmalar fərqli ola bilər. İnkişaf etmiş ölkələrdə hər bir dövlət orqanı özünün İT layihələrini və kompüter data mərkəzini tamamlamışdır. Belə ki, bu ölkələrdə e-dövlətin hissələri artıq bir çox dövlət müəssisələrində mövcuddur. E-dövlətin hər bir hissəsinə nəzər saldıqda, informasiya və bilik sistemlərinin səmərəli fəaliyyət göstərdiyini görmək olur. Burada əsas problem bu sistemlərin birləşdirilməsidir. E-dövlət yanaşmasını nəzərə alsaq, bu bütün mövcud informasiya və bilik sistemlərinin bir-birilə birləşdirilməsi yolu ilə əldə oluna bilər. Bu yanaşma paylanmış model kimi tanınır. Şəkil 1.1.1-də müxtəlif dövlət təşkilatları arasında müxtəlif məlumat standartlarının inteqrasiyasını həyata keçirən E-Məlumatların Mübadiləsi (EMM) modeli təsvir olunmuşdur [49, s.119]:



Şəkil.1.1.1 Paylanmış model



Şəkil.1.1.2 Mərkəzləşdirilmiş model

Yuxarıda qeyd olunduğu kimi, inkişaf etməkdə olan ölkələrdə e-dövlət inkişaf etmiş ölkələrdən fərqlidir, çünki informasiya və bilik sistemləri, informasiya texnologiyaları və infrastrukturu bu ölkələrdə yaxşı inkişaf etməmişdir. Ona görə də kompüter və informasiya texnologiyalarının ümumi standartları e-dövlət layihəsinin başlanğıc mərhələsində həll olunmalıdır. Dövlətin əsas vəzifəsi bütün dövlət orqanlarının e-dövlət konsepsiyasını, missiyasını, məqsədini, vəzifələrini, strategiya və taktikasını müəyyən etmək, hər bir orqanın eyni e-dövlət mənzərəsinə malik olduğuna əmin olmaqdır [25]. Mərkəzi planlaşdırma nəticəsində e-dövlət layihələrinin həyata keçirilməsi zamanı aydın mənzərə yaratmaq mümkündür. Bunun nəticəsində hər bir dövlət orqanı öz resurslarını "bir pəncərə" prinsipi əsasında mübadilə edə bilər. Bu kifayət qədər resursu olmayan ölkələr üçün nəzərə tutulub və mərkəzləşdirilimiş model adlanır (Şəkil 1.1.2). Bu model bir çox inkişaf etməkdə olan ölkələrdə daha münasibdir. Hazırda Azərbaycanda paylanmış modelin tətbiq olunduğu sahələr var və gələcəkdə bütövlükdə paylanmış modelə keçid nəzərdə tutulur.

Dövlət fəaliyyətinin səmərəliliyinin artırılmasına yönəlmiş e-dövlət təşəbbüsləri vətəndaşlara və bzines sektoruna göstərilən dövlət xidmətlərinin təkmilləşdirilməsi, dövlətin daxildən iş səmərəliliyinin artırılması, bununla yanaşı vətəndaşların qərar qəbuletmə prosesində iştirakının genişləndirilməsi məqsədi daşıyır. Lakin bir çox araşdırmalar göstərir ki, bütün dünyada e-dövlət layihələrinin həyata keçirilməsi üçün təşəbbüslərin böyük bir hissəsi qoyulmuş məqsədlərə hələ də tam çatmayıb. E-dövlət və ümumilikdə e-dünya üçün bir neçə əsas aspekt onların inkişafında üzləşdiyi əsas problemləri əks etdirir [57, s.11]:

1. Siyasi aspektlər
   - ✓ Məhdud büdcənin ayrılması;
   - ✓ Elektron Tranzaksiyalar Qanununun olmaması;
   - ✓ Qərar qəbuletmə prosesinin ləngliyi;
   - ✓ İslahat və inteqrasiya;
   - ✓ Bürokratiya.

2. Sosial aspektlər
   - ✓ Təhsil və mədəniyyətin olmaması;
   - ✓ Özəl sektorun zəifliyi;
   - ✓ Məlumatların məxfiliyi və gizliliyi;
   - ✓ Özünə xidmət (Self service) modellərinin əldə olunmasına vətəndaşların kifayət qədər cəlb olunmaması.
3. İqtisadi aspektlər
   - ✓ Büdcə çatışmazlığı;
   - ✓ İnfrastruktur investisiyalarının azlığı;
   - ✓ Korrupsiya.
4. Texniki aspektlər
   - ✓ Texniki sahədə ixtisaslaşmış kadrların azlığı;
   - ✓ Məlumatların beynəlxalq şəbəkəsindən istifadə xərclərinin yüksək olması;
   - ✓ İnformasiyadan istifadə və kommunikasiya standartlarının olmaması;
   - ✓ Proqram təminatı və texnikaya inam.

E-dövlət sahəsində mövcud problemlərdən bir neçəsi aşağıda ətraflı şəkildə nəzərdən keçirilmişdir:

Elektron identifikasiya, təhlükəsizlik və inam: E-dövlətdə əsas problemlərdən biri vətəndaşların e-dövlət xidmətlərinə inamının olmamasıdır. E-dövlət layihələrinin müvəffəqiyyətli olması üçün qurumlar, biznes və vətəndaşlar arasında inam yaratmaq lazımdır. Onlayn mühitdə inamın yaradılması isə çətin məsələlərdən biridir. Çünki bu mühitdə üz-üzə münasibətə nisbətən müxtəlif həmsöhbətlər arasında xeyirxahlıq və dürüstlüyün qiymətləndirilməsi daha mürəkkəbdir. E-dövlət xidmətlərinin dayanıqlığı təkcə xidmətlərin göstərilməsində kritik əhəmiyyət daşımır, həmçinin vətəndaşların inam və etibarının qurulmasında vacibdir. Boston konsaltinq qrupu tərəfindən aparılan araşdırmada müəyyən olunmuşdur ki, e-dövlət xidməti istifadəçilərinin 47% -i şəxsi məlumatları üçün daha çox zəmanətin olmasını istəyir. Burada qeyd olunur ki, Danimarka, Finlandiya, İsveç və Hollandiya e-dövlət orqanlarına inamın yüksək olduğu dövlətlər sırasındadır [53].

İnam mühitdən asılı olaraq tamlığı dəyişən elektron identifikasiya məlumatından asılıdır. Avropa İttifaqı tərəfindən bu məsələ ilə bağlı bir sıra səylər göstərilmişdir. 2011-ci ildə Polşanın Poznan Şəhərində Nazirlər səviyyəsində altıncı E-Hökumət Konfransında vətəndaşlar və hüquqi şəxslərə bir üzv dövlət tərəfindən verilmiş elektron şəxsiyyətin digər üzv dövlətlərdə qarşılıqlı tanınma prosesi vasitəsilə bütün dövlət və özəl əməliyyatlarda istifadə olunması təklif olunmuşdur. 2014-ci ildə Avropa Parlamenti və Şurası sərhədyanı xidmətlərdə qarşılıqlı fəaliyyəti asanlaşdırmaq üçün elektron identifikasiya qarşılıqlı tanınmanı nəzərdə tutan elektron identifikasiya və etimad xidmətləri üzrə 2014/910 Qərarı qəbul etmişdir.

E-dövlətin həyata keçirilməsində hər bir dövlətin üzləşdiyi əsas problemlər onun planlaşdırılması və idarə olunmasıdır. Maraqlı tərəflərin tələblərini və istəklərini əks etdirən sistemin inkişafına ehtiyac var. Müxtəlif tədqiqatçılar tərəfindən e-dövlətin "təklif tərəfi" ilə "tələb tərəfi" arasında uyğunsuzluq (fərq) vurğulanmışdır. Bu o deməkdir ki, e-dövlət xidmətlərinin mövcudluğu ilə onların faktiki mənimsəmə qabiliyyəti və istifadəsi arasında boşluq mövcuddur. Məsələn, Avstriyada onlayn ümumi dövlət xidmətlərinin demək olar ki, 100% -i mövcuddur, lakin fiziki şəxslər tərəfindən faktiki istifadə yalnız 50%-dən bir qədər çoxdur [31].

Texnoloji problemlər haqqında danışarkən, e-dövlət təşəbbüslərinin həyata keçirilməsi üçün ən əhəmiyyətli problemlərdən biri təhlükəsizlikdir. Təhlükəsizlik dizayn mərhələsində həll olunmalıdır. Təhlükəsizliyin pozulması e-dövlətə olan ictimai etimadı sarsıda bilər. Dövlət əldə etdiyi şəxsi məlumatların məxfililiyin qorunmasına məsuliyyət daşıyır [134]. Hökumətdə gündəlik əməliyyatlar vasitəsilə böyük miqdarda vətəndaşlara dair məlumatlar toplanır. Verilənlər bazasında saxlanılan fərdi məlumatlardan istifadə etdikdə onların məxfiliyinin qorunması vacib məsələlərdəndir. E-dövlət mühitində müxtəlif məzmunda (dövlət əleyhinə təbliğat, mentalitetə uyğun gəlməyən, milli mənəvi dəyərlərin əsaslarını sarsıdan, terrorizmi təbliğ edən informasiyanın yayılması və s.) informasiyanın vaxtında aşkarlanması dövlətin və cəmiyyətin təhlükəsizliyinin təmin olunması baxımından mühüm əhəmiyyət kəsb edir və günümüzün ən aktual elmi-nəzəri və praktiki problemlərindən biridir [23, 7, 177]. Heç də təsadüfi deyildir ki, e-dövlətin təhlükəsizliyi problemi

Avropa Komissiyası tərəfindən qəbul edilmiş eGovRTD2020 layihəsində e-dövlət sahəsində araşdırılması vacib olan 13 ən aktual elmi-tədqiqat istiqamətindən biri kimi qeyd olunmuşdur [168].

E-iştirak: E-iştirak vətəndaşların siyasi iştirakını asanlaşdırmaq məqsədilə bir-biriylə, vətəndaş cəmiyyəti, onların seçilmiş nümayəndələri və hökumət ilə əlaqə qurmağa kömək edən İT-dən istifadə etmək imkanıdır. Əvvəllər sadəcə sorğular və ərizələr vasitəsilə vətəndaşlarla məsləhələşmələr aparılırdısa, artıq e-iştirak vasitəsilə vətəndaşlar fəal siyasi prosesə cəlb olunaraq məsələlər qaldıra, gündəliyin dəyişdirilməsini və hökumət təşəbbüslərini dəyişə bilərlər. Vətəndaşların onlayn iştirakının artırılması siyasi qərarların keyfiyyətinin yaxşılaşdırılması və qəbul olunan qərarların legitimliyini artırmaq potensialına malikdir. Bəzi tədqiqatçılar e-dövlətin bu aspektini e-demokratiyanın yeni erası, şəffaflıq və hesabatlılığın artması kimi qiymətləndirirlər. Müxtəlif texnologiyalar, o cümlədən veb axını, sosial media (xüsusilə Facebook və Twitter), bloq, müzakirə forumları, qərar dəstək sistemləri və elektron səsvermə sistemləri e-iştiraka kömək edə bilər. Xüsusilə sosial media interaktiv, ikitərəfli rabitə və çox güclü şəbəkə effektləri ilə faydalı ola bilər. Lakin Avropanın e-iştirak layihələri üzrə 2010-cu il sorğularının nəticələri e-iştirak tətbiqlərinə nisbətən ümumi məqsədli İT alətlərinin daha çox istifadə olunduğunu aşkar etmişdir. Sorğu həmçinin e-iştirak layihələri üçün güclü dövlət dəstəyi, istifadəçi interfeysi, müxtəlif rabitə kanallarının (oflayn, həmçinin onlayn) istifadəsi, müvafiq təhlükəsizlik və gizlilik müddəaları (iştirakçıları tam olaraq müəyyən etmək üçün anonim cavabların müəyyənləşdirilməsi) və qeyri-mütəxəsislər tərəfindən siyasi məsələlərin anlaşılan şəkildə təqdim olunması daxil olmaqla, bir sıra uğurlu faktorları müəyyən etmişdir [53].

E-iştirak vasitəsilə vətəndaşların dövlətin idarə olunmasında rolunu artırmaq olar. Onların müxtəlif forum, e-dövlət portalı, bloqlarda yazdığı şərhləri analiz etməklə bir sıra dövlət xidmətlərini təkmilləşdirmək, onların e-dövlətdən məmnuniyyətini artırmaq olar [28]. Elektron bölgü (gəlir, elektron bacarıqlar və ya yaşayış yeri baxımından) göstərir ki, bəzi vətəndaşların digərlərinə (məsələn. gənc "elektron nəslə") nisbətən bu prosesdə iştirak etmək üçün məhdud imkanları var.

Belə ki, internetə çıxışı olmayan insanlar onlayn xidmətlərdən yararlana bilmir. Bəzi vətəndaşların onlayn şəkildə özünü identifikasiya etmək istəməməsi onların sərbəst və açıq şəkildə iştirakına mane olur [39]. Bu problemləri nəzərə alaraq e-iştirak hal-hazırda əsas tədqiqat və eksperiment sahələrindən biri olaraq qəbul olunur.

Big data: Big data rəqəmsal dövrdə dövlətin qarşısında duran yeni və mühüm məsələlərdən biridir. Biznes sektoru big data tətbiqlərinin inkişafında öndə olsa da, dövlət sektorunda da buna çox böyük ehtiyac yaranmışdır [92].

Bir çox dövlət qurumunda müxtəlif mənbələrdən məlumatlar toplanır. Bu məlumat axını e-dövlətdə big data mənbələrinin yaranmasına gətirib çıxarır. Bu məlumatlar yalnız ölçülərinə görə böyük deyil, həm də dinamik və qeyri-bircins xarakterlidir. Sürətlə böyüməkdə olan bu informasiyadan real vaxtda qərarların qəbulunun dəstəklənməsi, dövlətin vətəndaşlara xidmət göstərməsi, səhiyyə xərcləri, işsizlik, təbii fəlakətlər və terrorizmin artması kimi milli problemlərin aradan qaldırılmasında istifadə oluna bilər [4, 47]. Bu məlumatların həcminin çox böyük sürətlə artması müvafiq idarəetmə strukturunun həyata keçirilməsində bir sıra problemlər yaradır. Bu problemlərə elektron mühitdə eksponensial sürətdə yaradılan informasiya axınını analiz etmək üçün səmərəli üsulların inkişaf etdirilməsi tələbatı, e-məlumatların saxlanılması, big datanın idarə olunması, müxtəlif məlumat mənbələrinin inteqrasiyası, elektron gizlilik və təhlükəsizlik risklərinin idarə olunması kimi problemlər daxildir. Verilənlər eksponensial sürətlə artdığından ənənəvi məlumat (data) emalı vasitələri və texnologiyaları bu məlumatların analizində faydalı ola bilmir. Big data texnologiyaları bu problemlərin aradan qaldırılmasında əhəmiyyətli rol oynaya bilər. Big data analitikləri bunun text və data mining texnologiyaları vasitəsilə mümkün olduğunu qeyd edirlər.

Dövlət sektorunda informasiya və açıq məlumatlar: Açıq məlumat – sərbəst şəkildə heç bir məhdudiyyət qoyulmadan istənilən şəxs tərəfindən istənilən məqsədlə istifadə edilən, şəkli dəyişdirilərək başqa formaya salınan (sxem, xəritə, diaqram və s.) və paylaşılan məlumatlardır. Açıq hökumət məlumatları siyasətçilərə mürəkkəb problemləri həll etməyə, dövlət xidmətlərinin səmərəliliyinin artırılmasına, yeni, ən son tətbiqlər ilə iqtisadi artım yaratmağa, vətəndaşları siyasi inkişafa cəlb etməyə

kömək edə bilər. “Açıq Elektron Hökumət” yalnız metodoloji nöqteyi-nəzərdən deyil, eyni zamanda istifadəçilərin tələblərinə uyğun olaraq, dəqiq strukturlar əsasında təqdim edilir. Bu struktur özündə açıq məlumatlar, paylaşma, ünsiyyət və əməkdaşlıq prinsiplərini əks etdirir [81, 179].

Açıq Hökumət Məlumatları ilə bağlı göstərilən bir sıra təşəbbüslərə baxmayaraq, hələ də kifayət qədər problemlər mövcuddur. Bu problemlər əsasən texniki xarakterlidir. Açıq Hökumət Məlumatları ilə bağlı ən çox göstərilən problemlər içərisində metaməlumatların mövcud olmaması, məlumatlar və onlardan necə istifadə olunması haqqında kifayət qədər biliyin olmaması, terminologiyaların müxtəlifliyi, İT cihazlar tərəfindən oxuna bilən formatda toplanılmamasıdır.

Məlumatların qeyri-bircins təbiətli olması əsas problemlərdən biridir. Açıq hökumət portalı istifadəçilərinin azlığı göstərir ki, açıq məlumatlardan istifadə etməyə təsir edən faktorların müəyyən olunmasına ehtiyac var. Əgər proqnozlaşdırılan istifadəçilər bu məlumatlardan istifadə etmirlərsə, onda açıq hökumət təşəbbüsləri əbəsdir. Portalın müvəffəqiyyətli olması üçün istifadəçilər dərc olunmuş məlumatların aktuallığı və faydalılığı haqqında məlumatlandırılmalıdır [119]. [173]-də açıq məlumat istehlakında maraqlı tərəflərin iştirakına təsir edən amillərin müəyyənləşdirilməsi istiqamətində tədqiqat işi aparılmışdır. [42]-də açıq məlumat təşəbbüslərinin istənilən iştirak səviyyəsinə çatması üçün strategiyalar təklif olunmuşdur. [141]-də isə maraqlı tərəflərin iştirakının artırılması məqsədilə hansı dövlət xidmətlərinin təmin olunması istiqamətində tədqiqat işi aparılmışdır. Açıq məlumat tərəfdarları vətəndaşlar, biznes və vətəndaş cəmiyyəti tərəfindən hazırlanmış təkmilləşdirilmiş xidmətlər vasitəsilə xərclərə qənaət olunacağını iddia edirlər. Buna baxmayaraq informasiya keyfiyyətinin yaxşılaşdırılması, məlumatların interpretasiyası, yenidən istifadəsi ilə bağlı məsələlər əlavə xərclər yarada bilər [33,81].

Azərbaycanda bu sahə yeni inkişaf edir və 2015-ci il yanvar ayında istifadəyə verilmiş Açıq Hökumət Məlumatları portalında hazırda bir sıra dövlət qurumları tərəfindən 650-dən çox məlumat bazaları açıq elan edilmişdir. Bu məlumatlar əsasında artıq bir sıra veb və mobil proqramlar hazırlanmışdır. “Elektron hökumət”

portalında açıq məlumatların sayının və çeşidinin genişləndirilməsi istiqamətində mütəmadi olaraq işlər aparılır. Heç də təsadüfi deyil ki, Azərbaycan Respublikası Prezidentinin 2020-ci ildə təsdiq olunmuş sərəncamına əsasən "Açıq hökumətin təşviqinə dair 2020-2022-ci illər üçün Milli Fəaliyyət Plan"ında bir sıra mühüm məsələlər öz əksini tapmışdır. Bu məsələlərə dövlət xidmətlərinin təkmilləşdirilməsi sahəsində tədbirlər (Elektron Hökumət infrastrukturunun təkmilləşdirilməsi və Elektron Hökumət İnformasiya Sisteminə qoşulan informasiya sistemləri və ehtiyatlarının sayının artırılması, "Elektron məhkəmə" informasiya sisteminin funksionallığının artırılması, "Ağıllı şəhər" konsepsiyasının tətbiqi və s.), informasiya əldə edilməsinin təmin olunması sahəsində tədbirlər (dövlət qurumlarının rəsmi internet informasiya ehtiyatlarında "Açıq məlumatlar" bölməsinin yaradılması, informasiyanın açıqlanması prosedurlarının təkmilləşdirilməsi, informasiyanın açıqlanması məqsədilə beynəlxalq standartlara uyğun məlumat toplularının formalaşdırılması ilə bağlı qaydaların müəyyən edilməsi və s.), vətəndaş cəmiyyəti üzvlərinin fəaliyyətinin genişləndirilməsi, ictimai nəzarətin və ictimai iştirakçılığın artırılması sahəsində tədbirlər və s. daxildir.

Yeni texnologiyalar: İT sürətlə dəyişir və dövlət orqanları bu dəyişikliklərin tempinə uyğunlaşmalıdır. Cəmiyyətdə İT resurslarından və e-dövlət texnologiyalarından istifadə etmək bacarığı formalaşmalıdır. Məlumat təhlükəsizliyi, fərdi informasiya təhlükəsizliyi və digər resurslarının təhlükəsizliyi ilə bağlı əsas qorxunu ictimai şüurdan götürmək lazımdır. Müasir tələblərə cavab vermək üçün İT əməkdaşlarına mütəmadi olaraq treninqlər keçirilməlidir.

Uğurlu e-idarəetmə üçün əsas tələb mümkün olan ən yaxşı internet xidmətlərinin təmin olunmasıdır. Ölkənin bütün regionlarında bu imkanları genişləndirmək lazımdır. Mobil xidmətlər yüksək standartlara cavab verməlidir. Getdikcə daha çox vətəndaş və müəssisə e-dövlət ilə qarşılıqlı əlaqə yaratmaq üçün smartfon və planşet kimi mobil texnologiyalardan istifadə etməyə başlayıb. Məsələn, 2013-ci ildə 300 min-dən çox Fransa vətəndaşı mobil proqramlar vasitəsilə vergi ödənişlərini həyata keçirmək üçün smartfonlarlardan istifadə etmişdir. Hal-hazırda

Avropa ölkələrində elektron dövlət xidmətlərinin yalnız ¼ hissəsi mobil qurğular üçün nəzərdə tutulmuşdur [53].

E-dövlətdə potensial olaraq böyük təsirə malik olan texnologiyalardan biri də bulud texnologiyalardır. Bulud texnologiyaları – ümumi kompüter resursları toplusunu (məsələn, şəbəkə, serverlər, məlumat anbarları, proqramlar və servislər) sorğu əsasında şəbəkə vasitəsilə əldə etməyə imkan verən modeldir. Bulud texnologiyaları müxtəlif konfiqurasiyalarda (məsələn, özəl, ictimai və ya hibrid bulud) və müxtəlif yollarla (məsələn, infrastruktur, platforma və ya proqram səviyyəsində sifarişçiyə təhvil qabiliyyəti ilə) istifadə oluna bilər. Bulud əsaslı e-xidmətlər İT xərclərinin azaldılması, eyni zamanda yeni və innovativ xidmətlərin sürətli yayılmasını dəstəkləyərək dövlət üçün yeni perspektivlər təklif edir. 2011-ci ildə Avropada elektron texnologiya sənayesini təmsil edən "Elektron Avropa" e-dövlətin fəaliyyət planının mühüm elementi kimi bulud hesablamaların olmasını tövsiyə etmişdir [127].

Potensialı böyük olsa da, bulud texnologiyalarından istifadə etmək istəyən dövlət orqanları qarşısında bir sıra mühüm məsələlər dayanır. Bunlara təhlükəsizliyin təmin olunması, məxfiliyin qorunması və müxtəlif sistemlər arasında əlaqənin təmini kimi bir sıra məsələlər daxildir. Digər məsələlərə texniki standartları nəzərdə tutan hüquqi şərtlər, bulud texnologiyaları vasitəsilə dövlət xidmətlərinə inamın yaradılması daxildir. Bəzi məsələləri həll etmək üçün Horizon 2020 tədqiqat proqramına bulud texnologiyalardan istifadə etməklə dövlət sektorunda məhsuldarlıq və innovasiyanın artırılmasına yönəlmiş fəaliyyət daxil edilmişdir [177]. Qeyd olunanları ümumiləşdirərək e-dövlətə dair aşağıdakı məsələlər ilə bağlı tədqiqatların aparılmasına ehtiyac var:

- ✓ Vətəndaşların e-iştiraka cəlb olunması;
- ✓ Vətəndaş məmnuniyyətinin artırılması;
- ✓ E-dövlət portalından istifadənin artmasına təsir edən amillərin müəyyən olunması;
- ✓ Vətəndaşların əsas maraq dairəsinin müəyyən olunması;
- ✓ İctimai rəy analizi vasitələrinin təkmilləşdirilməsi;

- ✓ Eksponensial sürətlə artan məlumatların analizi vasitələrinin təkmilləşdirilməsi;
- ✓ E-dövlət mühitində gizli sosial şəbəkələrin aşkarlanması və analizi metodlarının işlənməsi;
- ✓ Vətəndaşların e-dövlət portalında təhlükəsizliyinin təmin olunması;
- ✓ Yeni texnologiyaların tətbiqi;
- ✓ Fərdi məlumatların mühafizəsi problemləri;
- ✓ Rəqəmsal fərqlilik (Digital divide);
- ✓ Açıq dövlət, e-dövlət 2.0.

Göstərilən bu problemlərin həllində bir sıra metod və texnologiyalardan istifadə oluna bilər. Bunları nəzərə alaraq fəsil 1.2 və 1.3-də e-dövlətin analizində text mining və sosial şəbəkə analizi texnologiyalarının rolu araşdırılmış, bu sahədə aparılan tədqiqatlar nəzərdən keçirilmişdir.

1.2. E-dövlətin analizində text mining texnologiyaları

E-dövlət dövlət xidmətlərinin daha yaxşı çatdırılması, biznes və sənaye müəssisələri ilə qarşılıqlı əlaqələrin inkişafı, informasiyaya çıxış vasitəsilə vətəndaş səlahiyyətlərinin genişləndirilməsi və daha səmərəli dövlət idarəetməsini nəzərdə tutur. Başqa sözlə, e-dövlət daxili əməliyyatları sadələşdirir və maksimum rahatlıq, az xərclə vətəndaşların bütün təbəqələrinin dövlət xidmətlərindən faydalanmasına kömək edərək dövlət idarələrinin fəaliyyətini yaxşılaşdırır. Lakin bütün bunlarla yanaşı qeyd etmək lazımdır ki, bir çox cəhdlərə baxmayaraq e-dövlət xidmətləri hələ də istənilən vətəndaş - yönümlü xidmətləri göstərmir və həllini gözləyən bir sıra mühüm məsələlər mövcuddur. Belə ki, artan tələbat, iqtisadi, sosial və siyasi həyatın müxtəlif aspektləri üzrə dövlət proseduralarının mürəkkəbliyi - informasiyanın toplanması, ötürülməsi və paylanmasına əsaslanan qabaqcıl (müasir) bilikləri tələb edir.

E-dövlət quruculuğunda vətəndaşların iştirakını təmin etmək və dövlət xidmətlərinin bütün növlərinin səmərəliliyinin və effektivliyinin artırılması mühüm məsələlərdən biridir. Məlumdur ki, demokratik cəmiyyət bilikli (informasiyalı)

vətəndaşlar tələb edir. Vətəndaş və hökumət arasında informasiya mübadiləsi düzgün, səmərəli olduqda demokratik prinsiplərdən danışmaq olar [109]. İT-in köməyi ilə idarəetmənin bütün prosesində vətəndaşların iştirakının təmini e-demokratiyanın təşəkkülünə gətirib çıxarmışdır. E-demokratiyanın inkişafı vətəndaşlar və hökumət arasında məlumatların fasiləsiz axınını yaradır, vətəndaşların cəmiyyətdə daha fəal rol oynamasına və qərarların qəbul edilməsi prosesinə kömək edir. Qərar qəbuletmə prosesində vətəndaşların rəyinin nəzərə alınması mühüm məsələlərdən biridir. Dövlət onlayn müzakirə forumlarından istifadə edərək vətəndaşları dövlət layihələrini müzakirə etməyə təşviq edə bilər. Onlayn müzakirə forumları vasitəsilə vətəndaşların fikirlərinin öyrənilməsi rəy sorğularına əsaslanan ənənəvi üsullardan daha etibarlı olar və qərarların qəbul edilməsi prosesinə kömək edə bilər [38]. Bu da əsaslı surətdə vətəndaş rəylərinin iterativ olaraq araşdırılması üsullarını, eləcə də onların qiymətləndirilmə dəqiqliyini dəyişə bilər. Bununla belə, struktursuz verilənlərin həcmi və analizinin mürəkkəbliyi bu işi çətinləşdirə bilər. Text mining bu cür verilənlərin analizində mühüm vasitələrdən biridir [60].

Qeyd edək ki, e-dövlət mühitində Veb, bioloji və sənaye sensorları, video, e-poçt, sosial media daxil olmaqla bir çox mənbələrdən informasiya əldə olunur və bu məlumatların analizində də text mining texnologiyalarından istifadə oluna bilər. Bu məlumatların analizi e-dövlətə səhiyyə, təhsil, milli təhlükəsizlik, hüquq-mühafizə orqanları və digər sahələrdə öz öhdəliklərini yerinə yetirməyə kömək edə bilər [5,92]. Bu texnologiyalar vasitəsilə e-dövlət mühitində verilənlərin toplanması və analizi ilə maliyyə göstəricilərini yaxşılaşdırmaq, gəlirləri və xərcləri optimal idarə etmək, şəffaflığı yüksəltmək mümkündür. Həmçinin dövlət təşkilatları daxilində mühüm qərarların qəbul edilməsində, ictimai asayişin qorunmasında, sosial təminat, milli təhlükəsizlik, səhiyyə məsələlərində, istehlakçıların alıcılıq qabiliyyətini öyrənməklə marketinq işlərinin yaxşılaşdırılmasında, insanların davranışlarının analizində bu texnologiyalardan istifadə oluna bilər [7].

Text mining statistika, maşın təlimi, məlumat-axtarış, data mining, linqvistika və təbii dilin emalı anlayışlarını bir araya gətirən nisbətən yeni sahədir. Text mining mətn sənədlərindən yeni, relevant biliyin çıxarılması və gələcək istifadədə

informasiyanın daha yaxşı təşkili üçün bu biliyin istifadə olunması prosesidir. Əgər bənzətmə aparsaq, mining dəyərsiz qayalardan qiymətli "filiz külçələrinin", text mining isə mətn verilənləri "dağından" gizlənmiş "qızılın" (biliyin) çıxarılması prosesidir. Bu avtomatik olaraq müxtəlif yazılı mənbələrdən informasiyanın çıxarılması ilə yeni, əvvəllər məlum olmayan informasiyanın kompüter tərəfindən aşkar olunması prosesidir. Text mining məqalələr, veb səhifələr, müzakirə forumları, bloqlar kimi mətnlərin geniş blokundan əsas terminləri, xüsusiyyətləri, hadisələri çıxarmaq və atributlar arasında münasibətlərin müəyyən olunmasında istifadə olunur. Text mining bir sıra xüsusiyyətlərinə görə data mining-ə oxşardır. Data mining əsasən ədədi strukturlaşdırılmış məlumatlarla, text mining isə struktursuz məlumatlarla işlədikdə faydalıdır [2, 3].

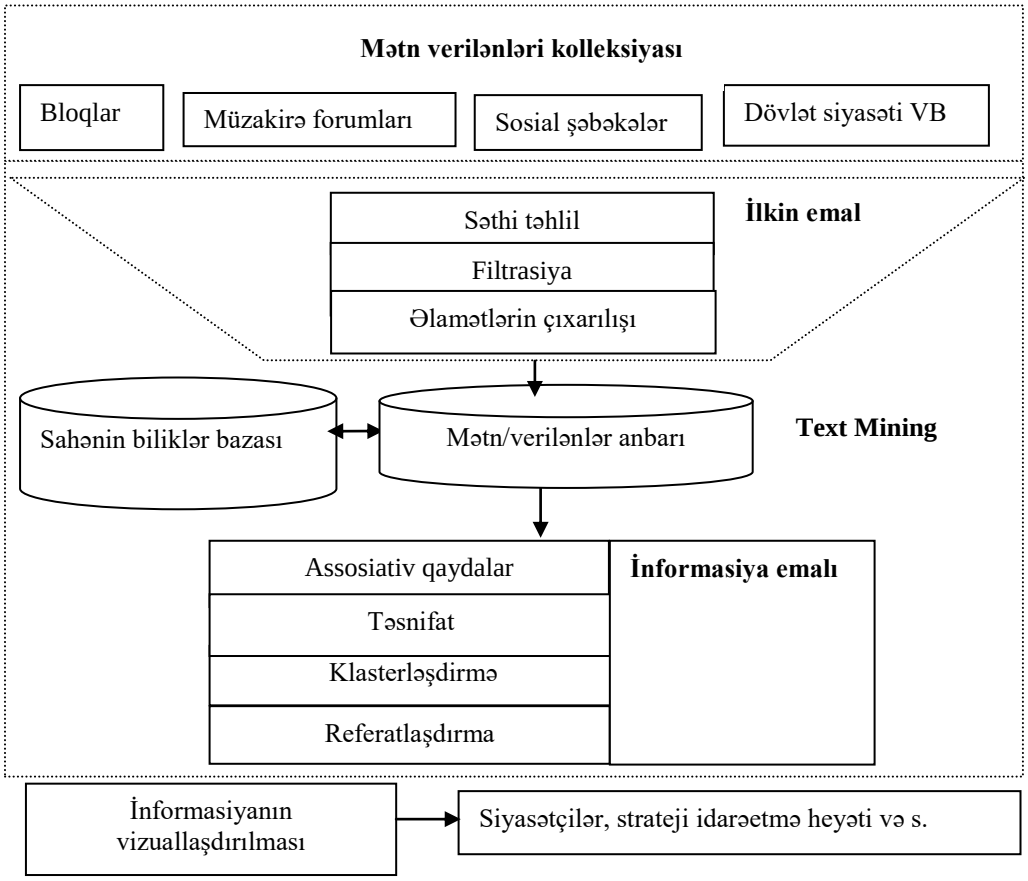
E-dövlətdə məlumatların artımı ilə əlaqədar olaraq həllini gözləyən bir çox mühüm məsələlər mövcuddur. Bir sıra tədqiqatçılar tərəfindən e-dövlətdə bu məsələlərin həllinə yönəlmiş text mining texnologiyasına əsaslanan metod və yanaşmalar təklif olunmuşdur. Onlardan bir neçəsi aşağıda nəzərdən keçirilmişdir:

[38]-də e-mesaj lövhələrində, e-məktublar və açıq müzakirə forumlarında vətəndaşlar tərəfindən yerləşdirilən məlumatları emal etmək qabiliyyətinə malik sistem təklif olunmuşdur. Bu sistem internet forumlardan mesajları toplayır, onları təsnif edir, ümumi xüsusiyyətləri və qanunauyğunluqları çıxardır. Sistemdə vətəndaşların rəyləri arasında olan tendensiyanı müəyyən etmək üçün assosiativ qaydalar texnologiyasından istifadə olunur. Bu qaydalar sistemin intellektual əsasını təşkil edir.

Qeyd edək ki, hər hansı intellektual sistemin həyata keçirilməsi üçün əsas addım lazım olan mənbələrin seçilməsidir. Bu mənbələrə dövlət siyasətinin verilənlər bazası, vətəndaşlara dövlət layihələrini müzakirə etməyə imkan verən onlayn müzakirə forumları, bloqlar və son dövrdə populyarlıq qazanmış sosial şəbəkə və ya sosial media daxil ola bilər. Burada, bir çox mənbələrdən və müxtəlif formatda (pdf, doc, docs, xml, jpg, html və s.) olan struktursuz informasiyadan bəhs olunduğuna görə sənədləri lazımi formata çevirmək üçün struktursuz və ya yarı strukturlaşdırılmış verilənləri idarə etmək bacarığına malik analiz sistemindən istifadə etməyə ehtiyac

yaranır. [131, s.272]-də e-dövlət qərarlarının dəstəklənməsi üçün vahid text mining texnologiyalarına əsaslanan vahid arxitektur təklif olunmuşdur. Bu arxitektur şəkil 1.2.1-də təsvir olunmuşdur. Bu tədqiqat işində siyasətçilərə kömək məqsədilə elektron ictimai forum, bloq və s. vasitəsilə vətəndaşların hökumət qərarları haqqında fikirləri, yazdıqları şərhlər ilə siyasət arasında əlaqələrin aşkarlanmasına cəhd göstərilir.

E-dövlətin prioritet məsələlərindən biri də vətəndaşların zərər və zorakılıqlardan qorunmasıdır. [106]-da vətəndaş – dövlət münasibətləri araşdırılmışdır. Bu işdə göstərilir ki, internet platforması cinayətlər barədə məlumat verilməsi üçün səmərəli və rahat üsuldur, cəmiyyət ilə hüquq-mühafizə orqanları arasında qarşılıqlı əlaqələri gücləndirir.



Şəkil 1.2.1 E-dövlət üçün text mining texnologiyasına əsaslanan qərar qəbuletmə sisteminin arxitekturu

Bu gün cinayət hadisələri barədə operativ məlumatların verilməsi üçün hüquq-mühafizə orqanları və qeyri-kommersiya təşkilatları tərəfindən müxtəlif tətbiqlər təklif olunur. Məsələn, Qısa Mesaj Xidməti (SMS), iPhone, iPad, və Android tətbiqləri anonim olaraq cinayət ipuclarını bildirmək üçün istifadə olunur. Başqa sözlə desək, vətəndaşlara şəxsiyyəti aşkar olmadan cinayət ipuclarını təqdim etməyə imkan verir. Bundan əlavə, hüquq-mühafizə orqanlarına vətəndaşlardan əldə olunan məlumatları toplamağa sərf olunan vaxta və resurslara qənaət etməyə kömək edir. Cinayətlər barədə onlayn verilənlər bazasında saxlanılan məlumatlar təbii dildə yazılır. Belə struktursuz mətnin həcmi artdıqca onun analizi problemləri yaranır [143]. Belə ki, geniş həcmli məlumatları filtirləmək, müqayisə etmək, fərqləndirmək böyük vaxt və zəhmət tələb edir. Buna görə də daha səmərəli həllər tələb olunur.

Mətn məlumatlarının analizi texnologiyalarından səmərəli istifadə olunması dövlət orqanlarına, xüsusilə hüquq-mühafizə orqanlarına informasiyanın daha səmərəli analizi və daha yaxşı qərarların dəstəklənməsində kömək edə bilər. [96]-da mətn analizi və təsnifat metodlarının e-dövlətin, xüsusilə hüquq-mühafizə orqanlarının səmərəliliyinin və effektivliyinin artırılmasında və qərar qəbul edənlərə vaxtında informasiya dəstəyinin verilməsində rolu araşdırılır. Bu tədqiqat işində cinayət analizinin asanlaşdırılması və avtomatlaşdırılması üçün təbii dil emalı üsulları, yaxınlıq ölçüləri və maşın təlimi üsullarından kombinasiya olunmuş Qərar Dəstəkləmə Sistemi (QDS) təklif olunmuşdur. Xüsusilə cinayətlər barədə məlumatların filtirlənməsi və bu məlumatların eyni və ya oxşar cinayət haqqında olub-olmamasını müəyyənləşdirmək mühüm məsələlərdəndir. Eyni cinayətə dair məlumatların aşkarlanması şübhəlilərin tutulması və ya tədbirlərin təkmilləşdirilməsində mühüm əhəmiyyət kəsb edir. Oxşar cinayətlərin tapılması cinayət tendensiyalarının və şəbəkə fəaliyyətinin analizi, hüquq-mühafizə resurslarının optimal bölüşdürülməsində vacibdir. QDS kimi sistemlərin yaradılması axtarış prosesinə və cinayətlər barədə məlumatların analizinə sərf olunan vaxta, resurslara qənaət edə, analitiklərin yükünü azalda, beləliklə, hüquq-mühafizə orqanları tərəfindən büdcə xərclərinin azaldılmasına səbəb ola bilər.

E-dövlət portalı ictimai xidmətlərin göstərilməsi və vətəndaş-dövlət münasibətlərində ən mühüm kanallardan birinə çevrilmişdir. E-dövlət sistemində dövlət orqanlarının saytları tərəfindən təqdim olunmuş böyük həcmli məlumatlar və zəngin informasiya resursları mövcuddur. Doğru qərar qəbul etmək və müəyyən məqsədlərə nail olmaq üçün bu informasiyadan ictimai maraq doğuran aktual məsələləri tez və vaxtında müəyyən etmək e-dövlətin əsas məsələlərindən biridir. E-dövlət mühitində bəzi informasiyalar üzərində insanların müraciətlərinin sayı və onların bu informasiyaya sərf etdikləri orta zamandan istifadə edərək insanların ən çox maraqlandıqları məsələləri müəyyən etmək olar. [164]-də e-dövlət mühtində insanların maraq dairəsində olan məsələləri müəyyən etmək üçün metod təklif olunmuşdur. Bu metodda istifadəçilərin e-dövlət portalında davranışını analiz etməklə ictimai maraq doğuran məlumatların təsnifatı həyata keçirilir. Burada, veb log mining və qeyri-səlis klaster analizi metodlarından istifadə olunmuşdur. Veb log mining vasitəsilə dövlət veb saytına daxil olan istifadəçilərin müraciət etdiyi informasiyalar müəyyən zaman periodu ərzində toplanır və obyektiv şəkildə bu müddət ərzində istifadəçilərin narahat olduğu aktual məsələlər müəyyən olunur. Qeyri-səlis klaster analizi vasitəsilə aktual məsələlər klasterləşdirilir.

İctimai-siyasi nöqteyi-nəzərdən istifadəçiləri e-dövlət xidmətlərindən istifadə etməyə motivasiya edən amillərin müəyyən olunması vacib məsələlərdən biridir. E-dövlət xidmətlərinin istifadəçilərinin sayının artmasına təsir göstərən amillərin müəyyən olunması dövlət orqanlarının fəaliyyətinin yaxşılaşdırılmasını həyata keçirəcək e-dövlət sistemləri üçün strategiyaları inkişaf etdirməyə imkan verə bilər. Bu amillərlə bağlı biliklər, onların ölçülməsi, bir-birilə bağlılığı uğurlu e-dövlət modelinin inkişafı, həyata keçirilməsi və idarə edilməsində böyük əhəmiyyət kəsb edir. Yeni texnologiyaların qəbulu və ya rədd edilməsi ilə bağlı qərarlar nəticə etibarilə fərdi istifadəçilər tərəfindən müəyyən olunur. Ona görə də hansı amillərin istifadəçinin e-dövlət sistemindən istifadə etmək qərarına təsir göstərdiyini bilmək vacibdir. Bu bilik e-dövlət xidmətlərinin tanınması və istifadəçilərinin sayının artırılmasında mühüm əhəmiyyət kəsb edir. [118]-də vətəndaşların e-dövlət xidmətlərindən istifadəsinə müxtəlif demoqrafik, idraki və psixoloji amillərin təsirini

müəyyən etmək üçün data mining üsullarına əsaslanan metod təklif olunmuşdur. Bu işdə hesablama nöqteyi-nəzərindən qərar ağacları, neyron şəbəkələri, induksiya qaydası, maşın təlimi və qraf vizuallaşdırma metodlarından istifadə olunur. Üç süni neyron şəbəkə modelləri (çoxlaylı neyron şəbəkələr, stoxastik neyron şəbəkələr, özünü təşkil edən neyron şəbəkələr) və üç maşın təlim üsulları (təsnifat və reqressiya ağacları, çoxdəyişənli adaptasiya reqressiya əyriləri və dayaq vektorlar metodu) standart statistik metod (xətti diskriminant analiz) ilə müqayisə olunur. Dəyişənlər çoxluğu olaraq cins, yaş, təhsil səviyyəsi, e-dövlət xidmətlərinin faydalılığı, istifadə rahatlığı, uyğunluq, etimad, vətəndaş rəyi və münasibətlər qəbul edilir. Bu tədqiqat işində aparılan araşdırma nəticəsində verilənlər bazasında kompleks (mürəkkəb) nümunələrin ortaya çıxarılması ilə e-dövlət xidmətlərindən istifadə davranışının müxtəlif aspektlərinin müəyyənləşdirməsinin mümkünlüyü və data mining üsullarının təsnifat bacarıqları müəyyən olunmuşdur.

E-dövlət sahəsində bir sıra dinamik e-idarəetmə saytları/portalları mövcuddur. İstifadəçilər bu saytlardan/portallardan əhəmiyyətli məlumatlar almaq, idarəetmə və xidmətlə bağlı bir sıra tələblər irəli sürmək üçün istifadə edə bilərlər. Vətəndaşların sərbəst və rahat şəkildə dövlət xidmətləri haqqında fikir və təcrübələrini yaza biləcəyi bloqların olması dövlət üçün daha məqsədə uyğundur. Belə ki, vətəndaş rəylərinin nəzərə alınması demokratik dövlətin inkişafına, xidmətlərin keyfiyyətinin yaxşılaşdırılmasına səbəb ola bilər. [158]-də mətn emalı və data mining üsullarından istifadə edərək prototip sistem təklif olunmuşdur. Bu sistem altı xidmət kateqoriyası üzrə vətəndaşların rəylərini təsnifləşdirir. Bu xidmət kateqoriyalarına təhsil, kənd təsərrüfatı, səhiyyə, nəqliyyat, İT və sənaye daxildir. İlk növbədə e-dövlət portalında yerləşdirilən vətəndaş rəyləri toplanır. Daha sonra verilən rəylər sonrakı təhlil üçün ilkin emal olunur. İlkin emal mətndən lazımsız və ya etibarsız məlumatların kənarlaşdırılmasından ibarət data mining proseslərindən biridir. Bundan sonra təsnifat məqsədilə açar sözlərin lüğəti yaradılır. Lazımsız sözlərin kənarlaşdırılması üçün ümumişlək sözlər lüğəti, rəyləri altı kateqoriya üzrə təsnif etmək üçün açar sözlərin lüğəti və qərar qəbuletmə məqsədilə tərif, tənqid və neytral (bitərəf) olmaqla rəyləri təsnif etmək üçün müsbət və mənfi sözlərin lüğəti yaradılır.

E-dövlətin mühüm məsələlərindən biri vətəndaşlara qərar qəbuletmə prosesində iştirak etmək imkanının verilməsidir. [149]-da hökumət qərarları haqqında vətəndaş rəylərinin toplanması və analizi üçün text və data mining üsullarından istifadə olunmuşdur. Burada, iki səviyyəli analizdən istifadə olunur. Yanaşmanın birinci hissəsi istifadəçi rəylərinə əsaslanaraq istifadəçilərin mövqeyini müəyyən edən ifadələrin müəyyən olunması və çıxarılmasından ibarətdir. İkinci hissədə müəyyən edilən ifadələrin müzakirə edilən məsələyə müsbət və ya mənfi münasibətini müəyyən etmək üçün analiz olunur. Məlumatları emal etmək üçün HTML parsingdən istifadə olunur. Məlumat bloku onun vasitəsilə kiçik hissələrə bölünür və bu qeyri-mətn elementlərinin aradan qaldırılmasına kömək edir. Qeyri-mətn elementləri dedikdə şəkillər, videolar, skriptlər, mətn qrafik vasitələri və s. nəzərdə tutulur. Daha sonra istifadəçi mətnlərinin leksik elementlərini çıxarmaq üçün leksik analiz (tokenization) tətbiq olunur. Sonra qərarlar ağacı təsnifat modulundan istifadə olunur.

Bütün bu tədqiqatlar göstərir ki, text mining texnologiyaları hal-hazırda e-dövlət mühitində məlumatların analizində mühüm rol oynayır və bu texnologiyalara əsaslanan metodların təkmilləşdirilməsinə ehtiyac var.

1.3. E-dövlətin analizində sosial şəbəkə analizi texnologiyaları

Veb-in getdikcə daha geniş tətbiqi dövlət-vətəndaş münasibətlərinin potensial inkişafına - dövlətlə vətəndaş, biznes və digər dövlət qurumları ilə qarşılıqlı əlaqədə İT-dən istifadənin genişlənməsinə gətirib çıxarmışdır. Veb 2.0 texnologiyalarının geniş tətbiqi vətəndaşlar və dövlət sektoru arasındakı münasibətlərdə sözün əsl mənasında yeni mərhələ açmışdır. İT və informasiya cəmiyyətinin inkişafı dövlət sektoru üçün aşağıdakı kimi müxtəlif potensial imkanlar yaratmışdır [1, s.32]:

- İnternet üzərindən informasiyanın daşınması;
- Dövlət orqanları ilə vətəndaşlar, biznes və digər dövlət qurumları arasında əlaqələrin genişləndirilməsi;
- Əməliyyatların və qeydiyyatların aparılması;

- İdarəetmə və birbaşa demokratiya mexanizmlərinin işlənməsi (onlayn məsləhətləşmələr, seçki, səsvermə və s.).

Dövlət qurumları və vətəndaşlar arasında informasiya mübadiləsi, sənəd axını dövlət şəbəkələrinin yaranması ilə nəticələnir [96,109]. Dövlət şəbəkələri "Təşkilatlararası şəbəkəyə" (TŞ) əsaslanır. "Təşkilatlararası şəbəkə"nin təpələri təşkilatların müxtəlif növ və səviyyələri, tilləri isə bu təpələr arasında istənilən əlaqələri əks etdirir. Şəbəkənin təpələrinə sosial şəbəkə analizində aktor deyilir [6, s.12]. Aktorlar kimi təşkilatların hər bir sinfi ola bilər: dövlət, biznes, qeyri-hökumət təşkilatları və onların qarışığı. Əlaqələr aktorlar arasında müxtəlif axınları xarakterizə edir: əməliyyat axınları, maddi resursların transferi (məsələn, pul və ya digər mallar), qeyri-maddi resursların axını (informasiya, sənədlər və s), bilik axtarışı və nəşri [50,129]. Dövlət aktorları dedikdə informasiyanın yaradılması, mübadiləsi və ötürülməsini həyata keçirən dövlət orqanları, təşkilatlar nəzərdə tutulur. Onlar informasiya axınını qanunvericilik mandatı, təşkilati təcrübə və ictimai ehtiyaca görə hazırlayır və layihələndirirlər. Təşkilatlararası şəbəkə gəlirlərin artırılması və ya xərclərin azaldılması məqsədilə "məhsul istehsalı və ya xidmətlərin göstərilməsində birlikdə qərar qəbul edən və öz səylərini inteqrasiya edən təşkilatların klasterləri" kimi müəyyən olunur.

Qeyd edək ki, klassik dövlət formatında vətəndaşlar təşkilatlar ilə qarşılıqlı əlaqə nəticəsində yaranan şəbəkədə aktiv, mərkəzi rol oynayırdı. Belə ki, düzgün sənəd axınının yaradılması üçün vətəndaşlar bir təşkilatdan digərinə getməli olurdu, bu isə daha çox zaman və maliyyə itkisi ilə nəticələnirdi. Bundan fərqli olaraq, müvafiq əməliyyat sistemləri ilə dəstəklənən daha yetkin e-dövlət xidmətlərinin göstərilməsində əsas diqqət ictimai xidmətlərin göstərilməsinə yönəlmişdir. Bu vətəndaş yönümlü e-dövlət adlanır. Bu situasiyada vətəndaş "bir pəncərə" prinsipi ilə lazım olan bütün xidmətləri alır. Belə ki, idarəetmə orqanları sənəd axını şəbəkəsində fəal rol oynayır və xidmətlərin həyata keçirilməsi proseslərində vətəndaşların yükünü azaldır. Dövlət şəbəkələrini analiz etmək üçün sosial şəbəkə analizi metodlarından çox geniş istifadə oluna bilər.

Sosial şəbəkə analizi şəbəkə daxilində aktorlar arasındakı münasibət və əlaqələrə əsaslanır. Şəbəkə analizində əsas aksiom aktorların müstəqil deyil, əksinə bir-birinə təsir edən olmasıdır. Sosial şəbəkə analizi sosial subyektlər arasında əlaqələrin strukturunu öyrənir və bu əlaqələrin nəticələri ilə əlaqədardır. Sosial şəbəkə analizi digər tərəfdən "məlumatın yığılması, statistik analizi və vizuallaşdırılması üçün konkret metodları əhatə edən riyazi metodlar və fərqli metodologiyalar çoxluğundan" istifadə edərək şəbəkədə "sosial əlaqələrin analizi" kimi müəyyən edilir [122]. Buna görə, sosial şəbəkə analizi "aktorlar arasında sosial əlaqələrin qanunauyğunluqlarının aşkarlanması və interpretasiyası" məqsədilə statistika ilə yanaşı güclü metodoloji alətə çevrilmişdir. Sosial şəbəkə analizi "bir-birilə əlaqəli aktorlar çoxluğu" [36]; "bir və ya daha çox əlaqəli sosial yönümlü aktorlar çoxluğu" [111], "aktorlar arasında bütün müvafiq əlaqələrin təsviri" kimi müəyyən olunur. Sosial şəbəkə analizi üzrə bir sıra alətlər mövcuddur. Bu alətlər tətbiq olunma mühitindən asılı olaraq elmi tədqiqat (R, ORA, Pajek), biznes (NetMiner, Nodexl, InFlow) və s. yönümlü olurlar. Bu alətlər tədqiqatçılara müxtəlif ölçülü şəbəkələrin – kiçikdən çox böyüyə qədər araşdırılmasına imkan verir. Müxtəlif alətlər şəbəkə modelinə tətbiq edilə bilən riyazi və statistik prosedurları təmin edir.

Qeyd edək ki, son on il ərzində sosial şəbəkə analizinin tədqiqində yeni mərhələ başlamışdır. Belə ki, əsas diqqət kiçik şəbəkələrin analizindən minlərlə və ya milyonlarla təpə nöqtəsi olan sosial şəbəkələrin analizinə yönəlmişdir. Bunun iki səbəbi var: 1) sosial media şəbəkələrinin kütləvi istifadəsi və genişlənməsi; 2) bu mediadan əldə olunan böyük verilənlər bazasının mövcudluğu. Sosial şəbəkələrə maraq artsa da, e-dövlətdə bu metodologiyanın tətbiqinə az diqqət ayrılmışdır. Bu sahədə bəzi tədqiqatçıların gördüyü işləri nəzərdən keçirək:

[83]-də e-dövlətin qiymətləndirilməsi üçün sosial şəbəkə analizinə əsaslanan yeni yanaşma təklif olunmuşdur. Burada əsas diqqət müxtəlif dövlət idarəetmə orqanları, vətəndaşlar və biznes müəssisələri arasında sənədlərin axını şəbəkəsinə yönəlmişdir. Bu tədqiqat işində şəbəkələrin yaradılması 4 mərhələdə həyata keçirilir:

1. Mərkəzi reyestr tərəfindən göstərilən xidmətlərin müəyyən edilməsi: bu məqsədlə veb saytlarda və normativ aktlarda verilən məlumatların analizindən istifadə olunur.

2. Hər bir xidmətin çatdırılması protokollarının formalaşdırılması: bu məqsədlə normativ aktlarda göstərilən məlumatların və xidmətlərin çatdırılması prosesinə cəlb edilən qərar qəbul edənlərin müsahibələrinin analizinin kombinasiyasından istifadə olunur.

3. Yuxarıda göstərilən hər bir xidmət üçün 2 şəbəkə yaradılır: islahatdan əvvəlki və sonrakı vəziyyəti göstərən şəbəkələr.

4. Ayrı-ayrı xidmətlərin şəbəkələrinin əvvəlki və sonrakı vəziyyətlərə uyğun şəbəkəyə qoşulması.

Bu şəbəkədə 3 növ aktor fərqləndirilir: xidmətlərin göstərilməsi prosesinin təşəbbüskarı kimi vətəndaş; xidmətlərin göstərilməsi prosesini həyata keçirən dövlət idarəetmə müəssisələri və digər təşkilatlar. Şəbəkədə bütün əlaqələr sənəd axınını təmsil edir; əlaqələrin (tillərin) sayı iki əlaqəli aktor arasında mübadilə olunan müxtəlif sənədlərin sayına uyğundur. Burada şəbəkənin yaradılması və vizuallaşdırılması üçün Pajek proqramından istifadə olunmuşdur.

E-dövlət saytları müxtəlif subyektiv ölçmələrin köməyi ilə konsaltinq şirkətləri, beynəlxalq təşkilatlar və elmi tədqiqatçılar tərəfindən mütəmadi olaraq qiymətləndirilir. [125]-də e-dövlətin qiymətləndirilməsində yeni metod təklif olunmuşdur. Bu tədqiqat işində əsas məqsəd Vebometrika və sosial şəbəkə analizi metodlarından istifadə edərək bu qiymətləndirmələrin yaxşılaşdırılmasından ibarətdir. Bu işdə təklif olunan metod pilot olaraq Kanada, ABŞ, Böyük Britaniya, Yeni Zelandiya və Çexiya dövlət saytlarında nəzərdən keçirilmişdir. Burada aktorlar arasında orta məsafədən, yolların uzunluğunun paylanması, giriş və çıxış dərəcələrindən istifadə olunmuşdur. Şəbəkə aktorlar çoxluğundan və aktorları cüt-cüt birləşdirən istiqamətlənmiş tillər çoxluğundan ibarətdir. Aktorlar İnternetdən əldə olunan sənədlər; əlaqələr isə bu sənədlərin arasında naviqasiya üçün istifadə oluna biləcək hiperlinklərdən ibarətdir. Hər bir aktorun giriş dərəcəsi aktora daxil olan tillərin sayını, çıxış dərəcəsi isə aktordan çıxan tillərin sayını göstərir. Giriş və çıxış

dərəcələrinin cəmi aktorun dərəcəsini göstərir. Bu ölçmələrdən saytın naviqasiyalılığı və mərkəziliyini (ing. nodality) müəyyən etmək üçün istifadə olunur. Burada aparılan analizlərdə ABŞ və Kanada saytlarının Böyük Britaniyadan daha yaxşı naviqasiyanı təmin etdiyi göstərilir.

[84]-də e-dövlətin daha yaxşı və səmərəli təhlilinə nail olmaq məqsədilə sosial və informasiya şəbəkələrinin analizi metodlarından istifadə olunmuşdur. Bu tədqiqat işində dövlətin Veb-də iştirakı və müxtəlif dövlət idarələrinin bir-biri ilə qarşılıqlı əlaqəsi analiz olunmuşdur. Bundan başqa dövlət orqanları ilə qeyri-hökumət təşkilatları arasında əlaqələr araşdırılmışdır. Bu işdə əsas məqsəd sosial şəbəkə analizi üsullarının köməyilə e-dövlət hiperlinkləri üzərində səmərəli tədqiqatın yerinə yetirilməsidir. Hiperlinklərin şəbəkəsinin yaradılması aşağıdakı şəkildə həyata keçirilmişdir: Əvvəlcə müxtəlif Veb təşkilatları arasında əlaqələrin verilənlər bazası yaradılır. Əldə olunmuş informasiya saxlanılır və qraf şəklində təqdim olunur. Təpələr veb səhifələri, istiqamətlənmiş tillər isə hiperlinki göstərir. Bütün belə təpələr və istiqamətlənmiş tillər çoxluğu veb qraf kimi təqdim olunur. Ola bilər ki, bu istiqamətlənmiş qraf güclü əlaqəli olmasın: elə səhifə cütü ola bilərki, birindən digərinə hiperlinklər vasitəsilə keçmək mümkün olmasın. Burada səhifə daxilində olan hiperlinklərə daxili linklər, səhifəyə kənardan olan linklərə xarici linklər kimi istinad olunur. Daha sonra yaradılmış şəbəkədən e-dövlətin daxildən və xaricdən necə əlaqə saxladığı araşdırılır. Dövlət hiperlinklərinin şəbəkəsinin yaradılması üçün VOSON sistemindən istifadə olunur. Bu hiperlink şəbəkələrin toplanması və təhlili üçün istifadə olunan şəbəkə proqram təminatıdır. Şəbəkə yaradıldıqdan sonra şəbəkə qraflarının araşdırılması üçün NodeXL proqram təminatından istifadə olunur. NodeXL ödənişsiz, açıq sistemdir [167]. Şəbəkənin analizi üçün əsas 4 mərkəzilik ölçüləri tətbiq olunur: dərəcə üzrə mərkəzilik, yaxınlıq üzrə mərkəzilik, vasitəçilik üzrə mərkəzilik və PageRank. Şəbəkədə ən "məşhur" subyekti (yəni, hansı subyektə daha çox təpə birləşdiyini) tapmaq üçün dərəcə üzrə mərkəzilikdən istifadə olunur. Şəbəkədə daha güclü təsirə malik subyekti (saytı) tapmaq üçün vasitəçilik üzrə mərkəzilikdən istifadə olunur. Təpələr arasında ən qısa məsafəni tapmaq üçün yaxınlıq üzrə mərkəzilikdən istifadə olunur. Ən əhəmiyyətli aktoru tapmaq üçün

PageRank istifadə olunur. Sonda şəbəkənin analizi nəticəsində hansı müəssisənin ən məşhur, güclü, əhəmiyyətli və mərkəzi olduğu müəyyən olunur.

İnternetin artan istifadəsi və texnologiyaların inkişafı dövlət siyasətində əsas alətlərdən biri olan "mərkəzilik" anlayışına öz təsirini göstərir, yəni hansı dövlət orqanı sosial və informasiya şəbəkələrinin mərkəzindədir [72]. [59]-da sosial şəbəkə analizi metodlarından Veb-də dövlətin mərkəziliyinin keyfiyyətli analizi üçün istifadə olunmuşdur. Burada ilk olaraq mərkəzilik, görünüş (ing. visibility) və naviqasiya (ing. navigability) anlayışları müzakirə olunur. Daha sonra üç ölkənin (ABŞ, Böyük Britaniya və Avstraliya) xarici işlər ilə əlaqədar müvafiq nazirliklərinin veb-saytlarına iki əsas metrika tətbiq olunur: struktur və istifadəçi göstəriciləri. Struktur göstəricilərindən iki təsadüfi səhifə arasında orta məsafə və saytların qarşılıqlı əlaqəsinin hesablanmasında istifadə olunur. Sosial şəbəkə analizi göstəricilərindən hər saytın naviqasiyasını müqayisəli qiymətləndirmək üçün istifadə olunur. Bu işdə əsas diqqət xidmət əməlyyatlarından daha çox axtarış əməliyyatlarına ayrılmışdır. Alınmış nəticələr beş istiqamət üzrə veb-saytın qiymətləndirilməsi üçün istifadə olunmuşdur: görünüş, ekstraversiya, əlyetərlilik, naviqasiya və rəqabətə davamlılıq.

Burada görünüş daxili link analizi vasitəsilə, ekstraversiya xarici link analizi vasitəsilə hesablanır. Saytda ana səhifədən kənara çıxan daxili linklərin faizi informasiyanın əlçatanlığının ölçüsü kimi qəbul olunur. Naviqasiya vebometrika və istifadəçi eksperimentləri vasitəsilə hesablanır.

[67]-də əsas diqqət dövlət orqanlarının (təşkilatların) Facebook vasitəsilə auditoriya ilə əlaqəsi, onların yazdığı postlara vətəndaş rəylərinin analizinə yönəlmişdir. Facebook formal hesablar çərçivəsində xüsusi sosial qruplar formasında fərdlər arasında söhbəti asanlaşdırır. Bu tədqiqat şəbəkənin üfüqi vəziyyəti, şəbəkə sıxlığı və şəbəkə üzvləri arasında şərh sayına əsaslanır. Bu üçtərəfli münasibətləri test etmək üçün dövlət orqanlarının yazdığı 4280 mövzu (post) analiz olunmuşdur. Nəticələr göstərir ki, 1) şəbəkənin üfüqiliyi istifadəçilərin sayının artmasına müsbət təsir göstərir; 2) Veb 2.0 əsaslı mühitdə istifadəçilərin yazdığı şərhlər təşkilatlar tərəfindən yayımlanan postlardan daha vacibdir 3) təşkilatlar tərəfindən qeyd olunan postların sayının artımı istifadəçilərin yazdığı şərhlərin sayına mənfi təsir göstərir.

[90]-da e-dövlət sahəsinin müəyyən aspektlərini təyin etmək üçün hibrid yanaşma (sosial şəbəkə analizi metodları və üçqat spiral göstəriciləri) təklif olunur. Bu tədqiqat işində təşkilat, ölkə və regionlar üzrə e-dövlətlə bağlı həmmüəllifli məqalələr əsasında şəbəkə qurulur və analiz olunur. Dərəcə üzrə mərkəzilik, sıxlıq və klaster daxil olmaqla sosial şəbəkə analizi metodlarından e-dövlət şəbəkəsinin gizli şəbəkə strukturları və xüsusiyyətlərini müəyyən etmək üçün istifadə olunur. Klasterlər NetDraw proqram paketində spring-embedding alqoritmi vasitəsilə təyin olunur. Burada linklər təşkilatlar, ölkə və regionlar üzrə həmmüəllifli məqalələrin sayına uyğundur. Şəbəkə xüsusiyyətləri NetMiner 3.3.0 vasitəsilə hesablanır. Heliks göstəriciləri ikitərəfli və üçtərəfli əlaqələri yoxlamaq üçün istifadə olunur.

Göründüyü kimi, e-dövlət sahəsində sosial şəbəkə analizinin tədqiqi böyük potensiala malikdir və bu sahədə tədqiqatların aparılmasına ehtiyac var.

Yuxarıda deyilənləri nəzərə alaraq, dissertasiya işində aşağıdakı məsələlərə baxılır:

1. e-dövlətdə terrorizmlə əlaqəli mətnlərin aşkarlanması üçün hibrid təsnifatlandırma metodun işlənilməsi;
2. e-dövlətdə terrorizmlə əlaqəli mətnlərin filtrasiyası üçün sentiment analiz texnologiyasına və Bayes klassifikatoruna əsaslanan metodun işlənilməsi;
3. e-dövlətdə gizli sosial şəbəkələrin aşkarlanması və analizi üçün sentiment analiz texnologiyasına əsaslanan metodun işlənilməsi;
4. e-dövlət xidmətlərindən vətəndaş məmnuniyyətinin avtomatik qiymətləndirilməsi üçün metodun işlənilməsi;
5. e-dövlətdə vətəndaşları (o cümlədən regionları) maraqlandıran aktual mövzuların müəyyən edilməsi üçün klasterləşmə və mövzu modelləşdirmə texnologiyalarına əsaslanan metodun işlənilməsi.

II FƏSİL. E-DÖVLƏTDƏ TERRORİZMİ TƏBLİĞ EDƏN MƏTNLƏRİN AŞKARLANMASI ÜÇÜN METODLAR

Müasir dövrdə kriminal qruplar təkcə real aləmdə deyil, həm də virtual mühitdə (İnternet, e-dövlət) də dövlət və cəmiyyət əleyhinə öz bədniyyətli fəaliyyətlərini həyata keçirirlər. Bu fəaliyyət növləri müxtəlif məqsədli olur: dövlət əleyhinə təbliğat, mentalitetə uyğun gəlməyən, milli mənəvi dəyərlərin əsaslarını sarsıdan, terrorizmi təbliğ edən informasiyanın yayılması və s.

E-dövlət mühitində bu məzmunda informasiyanın vaxtında aşkarlanması dövlətin və cəmiyyətin təhlükəsizliyinin təmin olunması baxımından mühüm əhəmiyyət kəsb edir və günümüzün ən aktual elmi-nəzəri və praktiki problemlərindən biridir. Heç də təsadüfi deyildir ki, e-dövlətin təhlükəsizliyi problemi Avropa Komissiyası tərəfindən qəbul edilmiş eGovRTD2020 layihəsində e-dövlət sahəsində araşdırılması vacib olan 13 ən aktual elmi-tədqiqat istiqamətindən biri kimi qeyd olunmuşdur [168].

E-dövlətin əsas funksiyalarından biri vətəndaşları ehtimal olunan zərər və zorakılıqlardan qorumaqdır. Təcrübə göstərir ki, bu əlverişli mühitdən cinayətkar qruplar da yaxşı "yararlanırlar" və onlar bu imkandan istifadə edərək dövlət və cəmiyyət üçün böyük təhlükə mənbəyinə çevrilirlər. Buna misal olaraq, 11 sentyabr 2001-ci il tarixində ABŞ-da həyata keçirilmiş terror hücumunu göstərmək olar. Terror hadisəsindən sonrakı təhlillər göstərdi ki, bu aktı həyata keçirən mütəşəkkil cinayətkar qrup bütün plan və fəaliyyətlərini İnternet şəbəkəsindən istifadə etməklə hazırlamış və koordinasiya etmişlər. Belə demək mümkünsə, virtual aləm cinayətkar qruplara öz əməllərini həyata keçirmək üçün çox əlverişli mühitdir [3].

Bunları nəzərə alaraq, dissertasiya işində e-dövlət mühitində terrorizmlə əlaqəli məqalələrin aşkarlanması və filtrasiyası üçün text mining texnologiyasına əsaslanan metodlar təklif olunmuşdur.

2.1. E-dövlətdə terrorizmi təbliğ edən mətnlərin təsnifatı üçün metodlar

Dövlətin mühüm vəzifələrindən biri də virtual mühitdə – İnternetdə və e-dövlətdə gizli fəaliyyət göstərən kriminal şəbəkələrin fəaliyyətini aşkarlamaq və analiz etməkdir. Bu mühit tez kommunikasiya yaratmaq və fəaliiyəti operativ koordinasiya etmək baxımından çox geniş imkanlara malikdir. Kriminal şəbəkənin üzvləri ünsiyyət qurmaq üçün veb-saytlardan, e-poçtdan, bloqlardan, onlayn çatdan və s. istifadə edir. Aydındır ki, belə kommunikasiya vasitələrində ötürülən informasiya növləri arasında mətnlər üstünlük təşkil edir. Ona görə də, mümkün ola biləcək terror aktlarının qarşısının alınması və dövlətin təhlükəsizliyinin təmin olunması üçün virtual mühitdə, o cümlədən e-dövlətdə dövr edən mətnlərin analizi mühüm əhəmiyyət kəsb edir. Hal-hazırda biliklərin idarə olunmasında, müxtəlif mənbələrdə toplanmış mətnlərin intellektual analizində text mining ən qabaqcıl və effektiv texnologiyalardan biri hesab olunur [9].

Text mining texnologiyalarının belə populyar və tətbiq sahəsinin geniş olmasının digər səbəblərindən biri də real və ya virtual mühitdə istehsal olunmasından asılı olmayaraq informasiya növləri arasında mətnlərin üstünlük təşkil etməsidir. Beynəlxalq verilənlər korporasiyasının (International Data Corporation) analitiklərinin verdiyi məlumata görə istehsal olunan informasiyanın təqribən 80%-dən çoxunu mətnlər təşkil edir. Deməli, e-dövlətin təhlükəsizliyinin təmin olunması baxımından bu mühitdə dövr edən mətnlərin intellektual analizi mühüm əhəmiyyət kəsb edir və elmi-tədqiqat nöqteyi-nəzərindən aktual məsələdir.

Qeyd edək ki, virtual mühitdə (İnternetdə, e-dövlətdə) kriminal və terrorizmlə bağlı informasiyanın aşkarlanması, identifikasiyası və izlənməsi üçün text mining texnologiyasına əsaslanan müxtəlif metodlar, alqoritmlər və modellər təklif edilmişdir. Məsələn, veb-də kriminal informasiyanın filtrasiyası və identifikasiyası məqsədilə sənədlər arasındakı oxşarlığı müəyyən etmək üçün [97]-də yeni alqoritmlər təklif edilmişdir. Ərəb dilində kriminal sənədlərin identifikasiyası sistemi üçün [27]-də text mining texnologiyasının informasiyanın çıxarılması və klasterləşdirmə metodlarından istifadə olunmuşdur. İnformasiyanın çıxarılması üçün qaydalara əsaslanan yaxınlaşma, sənədlərin klasterləşdirilməsi üçün isə özü-özünə təşkil olunan

neyron şəbəkə (Kohonen şəbəkəsi) tətbiq olunmuşdur. Kriminal sənədlərin tipinin identifikasiyası üçün [37]-də iki mərhələdən - sənədlərin aşkarlanması və onların klasterləşdirilməsindən ibarət metod təklif olunmuşdur. Birinci mərhələdə sənədlər əhəmiyyətsiz sözlərdən təmizlənir, sonra sənədləri əhəmiyyətli sözlərin vektoru kimi təsvir edib, onlar arsındakı yaxınlığı hesablamaq üçün metrika daxil edilir. İkinci mərhələdə klasterləşdirmə alqoritmini tətbiq etməklə sənədlər kriminal tiplərə görə qruplaşdırılır. İnternetdə terrorizmlə əlaqəli məqalələri aşkarlamaq üçün [48]-də mətnlərin analizinə əsaslanan yeni yanaşma təklif olunmuşdur. Bu yanaşma WordNet semantik şəbəkəsindən [116] istifadə etməklə terrorizmlə əlaqəli məqalələr çoxluğundan kontekst sözlərin (isimlərin) siyahısını yaradır. Sonra WUP [171] metrikasını tətbiq etməklə kontekst sözlərin əhəmiyyətlilik dərəcəsini hesablayır. Sonda isə biqramlardan və Keselj metrikasından [88, s.258] istifadə etməklə sənədləri təsnifatlandırır. Terrorizmlə əlaqəli çoxdilli sənədlərin aşkarlanması üçün [100]-də təsnifatlandırma metoduna əsaslanan yeni yanaşma təklif olunmuşdur. Bu yanaşma veb sənədlərin qraf təsviri modeli ilə C4.5 təsnifatlandırma alqoritminin kombinasiyasından istifadə edir. [138]-də təklif olunmuş metod data mining alqoritmlərinin köməyilə veb saytlardakı mətnləri analiz etməklə terrorçuların fəaliyyətini (profilini) öyrənir. Kriminal məzmunlu mətnləri təsnifatlandırmaq üçün [139]-da qeyri-səlis qrammatikanın evolyusiyası (evolving fuzzy grammar) metodu təklif olunmuşdur. Bu metodda seçilmiş mətn fraqmentləri qeyri-səlis strukturda təsvir olunur.

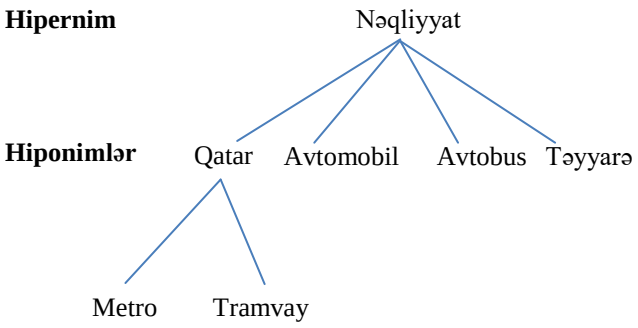
Beləliklə, problemin aktuallığını əsas tutaraq, e-dövlətdə şübhəli (terrorizmlə əlaqəli) mətnlərin aşkarlanması üçün text mining texnologiyalarına əsaslanan metod təklif etmişik. Bu metod [48]-də təklif olunmuş metoda oxşardır. Lakin təklif olunan metod bir neçə fərqli və üstün cəhətlərə malikdir:

- [48]-də təklif olunmuş metoddan fərqli olaraq, bu metodda sözlər arasındakı yaxınlığı hesablayarkən nəinki onlar arasındakı semantik yaxınlıq, həm də cümlənin sintaktik quruluşu, daha doğrusu sözlərin cümlədəki işlənmə ardıcıllığı nəzərə alınır;

- potensial şübhəli sənədləri daha dəqiq aşkarlamaq üçün sənədlər arasındakı yaxınlıq yeni iterativ üsulla hesablanır: əvvəlcə sözlərin yaxınlığı təyin edilir; sonra sözlər arasındakı yaxınlıqdan istifadə etməklə cümlələrin yaxınlığı hesablanır; nəhayət, cümlələr arasındakı yaxınlıqdan istifadə olunmaqla sənədlər arasındakı yaxınlıq hesablanır.
- cümlələr arasında yaxınlığı hesablamaq üçün hibrid yaxınlıq ölçüsü daxil edilir;
- təsnifatlandırma üçün yeni metod təklif olunur.

Təklif olunan metod: Təklif olunan metod bir neçə mərhələdən ibarətdir: 1) tədqiq olunan mühitdə dövr edən sənədlərin (informasiyanın) dilindən asılı olaraq, həmin dil üçün terrorizmlə əlaqəli terminlərin lüğət bazasının yaradılması; 2) baxılan dil üçün sözlərin semantik şəbəkəsinin yaradılması (metodun dəqiqliyi bu şəbəkədən çox asılıdır); 3) sözlərin morfoloji təhlili; 4) lügət bazasından istifadə etməklə sənədlərin ilkin filtrasiyası; 5) sözlər arasında semantik yaxınlığın hesablanması; 6) cümlələr arasında semantik yaxınlığın müəyyən edilməsi; 7) sənədlər arasında semantik yaxınlığın müəyyən edilməsi; 8) sənədin əvvəlcədən məlum olan siniflərdən birinə aid edilməsi (təsnifatlandırma) [23].

Tutaq ki, tədqiq olunan mühitin dili üçün baxılan mövzu (terrorizm) ilə bağlı lüğət bazası (VBase) yaradılmış və sözlərin semantik şəbəkəsi qurulmuşdur (ingilis dilində yaradılmış şəbəkəyə oxşar olaraq bu şəbəkəni WordNet ilə işarə edək). Qeyd etmək lazımdır ki, bu biliklər bazası sözlər arasındakı semantik münasibətləri müəyyən etməyə imkan verir. Məsələn, bu şəbəkənin köməyilə sinonimləri, hipernimləri, hiponimləri və s. asanlıqla tapmaq mümkündür (Şəkil 2.1.1).



Şəkil 2.1.1 Hipernim və hiponimlər

Təklif olunan yanaşmanın hər bir mərhələsi aşağıda ətraflı izah edilir.

1) Sənədlərin ilkin filtrasiyasi: Sənədlərin ilkin filtrasiyası aşağıdakı qaydada həyata keçirilir. Əvvəlcə sənəddən terminlər çıxarılır, onlar morfoloji təhlil edilir (bu sözün başlanğıc formasını tapmaq üçündür, çünki eyni bir söz qəbul etdiyi şəkilçilərdən asılı olaraq müxtəlif formalarda olur) və sənəd sözlər (terminlər) çoxluğu kimi təsvir olunur, $d = (\mathrm{t}_1, \mathrm{t}_2, ..., \mathrm{t}_m)$. Sonra Şimkeviç-Simpson ölçüsündən istifadə edərək VBase bazası ilə $d = (t_1, t_2, ..., t_m)$ çoxluğu arsındakı yaxınlıq hesablanır:

$$\mathrm{sim}_{\mathrm{S-S}}(d, \mathrm{VBase}) = \frac{|d \cap \mathrm{VBase}|}{|d|}, \tag{2.1.1}$$

burada $|A| - A$ çoxluğundakı elementlərin sayıdır.

Əgər $\mathrm{sim}_{\mathrm{S-S}}(d, \mathrm{VBase}) \geq \varepsilon$ olarsa, onda d sənədi şübhəli sənədlər çoxluğuna əlavə edilir və identifikasiya üçün növbəti mərhələyə keçid edilir. Burada ε eksperimental yolla müəyyən edilmiş hədd qiymətidir.

2) Sözlərin semantik yaxınlığı: Sözlər arasındakı semantik yaxınlıq aşağıdakı ardıcıllıqla təyin edilir:

1. İki söz t_1 və t_2 götürülür.
2. WordNet semantik şəbəkəsindən bu sözlərin kökü tapılır.
3. WordNet leksik bazasındann hər bir sözün sinonimləri və onların sayı təyin edilir;
4. WordNet şəbəkəsində istifadə etməklə, t_1 və t_2 sözlərinin ən yaxın ümumi (Least Common Subsume – LCS) kökü tapılır;
5. (2.1.2) və (2.1.3) düsturlarının köməyilə sözlər arasındakı semantik yaxınlıq hesablanır.

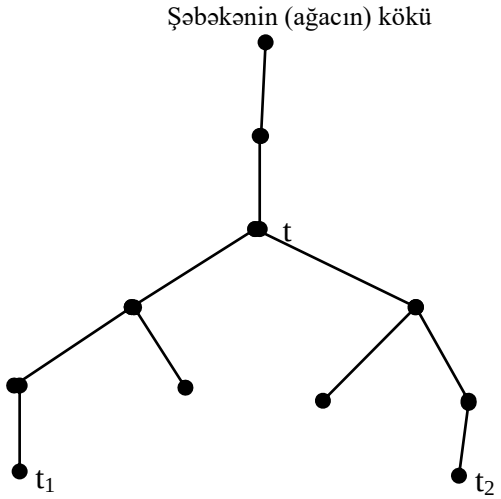
Sözlər arasındakı semantik yaxınlığı hesablamaq üçün əvvəlcə WordNet şəbəkəsindən istifadə etməklə, sözün informativ məzmunu (yükü) $\mathrm{IC}(\mathrm{t})$ təyin edilir [10, s.344]:

$$\mathrm{IC}(\mathrm{t}) = 1 - \frac{\log(\mathrm{synset}(\mathrm{t}) + 1)}{\log(\mathrm{t}_{\max})} \tag{2.1.2}$$

Sonra (2.1.2) düsturundan istifadə edərək sözlər arasındakı semantik yaxınlıq hesablanır [10, s.344]:

$$\mathrm{sim}_{\mathrm{IC}}(t_1, t_2) = \begin{cases} \dfrac{2 * \mathrm{IC}(\mathrm{LCS}(t_1, t_2))}{\mathrm{IC}(t_1) + \mathrm{IC}(t_2)}, & \text{əgər} \quad t_1 \neq t_2 \\ 1, & \text{əgər} \quad t_1 = t_2 \end{cases} \tag{2.1.3}$$

burada $\mathrm{LCS}(t_1, t_2)$ – WordNet şəbəkəsində t_1 və t_2 sözlərinin ən yaxın olduğu ortaq söz (məsələn, şəkil 2.1.2-də göstərilən hal üçün $\mathrm{LCS}(t_1, t_2) = t$), t_{max} – WordNet semantik şəbəkəsindəki sözlərin ümumi sayı, $\mathrm{synset}(t)$ – t sözünün sinonimlərinin sayıdır.



Şəkil 2.1.2 Sözlərin semantik şəbəkəsi

Sözlər arasındakı semantik yaxınlığı həm də WUP metrikasından [171, s.136] istifadə etməklə hesablayırıq:

$$\mathrm{sim}_{\mathrm{WUP}}(t_1, t_2) = \frac{2 * \mathrm{depth}(t)}{\mathrm{depth}(t_1) + \mathrm{depth}(t_2) + 2 * \mathrm{depth}(t)} \tag{2.1.4}$$

burada $\mathrm{depth}(t_1)$ – WordNet semantik şəbəkəsində (ağacında) t_1-dən t-yə qədər olan qovşaqların sayı; $\mathrm{depth}(t_2)$ – t_2-dən t-yə qədər olan qovşaqların sayı; $\mathrm{depth}(t)$ – t-dən şəbəkənin kökünə qədər olan qovşaqların sayıdır. Məsələn, şəkil 2.1.2-də göstərilən hal üçün $\mathrm{depth}(t_1) = \mathrm{depth}(t_2) = 3$ və $\mathrm{depth}(t) = 2$. Onda

$$\text{sim}_{\text{WUP}}(t_1, t_2) = \frac{2 * 2}{3 + 3 + 2 * 2} = 0{,}4$$

Beləliklə, sözlər arasında semantik yaxınlıq (2.1.3) və (2.1.4) düsturları ilə verilən metrikaların xətti kombinasiyası kimi təyin olunur:

$$\text{sim}(t_1, t_2) = \alpha \times \text{sim}_{\text{IC}}(t_1, t_2) + (1 - \alpha) \times \text{sim}_{\text{WUP}}(t_1, t_2) \qquad (2.1.5)$$

burada $0 \leq \alpha \leq 1$ – çəki əmsalıdır.

3) Cümlələrin yaxınlıq ölçüsü: Cümlələr arasındakı yaxınlığı hesablamaq üçün 3 metrikadan istifadə olunacaqdır: semantik, kosinus və sintaktik.

Semantik yaxınlıq. Cümlələr arasındakı semantik yaxınlıq sözlər arasındakı semantik yaxınlıqdan (2.1.5) istifadə edilərək hesablanır:

$$\text{sim}_{\text{semantic}}(s_1, s_2) = \frac{\sum_{t_1 \in s_1, t_2 \in s_2} \text{sim}(t_1, t_2)}{m_1 + m_2} \qquad (2.1.6)$$

burada m_1 və m_2 uyğun olaraq s_1 və s_2 cümlələrindəki sözlərin sayıdır.

Kosinus metrikası. Kosinus metrikası vektor modelinə əsaslanan metrikadır. Vektor modelinə əsasən cümlələr arasındakı yaxınlığı hesablamaq üçün əvvəlcə onların hər biri vektor şəklində təsvir olunur, sonra isə iki vektor arasındakı məsafə (yaxınlıq) hesablanır. Tutaq ki, s_1 və s_2 cümlələri verilmişdir. Ənənəvi yanaşmalarda cümlələri vektor şəklində təsvir edərkən, vektorun uzunluğu sənəddə (yaxud sənədlər çoxluğunda) rast gəlinən sözlərin sayına bərabər götürülür. Aydındır ki, bu cür təsvir zamanı vektorun uzunluğu cümlənin uzunluğundan (cümlədəki sözlərin sayından) dəfələrlə böyük olur və deməli, vektorun elementlərinin böyük əksəriyyəti 0-a bərabər olur. Bu isə hesablama baxımından effektiv təsvir üsulu deyil. Ona görə də burada iki cümlə arasındakı yaxınlığı hesablayarkən, sözlər çoxluğu yalnız bu cümlələrdə rast gəlinən müxtəlif sözlərdən yaradılır. $WS = \{t_1, t_2, ..., t_m\}$ ilə bu sözlər çoxluğunu işarə edək, burada m müxtəlif sözlərin ümumi sayıdır. İki cümlədəki sözlər çoxluğu aşağıdakı ardıcıllıqla yaradılır [10, 178]:

1. İki cümlə götürülür, s_1 və s_2.

2. s_1 cümləsindən götürülmüş hər bir t sözü üçün aşağıdakı işlər apaırılır:

2.1.WordNet leksik bazasından istifadə etməklə onun kökü (RW) təyin edilir.

2.2. Əgər RW sözü WS çoxluğunda iştirak edirsə, onda addım 2-yə keçməli və s_1-dən götürülmüş növbəti söz üçün prosesi davam etdirməli, əks halda 2.3 addımına keçməli;

2.3. Əgər sözün RW kökü sözlər çoxluğunda (WS) iştirak etmirsə, onda RW sözünü WS çoxluğuna əlavə edib, 2-ci addıma keçməli və prosesi s_1-dən götürülmüş növbəti söz üçün davam etdirməli. Proses s_1 cümləsindəki sözlər qurtarana kimi davam etdirilir.

Yuxarıdakı proses s_2 cümləsi üçün də təkrarlanır.

Cümlələr arasındakı yaxınlığı müəyyən etmək üçün semantik vektor modelindən istifadə edilir [16, 22]. Bunun üçün ilkin olaraq aşağıdakı əməliyyatlar yerinə yetirilir:

1. *Vektorun qurulması*. Vektorun hər bir elementi WS sözlər çoxluğundakı sözə uyğundur. Deməli, vektorun ölçüsü WS çoxluğundakı sözlərin sayına bərabərdir.
2. *Vektorun elementlərinin təyini*. Semantik vektorun hər bir elementi (sözün çəkisi) aşağıdakı qayda ilə təyin edilir:

2.1. Əgər WS sözlər çoxluğundan olan t sözü s_1 cümləsində iştirak edirsə, onda bu sözün vektordakı çəkisi 1 götürülür, əks halda növbəti addıma keçilir;

2.2. Əgər t sözü s_1 cümləsində iştirak etmirsə, onda (2.1.5) düsturunun köməyilə t sözü ilə s_1 cümləsindəki sözlər arasındakı yaxınlıq hesablanır.

2.3. Əgər sözlər arasında yaxınlıq sıfırdan fərqlidirsə, onda t sözünün vektordakı çəkisi kimi bu qiymətlərdən ən böyüyü götürülür. Əks halda növbəti addıma keçid edilir;

2.4. Əgər sözlər arasında yaxınlıq sıfra bərabərdirsə, onda t sözünün vektordakı çəkisi 0 götürülür.

Beləliklə, kosinus metrikasından istifadə etməklə, iki vektor arasındakı yaxınlıq aşağıdakı kimi hesablanır:

$$\text{sim}_{\cos}(s_1, s_2) = \frac{\sum_{j=1}^{m}(w_{1j} \times w_{2j})}{\sqrt{\sum_{j=1}^{m} w_{1j}^2} \times \sqrt{\sum_{j=1}^{m} w_{2j}^2}} \quad (2.1.7)$$

burada $s_1 = (w_{11}, w_{12}, ..., w_{1m})$ və $s_2 = (w_{21}, w_{22}, ..., w_{2m})$ – s_1 və s_2 cümlələrinə uyğun semantik vektorlar; w_{pj} – s_p vektorunda t_j sözünün çəkisi; m isə sözlərin ümumi sayıdır.

Sintaktik yaxınlıq. Cümlənin semantik yükü yalnız sözlərin semantik yükü ilə deyil, həm də sözlərin işlənmə ardıcıllığından, yəni sözün cümlədəki mövqeyindən də birbaşa asılıdır. Məsələn, yuxarıdakı yaxınlıq ölçüsünə (semantik yaxınlıq) görə "Əli Həsənə zəng etdi" və "Həsən Əliyə zəng etdi" cümlələri oxşar cümlələr kimi qiymətləndiriliəcəkdir, çünki onlar eyni sözlərdən təşkil olunmuşdur. Buna görə də cümlələrin semantik yaxınlığını hesablayan zaman sözlərin cümlədəki işlənmə ardıcıllığı (onların cümlədəki mövqeyi) da mütləq nəzərə alınmalıdır. Beləliklə, cümlələrin sözlərin cümlədəki rast gəlmə mövqeyinə əsaslanan yaxınlığını hesablamaq üçün sintaktik-vektor yanaşmasından istifadə olunur [105]. Bunun üçün əvvəlcə aşağıdakı əməliyyatlar yerinə yetirilir [10]:

1. Vektorun qurulması. Sintaktik vektorun qurulması üçün WS çoxluğundakı və cümlədəki sözlərdən istifadə edilir. Sintaktik-vektorun uzunluğu WS çoxluğundakı sözlərin sayına bərabərdir.

2. Vektorun elementlərinin təyini. Sintaktik-vektorun hər bir elementi sözün çəkisini ifadə edir və o, sözün cümlədəki mövqeyinə bərabərdir. Bu çəki aşağıdakı kimi müəyyən edilir:

2.1. Əgər t sözü s_1 cümləsində rast gəlinirsə, onda onun vektorda çəkisi cümlədəki mövqeyinə bərabər götürülür, əks halda növbəti addıma keçilir;

2.2. Əgər t sözü s_1 cümləsində rast gəlinmirsə, onda (2.1.5) düsturunun köməyilə s_1 cümləsindəki sözlərlə t sözü arasındakı yaxınlıq hesablanır.

2.3. Əgər sözlər arasında yaxınlıq sıfırdan fərqlidirsə, onda vektorda uyğun elementin qiyməti (sözün çəkisi) olaraq s_1 cümləsində ən böyük çəkiyə malik sözün

mövqeyi götürülür.

2.4. Əgər sözlər arasında yaxınlıq sıfra bərabərdirsə, onda vektorda uyğun elementin qiyməti 0 götürülür.

Cümlələrin, sözlərin cümlədəki mövqeyinə əsaslanan yaxınlığını hesablamaq üçün aşağıdakı düsturdan istifadə olunur [10]:

$$\mathrm{sim}_{\mathrm{wordorder}}(s_1, s_2) = 1 - \frac{\|\mathrm{o}_1 - \mathrm{o}_2\|}{\|\mathrm{o}_1 + \mathrm{o}_2\|} \tag{2.1.8}$$

burada $\mathrm{o}_1 = (w_{11}, w_{12}, ..., w_{1m})$ və $\mathrm{o}_2 = (w_{21}, w_{22}, ..., w_{2m})$ – s_1 və s_2 cümlələrinə uyğun sintaktik-vektorlar; w_{pj} isə o_p vektorunda t_j sözünün çəkisi, $\|\cdot\|$-Evklid məsafəsidir.

Xətti kombinasiya. Cümlələr arasında yaxınlığı hesablamaq üçün semantik, kosinus və sintaktik ölçülərin xətti kombinasiyası istifadə olunur:

$$\mathrm{sim}_{\mathrm{sentences}}(s_1, s_2) = \beta_1 \cdot \mathrm{sim}_{\mathrm{semantic}}(s_1, s_2) + \beta_2 \cdot \mathrm{sim}_{\mathrm{wordorder}}(s_1, s_2) + \beta_3 \cdot \mathrm{sim}_{\cos}(s_1, s_2) \tag{2.1.9}$$

burada β_i $(0 \le \beta_i \le 1,\ i = 1,2,3)$ çəki parametrləridir və aşağıdakı şərti ödəyirlər:

$$\beta_1 + \beta_2 + \beta_3 = 1 \tag{2.1.10}$$

4) Sənədlərin yaxınlıq ölçüsü: Sənədlər arasındakı yaxınlığı hesablamaq üçün cümlələr arasındakı yaxınlıqdan (2.1.9) istifadə olunur:

$$\mathrm{sim}_{\mathrm{documents}}(d_1, d_2) = \frac{\sum_{s_1 \in d_1, s_2 \in d_2} \mathrm{sim}_{\mathrm{sentences}}(s_1, s_2)}{n_1 + n_2} \tag{2.1.11}$$

burada n_1 və n_2 uyğun olaraq d_1 və d_2 sənədlərindəki cümlələrin sayıdır.

Sadəlik üçün aşağıda $\mathrm{sim}_{\mathrm{documents}}(d_1, d_2)$ əvəzinə $\mathrm{sim}(d_1, d_2)$ yazılışından istifadə ediləcəkdir.

5) Sənədlərin təsnifatlandırılması: Tutaq ki, $\mathbf{C} = (C_1, ..., C_k)$ siniflər çoxluğu məlumdur. Verilmiş $\mathbf{D} = (d_1, ..., d_N)$ sənədlər çoxluğunun bu siniflərdən $\mathbf{C} = (C_1, ..., C_k)$ hər hansı birinə (yaxud bir neçəsinə) aid edilməsi prosesinə sənədlərin təsnifatı deyilir. Bu halda sənəd siniflərdən hansına daha çox yaxındırsa, onda o, həmin sinif(lər)ə aid edilir. Ədəbiyyatda kifayət qədər təsnifatlandırma metodları təklif edilmişdir. Burada d_i sənədinin C_q sinfinə aid olma dərəcəsini

müəyyən etmək üçün *k*NN (k-Nearest Neighbor – k-ən yaxın qonşu) [10], Bayes [56, s.9-12] və yeni təklif olunan RG metodundan [23] istifadə olunur.

***k*NN metodu**. Bu metoda görə d_i sənədinin C_q sinfinə aid olması aşağıdakı düsturla hesablanmış kəmiyyətin qiyməti ilə müəyyən edilir:

$$\text{score}_{k\text{NN}}(d_i \mid C_q) = \sum_{d \in k\text{NN}_q(d_i)} \text{sim}(d_i, d), \quad i = 1,2,...,N; \quad q = 1,2,...,k \qquad (2.1.12)$$

burada $k\text{NN}_q(d_i)$ – C_q sinfində d_i sənədinə ən yaxın olan k sayda sənədlər çoxluğudur.

d_i sənədi ən böyük $\text{score}_{k\text{NN}}(d_i \mid C_q)$ qiymətinə malik sinfə aid edilir, başqa sözlə $d_i \in C_{k^*}$, əgər $k^* = \arg\max_q \text{score}_{k\text{NN}}(d_i \mid C_q)$.

Modifikasiya olunmuş Bayes metodu. Bayes metoduna görə d_i sənədinin C_q sinfinə aid olma dərəcəsi aşağıdakı şərti ehtimalın qiyməti ilə müəyyən olunur:

$$\text{P}(C_q \mid d_i) = \frac{\text{P}(C_q)\text{P}(d_i \mid C_q)}{\text{P}(d_i)} \qquad (2.1.13)$$

Təsnifatlandırma prosesində $\text{P}(d_i)$ sabit qaldığından, (2.1.13) ifadəsində kəsrin məxrəcini nəzərə almamaq olar.

burada $\text{P}(C_q)$ – sənədlərin C_q sinfində olma ehtimalıdır. Bu aprior ehtimal aşağıdakı kimi təyin edilir:

$$\text{P}(C_q) = \frac{\sum_{i=1}^{N} \text{P}(C_q \mid d_i)}{N} \qquad (2.1.14)$$

$\text{P}(d_i \mid C_q)$ kəmiyyətini hesablamaq üçün fərz olunur ki, sözlərin sənədlərdə işlənməsi bir-birindən asılı deyildir. Onda $\text{P}(d_i \mid C_q)$ ehtimalı aşağıdakı düsturun köməyilə hesablanır:

$$\text{P}(C_q \mid d_i) = \text{P}(C_q) \prod_{j=1}^{m} \left(\text{P}(\text{t}_j \mid C_q)\right)^{w_{ij}} \qquad (2.1.15)$$

burada $\text{P}(\text{t}_j \mid C_q)$ – t_j sözünün C_q sinfində olma ehtimalı, m – **D** sənədlər çoxluğundakı sözlərin sayıdır:

$\mathrm{P}(\mathrm{t}_j|C_q)$ ehtimalı aşağıdakı kimi hesablanır:

$$\mathrm{P}(\mathrm{t}_j|C_q) = \frac{\sum_{d_i \in C_q} w_{ij}}{\sum_{j=1}^{m} \sum_{d_i \in C_q} w_{ij}}, \ j = 1,..., m \qquad (2.1.16)$$

burada $w_{ij} - \mathrm{t}_j$ sözünün d_i sənədində çəkisidir.

Əgər t_j sözünün C_q sinfində çəkisi sıfra bərabər olarsa, onda (2.1.16) düsturundan alınır ki, $\mathrm{P}(\mathrm{t}_j|C_q)$ ehtimalı da sıfra bərabər olacaqdır. Beləliklə, (2.1.15) düsturundan asanlıqla alınır ki, bu hal üçün $\mathrm{P}(C_q|d_i)$ ehtimalı da sıfra bərabər olacaqdır. Ona görə də praktikada aşağıdakı düsturdan istifadə olunur:

$$\mathrm{P}(\mathrm{t}_j|C_q) = \frac{1 + \sum_{d_i \in C_q} w_{ij}}{m + \sum_{j=1}^{m} \sum_{d_i \in C_q} w_{ij}}, \ j = 1,..., m \qquad (2.1.17)$$

(2.1.15) düsturunda loqarifmik miqyasa keçib, nəticəni d_i sənədindəki sözlərin ümumi çəkisinə (w_i) bölsək aşağıdakı bərabərliyi alarıq:

$$\mathrm{score}_{\mathrm{M\,Bayes}}(C_q|d_i) = \mathrm{P}(C_q|d_i) = \frac{\log \mathrm{P}(C_q)}{w_i} + \sum_{j=1}^{m} \mathrm{P}(\mathrm{t}_j, d_i) \log \mathrm{P}(t_j|C_q) \qquad (2.1.18)$$

burada $P(t_j, d_i) = w_{ij} / w_i$ – t_j sözünün d_i sənədində işlənmə ehtimalıdır, $w_i = \sum_{j=1}^{m} w_{ij}$, $i = 1,..., n$; $q = 1,..., k$.

(2.1.18) düsturunda *k*NN metoduna oxşar olaraq $\mathrm{score}_{\mathrm{Bayes}}(C_q|d_i) = \mathrm{P}(C_q|d_i)$ işarələməsi qəbul edilmişdir. Bu modelə görə d_i elə sinfə aid edilir ki, bu sinif üçün $\mathrm{P}(C_q|d_i)$ ehtimalı ən böyük qiymətə malik olsun, $d_i \in C_{k*}$, burada $k^* = \arg \max_{1 \le q \le k} \mathrm{score}_{\mathrm{M\,Bayes}}(C_q|d_i)$.

RG metodu. Bu metodun köməyilə d_i sənədinin C_q sinfinə aid olma dərəcəsi aşağıdakı düstürla müəyyən edir:

$$\text{score}_{\text{RG}}(d_i \mid C_q) = \lambda \times \frac{\text{sim}(\text{O}_{d_i}, \text{O}_{C_q})}{\sum_{p=1}^{k} \text{sim}(\text{O}_{d_i}, \text{O}_{C_p})} + (1-\lambda) \times \frac{\sum_{v \in C_q} \text{sim}(\text{O}_{d_i}, \text{O}_v)}{\sum_{p=1}^{k} \sum_{d \in C_p} \text{sim}(\text{O}_{d_i}, \text{O}_d)} \tag{2.1.19}$$

burada $\text{sim}(\text{O}_{d_i}, \text{O}_{C_q})$ – d_i sənədinin O_{d_i} obrazı ilə C_q sinfinin O_{C_q} obrazı arasında yaxınlıq ölçüsü; $\text{sim}(\text{O}_d, \text{O}_v)$ – d və v sənədlərinin O_d və O_v obrazları arasındakı yaxınlıq ölçüsü; λ isə çəki əmsalıdır, $0 \le \lambda \le 1$.

O_{C_q} obrazı C_q sinfinin mərkəzi kimi təyin edilir, $\text{O}_{C_q} = (w_1^q, w_2^q, ..., w_m^q)$:

$$w_j^q = \frac{1}{|C_q|} \sum_{d \in C_q} w_j^{q,d}, \; q = 1,..., k, \; j = 1,..., m \tag{2.1.20}$$

burada $|C_q|$ – C_q sinfindəki sənədlərin sayını, $w_j^{q,d}$ isə C_q sinfinə daxil olan d sənədindəki j-ci sözün çəkisini göstərir.

Analoji qayda ilə O_d obrazı d sənədinin mərkəzi kimi təyin edilir, $\text{O}_d = (w_1^d, w_2^d, ..., w_m^d)$:

$$w_j^d = \frac{1}{|d|} \sum_{s \in d} w_j^{d,s}, \; j = 1,..., m \tag{2.1.21}$$

burada $|d|$ – d sənədindəki cümlələrin sayı, $w_j^{d,s}$ isə d sənədinə aid olan s cümləsindəki j-ci sözün çəkisidir.

Hibrid metod. Yekun təsnifatlandırma metodu kimi (2.1.12), (2.1.18) və (2.1.19) düsturlarının köməyilə alınmış nəticələrin xətti kombinasiyasından istifadə olunması təklif edilir:

$$\begin{aligned} &\text{score}(d^{\text{new}} \mid C_q) = \gamma_1 \cdot \text{score}_{k\text{NN}}(d^{\text{new}} \mid C_q) + \gamma_2 \cdot \text{score}_{\text{Bayes}}(d^{\text{new}} \mid C_q) \\ &+ \gamma_3 \cdot \text{score}_{\text{RG}}(d^{\text{new}} \mid C_q) \end{aligned} \tag{2.1.22}$$

burada $0 \le \gamma_i \le 1, \; (i = 1,2,3)$ çəki əmsallarıdır və aşağıdakı şərti ödəyirlər:

$$\gamma_1 + \gamma_2 + \gamma_3 = 1. \tag{2.1.23}$$

Beləliklə, d^{new} sənədi elə C_{k^*} sinfinə aid edilir ki, bu sinif üçün o ən böyük $\text{score}(d^{\text{new}} \mid C_q)$ qiymətinə malik olsun, $d^{\text{new}} \in C_{k^*}$, burada.

6) Qiymətləndirmə: Təsnifatı qiymətləndirmək üçün dəqiqlik (accuracy), həssaslıq (precision), tamlıq (recall) və F-ölçü (F-measure) meyarlarından istifadə olunur: $k^* = \arg\max_q \text{score}(d^{\text{new}} \mid C_q)$

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{2.1.24}$$

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{2.1.25}$$

$$\text{Recall} = \frac{T_p}{T_p + F_n} \tag{2.1.26}$$

$$\text{F} - \text{measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.1.27}$$

burada T_p – doğru təsnif edilmiş terrorizmlə əlaqəli sənədlərin sayı; F_p – səhv təsnif edilmiş terrorizmlə əlaqəli sənədlərin sayı; T_n – doğru kimi təsnif edilmiş terrorizmlə əlaqəli olmayan sənədlərin sayı; F_p – səhv kimi təsnif edilmiş terrorizmlə əlaqəli olmayan sənədlərin sayıdır.

Beləliklə, e-dövlət mühitində terrorizmlə əlaqəli sənədlərin aşkarlanması üçün text mining texnologiyası metodlarına əsaslanan kompleks yanaşmanın hər bir mərhələsi izah edildi. Lakin burada öz həllini gözləyən bir neçə problem var:

- ✓ tədqiq olunan dil üçün semantik şəbəkənin qurulması;
- ✓ kriminal qrupun kommunikasiyada xüsusi jarqon sözlərdən istifadə etməsi;
- ✓ sözlərin qrammatik cəhətdən düzgün yazılışı;
- ✓ mətnlərin müxtəlif dillərdə olması;
- ✓ mətndən terminlərin çıxarılması;
- ✓ təsnifat metodunun seçilməsi və dəqiqliyi.

Bütün bunlar təklif olunan metodun dəqiqliyinə birbaşa təsir edən amillərdir. Digər mühüm məsələ, siniflərin (mövzular) əvvəlcədən məlum olması və hər yeni

sənədin bu sinflərdən birinə aid edilə bilməsi fərziyyəsidir. Lakin bu həmişə belə olmur və siniflər dinamikidir. Zaman keçdikcə yeni mövzular meydana çıxa bilər və bu mümkün haldır. Ona görə də burada ən düzgün yanaşma sənədləri avtomatik qruplaşdırıb, sonra identifikasiya etməkdir. Bütün bunlar onu göstərir ki, gələcək tədqiqatlar üçün kifayət qədər ciddi problemlər var. Bu problemlərin bir çoxu (məsələn, sözlərin avtomatik morfoloji təhlili üçün qaydaların yaradılması, sözlərin semantik şəbəkəsinin yaradılması) mutidissiplinar xarakterlidir.

2.2. E-dövlətdə terrorizmi təbliğ edən mətnlərin aşkarlanması üçün metod

Dövlət orqanları arasında informasiya mübadiləsinin aparılması e-dövlət arxitekturasının inkişafı üçün ən vacib məsələlərdən biri hesab olunur. Bu cür informasiya mübadiləsi vasitəsilə dövlət orqanları terror təhlükəsi, təhlükəsizlik riskləri və hücumlarını təxmin edə bilərlər. Dövlət orqanları arasında şəbəkə əlaqələri daha sürətli məlumat mübadiləsini təmin edir. Bu müxtəlif dövlət orqanlarının terror fəaliyyəti haqqında məlumatı eyni zamanda ala bilməsi üçün vacibdir. Dövlət müəssisəsinin əməkdaşları vətəni, hökuməti və vətəndaşları ənənəvi və kiber terror hücumlarından qorumaq üçün təhlükəsizlik tədbirləri tətbiq etmək məqsədilə bu məlumatlardan istifadə edə bilərlər. Ənənəvi terror hücumları zamanı polis orqanları şübhəli şəxsləri izləmək və hücumların qarşısını almaq məqsədilə əvvəl baş vermiş terror fəaliyyəti haqqında və biometrik sənədlərdən toplanan məlumatlara əsaslana bilər. Həmçinin fərdləri, məkanları və danışıqları izləmək üçün mobil telefon şəbəkəsindən istifadə edilə bilər.

İnsanları terrorizmə nəyin yönəltdiyini müəyyən etmək asan deyil. Bu gün terroristlərin mülkiyyəti hesab olunan kritik veb saytlardan istifadə edən istifadəçilərin monitorinqi faydalı məlumatlara səbəb ola bilər. Müxtəlif terror təşkilatlarından hücumların qarşısını almaq üçün dövlət orqanları öz aralarında müvafiq məlumat mübadiləsi aparmalı və əməkdaşlıq etməlidirlər. Ədəbiyyatda terrorist veb saytlarının və sosial şəbəkə profillərinin aşkarlanması, monitorinqi və bloklanması terrorizmlə mübarizədə üç ən əhəmiyyətli fəaliyyət hesab olunur [154,161]. Burada ilk növbədə mümkün terrorist veb sayt və ya istifadəçi profili

tapılır. Daha sonra monitorinq prosesində orqanlar tərəfindən veb saytın sahibi və istifadəçiləri haqqında bütün vacib məlumatlar toplanılır. Bu cür məlumatlar vasitəsilə terrorist siyahısına bir neçə internet istifadəçisi daxil oluna bilər. Sonda veb sayt bloklanır. İnternet xidmət provayderləri bu iş üçün hökumətin sifarişi ilə məsuliyyət daşıyırlar.

Hər bir ölkənin təhlükəsizliyi bir çox faktlardan asılıdır. Terrorizmə qarşı qanun müxtəlif ölkələrdə fərqlidir, lakin eyni fikirləri hədəfləyir. Bu qanun dövlət orqanlarının hüquq və vəzifələrini müəyyənləşdirir, hansı hallarda bir şəxsin və ya təşkilatın terrorist təhdidi kimi qeyd oluna biləcəyini müəyyən edir. Dövlət orqanları vətəndaşlar üçün müvafiq təhlükəsizlik səviyyəsini təmin etmək üçün terrorla mübarizədə ən yaxşı həll yolunu tapmağa çalışırlar. Mümkün terrorist metodlarının çoxluğu səbəbiylə təşkilatlar bu mübarizədə birlikdə çalışmalıdırlar. Müxtəlif dövlət xidmətləri arasında şəbəkə əlaqələri, elektron monitorinq və şübhəli veb saytlarda olan fərdlərin yoxlanılması gələcək hücumların qarşısını ala bilər.

Terrorizm və internet: Günbəgün internetdən istifadənin artması ilə əlaqədar olaraq müxtəlif dinamiki platformalarda böyük miqdarda yeni texnologiyalar meydana çıxır və bu da bədniyyətli insanların cəmiyyət və insanlara zərər vermək üçün istifadə etməsinə şərait yaradır. Belə ki, internetə geniş rəqəmsal kitabxana kimi baxılsa, burada milyard səhifəlik əlçatan informasiyaların mövcud olduğunu görərik və bunların bir çoxu terror təşkilatlarının marağındadır. Məsələn, terroristlər nəqliyyat vasitələri, nüvə elektrik stansiyaları, ictimai binalar, aeroportlar və limanlar, hətta antiterror tədbirləri barədə hədəflər haqqında internetdən geniş məlumat ala bilərlər.

Onlayn platformalardan terroristlərin istifadəsi yeni deyil. Belə ki, 11 sentyabr hadisələrindən və bununla əlaqədar antiterror kampaniyasından sonra, çox sayda terrorçu qrup mesajlarını və fəaliyyətlərini təbliğ edən minlərlə veb sayt quraraq, kiber fəzaya keçdi. Bir çox terrorist saytları kəşfiyyat və hüquq-mühafizə orqanları, antiterrorist xidmətləri tərəfindən hədəfləndi.

Terroristlərin sosial mediadan istifadə etməsində bir neçə əsas səbəb mövcuddur. Bunlardan birincisi, bu kanalların öz əhatə dairəsində məşhur olması və

terrorisitlərə əsas kütlənin bir parçası olmağa icazə verməsidir. İkincisi, sosial media kanallarının istifadəçilərin dostu, etibarlı və ödənişsiz olmasıdır. Nəhayət, sosial şəbəkələrin terroristlərə öz hədəf kütləsinə çatmağına icazə verməsidir.

Terrorist qruplarının əksəriyyəti, xüsusilə əl-Qaidə kimi "dünya markaları" terrorizmi təbliğ etmək və yaymaq üçün internetdən istifadə edirlər [157]. Kriminal qruplar, xarici kəşfiyyat xidmətləri e-oğurluq və təxribat qabiliyyətlərini internet vasitəsilə nümayiş etdirirlər. Terroristlərin onlayn mühitdə ən mühüm məqsədi təbliğatdır. Terror təşkilatları və onların tərəfdarları internetin tənzimlənməmiş, anonim və asanlıqla istifadə edilə bilən təbiətindən istifadə edərək yüzlərlə veb saytı əldə saxlayaraq, bir sıra auditoriyaya mesajlarını çatdırırlar. Bundan əlavə, onlar onlayn qruplar vasitəsilə potensial insanların siyahısını tərtib edə bilərlər. Belə ki, terrorçu qruplar insanların profillərini izləyərək, kimi hədəf alacaqlarını və hər bir fərdə necə yaxınlaşacaqlarını müəyyən edirlər. Bu gün demək olar ki, hər bir terror təşkilatı bir çox müxtəlif dillərdə yaradılan veb saytlara malikdir [169]. Terror təşkilatları səmimi olmaq üçün şüarlardan (videoyazılar, audiokassetlər, futbolkalar, nişanlar və bayraqlar) istifadə edərək lokal (yerli) dildə olan veb saytlar vasitəsilə yerli tərəfdarları hədəf alırlar.

İnternet bir çox cəhətdən terror təşkilatlarının fəaliyyəti üçün ideal məkandır. Bunlar əsasən aşağıdakılardan ibarətdir [165]:

- Asan giriş;
- Dövlət nəzarətinin az olması;
- Bütün dünyaya yayılan böyük auditoriya;
- Kommunikasıyanın anonimliyi;
- Sürətli məlumat axını;
- Ucuz başa gəlməsi;
- Multimedia mühiti (mətn, qrafika, audio və video; istifadəçilərə film, mahnı, kitab, plakatlar və s. yükləmək imkanının verilməsi);
- Ənənəvi kütləvi informasiya vasitələrində əhatə dairəsini formalaşdırmaq qabiliyyəti.

Tədqiqatçılar terroristlərin internetdən istifadə etməsini bir neçə amillə bağlayır [51]. Cədvəl 2.2.1-də bunlar əks olunmuşdur.

Cədvəl 2.2.1

Furnell & Warren (1999) [62]	Cohen (2002) [61]	Thomas (2003) [157]	Weimann (2004) [165]
▪ Təbliğat & Təqdimat ▪ Maliyyə ▪ İnformasiya yayımı ▪ Təhlükəsiz rabitə	▪ Planlaşdırma ▪ Maliyyə ▪ Koordinasiya və əməliyyatlar ▪ Siyasi hərəkat ▪ Təbliğat	▪ Profil yaratmaq ▪ Təbliğat ▪ Anonim/Gizli Əlaqə ▪ "Kiber qorxu" yaratmaq ▪ Maliyyə ▪ Komanda və nəzarət ▪ Səfərbərlik və işəgötürmə ▪ Məlumat toplanması ▪ Risklərin azaldılması ▪ Oğurluq/ verilənlərin manipulyasiya ▪ Yanlış məlumat	▪ Psixoloji müharibə ▪ Təqdimat və təbliğat ▪ Data Mining ▪ -Maliyyə ▪ İşə qəbul və səfərbərlik ▪ Şəbəkələrin yaradılması ▪ Məlumatların ötürülməsi ▪ Planlaşdırma və koordinasiya

Göstərilən amillərdən bəziləri aşağıda ətraflı şəkildə qeyd olunmuşdur.

Məlumat toplanması: İnternet terrorist qrupları tərəfindən istifadə oluna biləcək məlumatlarla doludur. Belə qruplar müxtəlif xəritələr, peyk fotoşəkilləri və planları tapa bilərlər. Bununla yanaşı, terrorçular rabitə və nəqliyyat infrastrukturu, su təchizatı sistemləri, partlayıcı maddələrin istehsalı, saxta pasportların yaradılması və müxtəlif silahlarla bağlı məlumatların əldə edilməsi üçün internetdən istifadə edə bilərlər. Terrorist qruplar öz aralarında müxtəlif növ məlumatları bölüşürlər. Onlar yeni xəbərləri, logistika və taktiki məlumatları paylaşırlar. Əksər hallarda onlar parolla qorunmuş forumlara, çatlara və məlumat lövhələrinə güvənirlər. Son zamanlar bir sıra geniş miqyaslı terrorçu qruplar daha az mərkəzləşmiş və daha geniş şəkildə şəbəkəyə çevrilmişdir. Onlar ünsiyyət qurmaq üçün sosial şəbəkələrdən istifadə edirlər [144].

Təbliğat: Terror təşkilatları üçün yeni fəalları terrorizmə cəlb etmək çox əhəmiyyətlidir. Bir çox hallarda bu iş üçün siyasi və ya dini ritorika istifadə olunur. İnternet istifadəçilərinin əksəriyyətini gənclər təşkil edir və yaş hədləri nəzərə alınaraq, marketinq və təbliğat əsasən gənclərə yönəldilir.

İşə qəbul və səfərbərlik: Terror təşkilatlarının veb saytları tez-tez ziyarətçilərə müşahidə edilən və tədqiq edilən digər ünvana yönəldən linklər təklif edir. Daha çox səhifəni ziyarət edən və daha uzun müddət veb saytda qalanlarla əlaqə saxlanılır.

"Kiber qorxu" yaratmaq: Burada əsas diqqət gizlilik məlumatlarının açıqlanması və ya istifadəsi məqsədilə zərər çəkmiş kompüterlərə hücuma yönəlib. Potensial hədəflər telekommunikasiya sistemləri, müdafiə sistemləri, tibb müəssisələri, enerji şəbəkələri, nəqliyyat, ictimai şəxslər və siyasətçilərdir.

İnternetdə terroristlərin fəaliyyətinin aşkarlanması ilə terror hücumlarının qarşısı alına bilər. 2001-ci ildən etibarən müxtəlif dövlətlər və kəşfiyyat orqanları gələcəkdə terror hadisələrinin qarşısını almaq üçün internetdə terror fəaliyyətini müəyyən etməyə çalışırlar [76]. İnternetdə terror fəaliyyətini aşkarlamanın bir yolu internet-protokolunun ünvanı (IP) əsasında istifadəçiləri aşkar etmək üçün terror qrupları və təşkilatları ilə əlaqəli veb saytların bütün trafiklərinin izlənilməsidir [58]. Təəssüf ki, bu saytlardan terror qrupları sabit olaraq istifadə etmədikləri üçün terrorist saytları aşkar etmək çətindir. Terroristlər onların müəyyən olunmasını daha da çətinləşdirmək məqsədilə tez-tez dəyişən müxtəlif IP ünvanlarından istifadə edirlər. Həmçinin terrorizmi təbliğ edən saytların əksəriyyəti onların hökumət tərəfindən trafik izləmələrinin qarşısını almaq üçün mütəmadi olaraq URL-lərini dəyişirlər. Göründüyü kimi, spesifik IP-dən istifadə edərək hər bir istifadəçinin şəxsiyyətini bildiyi İnternet-xidmət provayderinin (ISP) və ya sistem administratorlarının təşkilat şəbəkəsinin trafikini izləməklə terrorla əlaqəli saytlara çıxışın aşkar edilməsi üçün yeni üsullar və texnologiyaların işlənilməsinə ehtiyac var. Bir çox tədqiqatçı və elm adamları bu problemi həll etmək üçün geniş tədqiqatlar aparmışdır.

Terrorizm və e-dövlət: E-dövlətin praktiki cəhətdən dinamik xarakterli olması ilə əlaqədar olaraq onun müxtəlif təsvirləri mövcuddur. Aşağıdakı 2 təsvir onun məzmunu haqqında ümumi fikir yaradır. İlk olaraq e-dövlət dedikdə proses və

əlaqələr nəzərə alınır. Belə ki, e-dövlət texnologiya, internet və yeni media vasitəsi ilə daxili və xarici əlaqələri dəyişdirərək, xidmətin çatdırılması və idarəetmə proseslərinin davamlı optimallaşdırılmasıdır. İkinci tərif, e-dövləti qarşılıqlı əlaqəni təşkil etmə üsulu kimi qiymətləndirir. Burada e-dövlət dedikdə müasir informasiya və kommunikasiya texnologiyalarının tətbiqi ilə hökumət və vətəndaşlar, şirkətlər, müştərilər və dövlət qurumları arasında qarşılıqlı əlaqələri birləşdirən təşkilat forması nəzərdə tutulur. Texnologiya, dövlət qurumları və tərəflər arasında mübadilələr hər iki tərif üçün ortaq cəhətlərdəndir.

E-dövlət milli təhlükəsizliklə bağlı məsələlərdə mühüm rol oynayır. Hökumətin, vətəndaşların və özəl sektorun qorunmasında e-dövlətin istifadə oluna biləcək xüsusiyyətlərindən biri informasiya mübadiləsidir. Hökumət daxilində və digər şəxslər arasında informasiya mübadiləsi e-dövlət kanalları və texnologiyaları ilə məhdudlaşmır, lakin bu resurs informasiya mübadiləsinin vacib bir vasitəsidir. Belə ki, e-dövlət resurslarından istifadə etməklə, hökumət rəsmiləri, məsələn, məlumatları real vaxtda ötürə; eyni zamanda çoxsaylı ofislərə məlumat verə; verilənlər bazalarını müqayisə edə və ya birləşdirə, həmçinin görünən fərqli məlumat hissələrini inteqrasiya edə bilərlər. Dövlətin terrorla mübarizəsi e-dövlətin bir çox istiqamətinin anlaşılmasını və qiymətləndirmə qabiliyyətini artırmışdır. E-dövlət terrorizmlə mübarizədə mənbə kimi istifadə edilir və hücumların qarşısının alınmasına kömək edir. Bu həmçinin hücumlardan sonra bərpa vaxtı faydalı ola bilər. E-dövlət hökumət tərəfindən terror hücumlarının qarşısının alınması, terrorizmə qarşı "zəifliyi" azaltmaq, zərərləri minimuma endirmək və hücumlardan qurtulmaq üçün birgə milli səy kimi müəyyən olunur [66].

Bütün bunlarla yanaşı təəssüf ki, e-dövlətin özü də terrorçuların hədəfi ola bilər. Terroristlər, potensial hədəfləri tapmaq, zəiflikləri müəyyən etmək və ya istismar etmək, fərqli məlumatları əldə etmək və inteqrasiya etmək üçün hökumət veb saytlarının məzmunundan istifadə edə bilərlər. Dövlət veb saytlarına kiber hücumlar və infrastrukturun (kompüter sistemi, güc təmin edən elektrik şəbəkəsinin) zədələnməsi və ya məhv edilməsi e-dövlətə zərər verə bilər [8]. E-dövlət həmçinin informasiyalarla potensial bir hədəf kimi cəlbedicidir. Belə ki, burada mövcud olan

informasiyalardan terroristlər hücumların planlanması və zəif cəhətlərin müəyyənləşdirilməsində istifadə edə bilərlər. E-dövlətə olan kiber təhlükələr aşağıdakılardır [68]:

İnsayderlər: narazı insayderlər (keçmiş əməkdaş) kompüter cinayətinin əsas mənbəyidir;

Haker: Hakerlər qərəzli məqsədlər və maliyyə mənfəəti üçün şəbəkələrə hücum edirlər;

Haktivizm: siyasi mesajlar göndərmək, saxta görüntü yaratmaq və ya məlumat vermək, mətni dəyişdirmək üçün veb səhifələrə və ya e-poçt xidmətlərinə siyasi əsaslı hücumlardır;

Virus yazıçıları: dağıdıcı kompüter viruslarının yayılması dünyada şəbəkələr və proqram təminatları üçün getdikcə ciddi təhlükə yaradır;

Kriminal qruplar: cinayətkarlar tərəfindən mənfəət əldə etmək üçün sistemlərə hücum cəhdləridir;

Terroristlər: bu qruplar informasiya texnologiyaları və internetdən planlar hazırlamaq, pul vəsaitlərini artırmaq, təbliğat yaymaq və təhlükəsiz ünsiyyət qurmaq üçün istifadə edirlər;

İnformasiya müharibəsi: xarici qüvvələrin kritik infrastrukturlara qarşı "informasiya müharibəsi" milli təhlükəsizliyə ciddi təhdiddir.

E-dövlətin qorunması hökumətin prioritetlərindən biri olmuşdur. E-dövlətə olan bu cür hücumların qarşısının alınması və vaxtında müəyyən olunması üçün yeni metod və alqoritmlərin işlənilməsinə ehtiyac vardır.

Terrorizm və big data analika: Terrorizm dünyanın bir çox yerlərinə kök salmışdır və demək olar ki, gündəlik həyatımızın “bir hissəsinə” çevrilmişdir. Belə ki, hər gün dünyanın müxtəlif bölgələrində terror hadisələri baş verir. İnternetdəki çıxışlar və videolar terrorizmin yayılmasının əsas mənbəyi hesab olunur. Terror təşkilatları internetdən fərdlərin beynini "yumaq" və “zəif” insanları veb səhifələr vasitəsilə terror fəaliyyətinə qatılmağa təşviq etmək üçün istifadə edirlər.

Terror qrupları hədəf aldıqları ölkə/region, ideologiya və hücumun tipinə görə müxtəlif davranış qanunauyğunluqlarına malikdirlər. Terror hücumunun növü

terrorist ideologiyası və gələcək cəhdlərin predmeti ilə bağlıdır. Hücum növləri aşağıdakı kimi qruplaşdırılır [147]:

Sui-qəsdlər: əsas məqsəd bir və ya daha çox əhəmiyyətli insanları öldürməkdir;

Silahlı hücumlar: ümumiyyətlə silah kimi güclü odlu silahlarla və ya bıçaq, fərdi istifadəsi olan partlayıcılar, bombalar və s. ilə törədilən hücumlardır;

Qaçırma: təyyarə, avtobus, qatar, gəmi, vertolyot, ümumiyyətlə, hərəkət edə bilən vasitələrin idarəsinin ələ keçirilməsidir;

Bombalar, partlayışlar: partlayıcı və ya kimyəvi maddələr və qazlardan yaradılan partlayışlardır;

Adam oğurluğu: pul əldə etmək və ya siyasi imtiyazlar məqsədilə insanların girov götürülməsidir;

İnfrastruktur hücumları: yol, körpü, tikinti, qatar, neft / qaz boru kəməri, kanal, qatar və ya qatar yolları, yük maşını, neft tankeri, anbar, su anbarı və s. kimi hədəflərə hücumlardır.

1970-cı ildən bəri qlobal miqyasda təxminən 125.000 tanınmış terror hadisəsi baş vermişdir [148, 132]. Hazırkı dövrdə terror fəaliyyətinin artması ilə yanaşı, terrorizmin qarşısını almaq və yayılmasını dayandırmaq vacib məsələlərdən birinə çevrilmişdir. Belə ki, Pakistan (150 məktəbli uşağın ölümü), Kanada (Ottava və Kvebekdə hökumətin hədəflənməsi), həmçinin Avstraliyadakı (Sidneydə Lindt Café mühasirəsi) son hücumlar hökumət, biznes və əhalinin qlobal terrorizmdən necə təsirləndiyini, terrorizmlə əlaqəli risklərin necə və harada olduğunun müəyyən olunmasının faydalı olduğunu göstərir [148].

Terrorizm özü ilə birlikdə bəşəriyyətə bərpa edilə bilməyən mənəvi, eyni zamanda maddi zərər vurur. Hazırda terrorizmlə bağlı qarşılaşdığımız təhlükənin "Soyuq Müharibə" təhdidlərindən çox fərqləndiyi aşkardır və bu məlumatların toplanması, təhlili üçün yanaşmalarımıza düzəlişlər edilməsini tələb edir. Soyuq Müharibə illərindən fərqli olaraq, hazırda terrorçular qeyri-iyerarxik quruluşda təşkil edilirlər. Bütün ənənəvi kəşfiyyat toplama metodları əhəmiyyətli olmasına baxmayaraq, terrorçuları anlamaq və hərəkətlərini proqnozlaşdırmaq üçün çox sayda

məlumatları analiz etməyə ehtiyac yaranır. Soyuq müharibədə mühüm məsələlərdən biri vacib informasiyaların gizli olaraq ələ keçirilməsi və onların qorunmasından ibarət idisə, indiki dövrdə bu cür informasiyaları tapmaq daha da çətinləşib. 11 sentyabr 2001-ci ildə olan hücumlar bu vəziyyəti aşkar göstərir.

Hazırda terroristlərin əməliyyatları haqqında toplanılan məlumatlar böyük həcmlidir və getdikcə çoxölçülü hala gəlməkdədir. Terrorist qrup davranışlarının faktorlarının anlaşılmasının mürəkkəb olması, məlumatların yüksək sürətdə artması və sübutların dünya miqyasında dağınıq olması səbəbiylə bu big data analitikası məsələlərindən birinə çevrilmişdir. Big data analitika terrora qarşı mübarizə üsulları üzərində böyük təsirə malikdir. Terrorla bağlı məlumatların artması onu data mining metodlarının tətbiqi üçün əlverişli hala gətirir. "Data mining" terrorizmlə mübarizədə əhəmiyyətli potensiala malik olan üsullardan biridir. Data mining və avtomatlaşdırılmış məlumat analiz üsulları yeni deyil; onlar artıq özəl sektor və dövlət orqanlarında effektiv istifadə olunur [55, s.5]. Data mining böyük məlumat yığımlarından faydalı məlumatların çıxarılması və əldə edilmiş nəticələrdən effektiv istifadə olunması metodudur. Data mining böyük verilənlər bazasında nümunələrin və əlaqələrin avtomatik aşkarlanması ilə əlaqəlidir. Data mining metodları bir sıra sahələrdə, məsələn, kriminal məlumatların analizi, bank dələduzluqları və başqa kritik problemlərin həllində daha yüksək təsirə malikdir. Bundan başqa, data mining metodları cinayətlərin, terror fəaliyyətinin və saxta davranışların aşkarlanmasında da istifadə oluna bilər [99, 48]. Eyni zamanda qeyd edək ki, müxtəlif platformalarda yaradılan veb səhifələr fərqli data strukturuna malik olduğundan onların tək bir alqoritm vasitəsilə analizi çətindir. Böyük miqdarda veb məlumatlarının analizi üçün effektiv veb mining alqoritmlərinə ehtiyac vardır. Veb səhifələrdəki mətn məlumatlarının mənimsənilməsi və onların terrorizmlə bağlılığını aşkar etmək üçün veb mining alqoritmləri istifadə oluna bilər. Bu yolla veb səhifələrə hakim ola və onların terrorizmi təbliğ etdiyi yoxlanıla bilər [94]. Veb mining və data mining hər ikisi mətnin intellektual analizi üçün geniş istifadə olunur. Data mining alqoritmləri strukturlaşdırılmış verilənlər yığımının manipulyasiyasında effektiv olduğu halda,

veb mining alqoritmləri internetdə struktursuz veb səhifələr və mətn məlumatlarının analizində istifadə olunur.

Data mining ilə yanaşı klasterləşmə alqoritmləri də terrorizmin qarşısının alınması və həll prosesinin sürətləndirilməsində istifadə oluna bilər. Belə ki, internet istifadəçilərə bir çox qiymətli məlumatlar təqdim edir və bu məlumatların həcmi getdikcə artır. Bu potensial təhlükəli sənədləri müəyyən etməyi çox çətinləşdirir. Yalnız sənədlərdən açar sözləri çıxarmaq məzmunu təsnif etmək üçün kifayət deyil. Veb sənədlərinin klassifikasiyası terrorizmə dair sənədlərin aşkar edilməsində uğurla tətbiq oluna bilər [89].

Bütün bu faktlar terrorizmlə mübarizədə big data analitikasının necə mühüm rol oynadığını və bu məsələlərin geniş şəkildə araşdırılmasına ehtiyac olduğunu göstərir.

Təklif olunan metod: Hazırda sosial media kanalları günbəgün artmaqdadır və hər gün milyonlarla insan bu kanalların faydalarından yararlanır. Sosial media istifadəçilərə məzmun yaratmaq və mübadilə etməyə imkan verən müxtəlif onlayn platformalardan ibarətdir. Sosial media kanallarına əsasən aşağıdakılar daxildir: Sosial şəbəkələr (məsələn, Facebook və LinkedIn), sosial platformalar (e-gov), bloqlar (Blogger və WordPress), mikro bloqlar (məsələn, Twitter və Tumblr), sosial xəbərlər (məsələn, Digg və Reddit), media mübadiləsi-media sharing (məsələn, Instagram və YouTube), wikilər (Vikipediya və Wikihow), sual-cavab saytları (məsələn, Yahoo! Answers və Ask.com) və s. Bu sosial media kanallarının tərkibində çoxsaylı informasiyalar cəmləşir və zaman keçdikcə məlumatların artması onların analizini çətinləşdirir. Məhz buna görə də bu məlumatların analizi üçün yeni texnologiyaların işlənməsinə zərurət yaranır [63].

Yuxarıda göstərildiyi kimi, e-dövlət son vaxtlarda ən çox istifadə olunan sosial media kanallarından birinə çevrilmişdir. E-dövlət interaktiv mühitdir və hər kəs bu mühitdən yararlana bilər. Bu mühitdən xoş məramlı olduğu kimi bəd niyyətli insanlar da öz məqsədləri üçün istifadə edə bilərlər. Belə ki, terror təşkilatının üzvləri dövləti hədələmək, siyasi mesaj göndərmək, özlərindən sonra iz qoymaq məqsədilə eyni vaxtda dövlətə hədələyici məktublar yaza bilərlər. Lakin bu məktublar, rəylərdə

fikirlər sətiraltı məna daşıyıb şifrələnmiş ola bilər. Eyni zamanda çoxsaylı hədələyici məktubların gəlməsi onların analizini çətinləşdirə bilər. Təklif olunmuş metodda əsas məqsəd yazılan rəyləri avtomatik olaraq analiz etmək və potensial təhlükəli rəyləri aşkarlamaqdır. Bunun üçün isə text mining metodlarından istifadə olunması təklif olunur. Text mining big data analitikasının üsullarından biridir və mətn məlumatlarından bilik əldə etmək məqsədilə istifadə olunur. Text mining metodlarına informasiyanın çıxarılması (Information extraction), mətnin referatlaşdırılması (Text summarization), suala cavab (Question answering), sentiment analiz (opinion mining) və s. daxildir. Təklif etdiyimiz metodda bu üsulların bir neçəsindən istifadə edəcəyik. Qeyd edək ki, təklif olunan metodda əsas məqsəd yazılan rəylərin analizi əsasında şübhəli rəyləri müəyyən etməkdir. Bunun üçün əvvəlcə terrorla əlaqəli sözlərdən ibarət lüğət bazası yaradılır. Daha sonra rəyləri analiz edərək onların polyarlığı, yəni müsbət, mənfi və neytral fikri ifadə etdiyi müəyyən olunur. Vaxta və yerə qənaət etmək məqsədilə yalnız mənfi fikri ifadə edən rəylərə baxılır. Daha sonra rəy ilə terrorla əlaqəli yaradılmış lüğət bazası arasında yaxınlıq hesablanılır. Əgər yaxınlıq müəyyən olunmuş həddən böyükdürsə, onda bu rəy şübhəli rəylər çoxluğuna daxil olur və onu yazan istifadəçi müvafiq orqanlar tərəfindən nəzarətə götürülür [17].

Təklif olunan metodda mətnə sözlər çoxluğu kimi baxılır və cümlələrdəki hər bir sözün semantik və sintaktik strukturu nəzərə alınır. Metod aşağıdakı mərhələlərdən ibarətdir:

- İlkin emal;
- Rəylərin polyarlığının müəyyən olunması;
- Terrorla əlaqəli şübhəli rəylərin seçilməsi.

Hər bir mərhələ aşağıda ətraflı şəkildə izah olunmuşdur:

***İlkin emal*:** İlkin emal zamanı hər bir mətn (rəy) ".", "!" və "?" vasitəsilə cümlələrə ayrılır. Mətn ümumişlək sözlərdən təmizlənir. Hər bir söz müxtəlif formalarda şəkilçilər qəbul etdiyi üçün bütün sözlər ilkin variantına qaytarılır. Daha sonra hər bir sözün sinonimləri tapılır və o, çoxluq şəklində təsvir olunur.

Rəylərin polyarlığının müəyyən olunması: Bu mərhələdə rəyin polyarlığı müəyyən olunur. Bunun üçün isə sentiment analizdən istifadə olunması təklif olunur. Yazılan rəyin polyarlığını tapmaq üçün aşağıdakı mərhələlərdən istifadə olunur:

Əvvəlcə rəy $T = \{S_1, S_2, ..., S_N\}$ cümlələr çoxluğu şəklində təsvir olunur. Burada N cümlələrin sayıdır. Daha sonra aşağıdakı düsturdan istifadə edərək hər bir rəyin polyarlığı müəyyən olunur:

$$score(T) = sign(\sum_{S \in T} score(S)) \tag{2.2.1}$$

burada $score(S)$ - rəyə daxil olan S cümləsinin polyarlıq dərəcəsidir. $sign(x)$ işarə funksiyası aşağıdakı kimi təyin olunur:

$$sign(x) = \begin{cases} 1, & \text{əgər} \quad x > 0 \\ 0 & \text{əgər} \quad x = 0 \\ -1 & \text{əgər} \quad x < 0 \end{cases} \tag{2.2.2}$$

Rəyə daxil olan cümlələrin polyarlıq dərəcələrinin cəmi sıfırdan böyük olarsa, rəy pozitiv, sıfra bərabər olarsa, neytral, sıfırdan kiçik olarsa, neqativ sinfə daxil olur. Cümlənin polyarlıq dərəcəsi $score(S)$ aşağıdakı düstur vasitəsilə hesablanılır:

$$score(S) = \sum_{w \in S} score(w) \tag{2.2.3}$$

w sözünün polyarlığı [11]-də göstərilən qayda ilə hesablanır. Belə ki, w sözünün WordNet şəbəkəsindən istifadə etməklə sinonimlər çoxluğu tapılır, daha sonra hər bir sözün sentiment lüğətdə balına baxılır və onların müsbət və mənfi polyarlığı müəyyən olunur:

$$score(w) = \frac{1}{n} \sum_{sw \in synset(w)} Pos_{sw} - \frac{1}{m} \sum_{sw \in synset(w)} Neg_{sw} \tag{2.2.4}$$

burada $synset(w)$- w sözünün WordNet şəbəkəsindən götürülmüş sinonimlər çoxluğu, Pos- sw sözünün müsbət, Neg - sw sözünün mənfi balını, n və m isə uyğun olaraq müsbət və mənfi sözlərin sayını göstərir.

Terrorla əlaqəli şübhəli rəylərin seçilməsi: Növbəti mərhələdə seçilmiş mənfi rəylərin terrorla əlaqəli olub-olmadığını müəyyən etmək üçün onların əvvəlcədən yaradılmış lüğət bazasında olan sözlərlə ilkin müqayisəsi aparılır. Əgər yaxınlıq

müəyyən olunmuş həddən böyükdürsə, onda növbəti mərhələdə yazılmış rəy ilə lüğət bazası arasında daha detallı müqayisə aparılır.

Tutaq ki, V_{terror} – terrorla bağlı sözlər və onların genişlənməsindən ibarət lüğət bazasıdır. Burada genişlənmə dedikdə sözlərin sinonimləri çoxluğu nəzərdə tutulur. Seçilmiş mənfi rəylər sözlər çoxluğu $T = \{w_1, w_2, ..., w_m\}$ şəklində təsvir olunur. Hər bir söz isə sözlər çoxluğu $w_i \to synset(w_i), i = 1, 2, ..., m$ kimi təsvir olunur. Yazılmış rəyin lüğət bazasında olan sözlərlə yaxınlığını hesablamaq üçün aşağıdakı düsturdan istifadə olunması təklif olunur:

$$SW_i = synset(w_i) \cap V_{terror} = \{w_{i1}, w_{i2}, ..., w_{im_i}\},\ i = 1, ..., m \qquad (2.2.5)$$

Daha sonra (2.2.5) düsturu vasitəsilə hesablanan sözlərin bir-birinə nə dərəcədə yaxın olduğu müəyyən olunur:

$$\theta_i = \frac{2}{m_i(m_i - 1)} \sum_{j=1}^{m_i - 1} \sum_{k=j+1}^{m_i} sim(w_{ij}, w_{ik}),\ i = 1, ..., m \qquad (2.2.6)$$

Sözlər arasındakı semantik yaxınlığı hesablamaq üçün [23]-də təklif olunan metoddan istifadə olunur. Belə ki, əvvəlcə WordNet şəbəkəsindən istifadə etməklə, sözün informativ məzmunu (yükü) $IC(w)$ təyin edilir [10]:

$$IC(w) = 1 - \frac{\log(|synset(w)| + 1)}{\log(w_{\max})} \qquad (2.2.7)$$

burada $w_{\max}$ – WordNet semantik şəbəkəsindəki sözlərin ümumi sayı, $|synset(w)|$ – w sözünün sinonimlərinin sayıdır.

Sonra (2.2.8) düsturundan istifadə edərək sözlər arasındakı semantik yaxınlıq hesablanır:

$$sim_{IC}(w_1, w_2) = \begin{cases} \dfrac{2 * IC(LCS(w_1, w_2))}{IC(w_1) + IC(w_2)}, & \text{əgər} \quad w_1 \neq w_2 \\ 1, & \text{əgər} \quad w_1 = w_2 \end{cases} \qquad (2.2.8)$$

burada $LCS(w_1, w_2)$ – WordNet şəbəkəsində w_1 və w_2 sözlərinin ən yaxın olduğu ortaq sözdür.

Sözlər arasındakı semantik yaxınlığı WUP metrikasından istifadə etməklə də hesablamaq mümkündür:

$$sim_{WUP}(w_1, w_2) = \frac{2 * depth(w)}{depth(w_1) + depth(w_2) + 2 * depth(w)} \quad (2.2.9)$$

burada $depth(w_1)$ – WordNet semantik şəbəkəsində w_1 - dən w-ya qədər olan qovşaqların sayı; $depth(w_2)$ – w_2 -dən w-ya qədər olan qovşaqların sayı; $depth(w)$ – w -dan şəbəkənin kökünə qədər olan qovşaqların sayıdır.

Beləliklə, sözlər arasında semantik yaxınlıq (2.2.8) və (2.2.9) düsturları ilə verilən metrikaların xətti kombinasiyası kimi təyin olunur:

$$sim(w_1, w_2) = \alpha \times sim_{IC}(w_1, w_2) + (1-\alpha) \times sim_{WUP}(w_1, w_2) \quad (2.2.10)$$

burada $\alpha \in [0,1]$ – çəki əmsalıdır.

Daha sonra (2.2.6) düsturu vasitəsilə tapılan sözlərin bir-birinə yaxınlıq dərəcələri θ_i-lər toplanır. Hesablanan cəm əvvəlcədən təyin olunmuş t həddən böyük olarsa, rəyi yazan istifadəçi terrorla əlaqəli şübhəli hesab olunur və nəzarətə götürülür. Aşağıdakı düstur bu məqsədlə istifadə olunur:

$$P(T) = \begin{cases} 1, & \text{əgər} \quad \frac{1}{m}\sum_{i=1}^{m} \theta_i \geq t \\ 0, & \text{əks halda} \end{cases} \quad (2.2.11)$$

burada “1” yazılan rəyin terrorla bağlı şübhəli olduğunu, ”0” isə olmadığını göstərir.

Şübhəli bilinən istifadəçi rəyinin terrorla bağlılığının ehtimalını müəyyən etmək mümkündür. Bunun üçün Naive Bayes klassifikasiya metodundan istifadə olunması təklif olunur. Şübhəli bilinən rəyin terrorla bağlılığının ehtimalını hesablamaq üçün bu rəyə daxil olan hər bir sözün terrorla əlaqəli olma ehtimalı hesablanır:

$$P(T) = P(w_1, w_2, ..., w_n) \quad (2.2.12)$$

$$P(w_1, w_2, ..., w_n) = \prod_{i=1}^{n} P(w_i) \quad (2.2.13)$$

burada $P(T)$-istifadəçi rəyinin ehtimalını, $P(w_i)$ isə bu rəyə daxil olan hər bir sözün terrorla əlaqəli olma ehtimalını göstərir.

Qeyd edək ki, burada hər bir sözün genişlənmiş sinonimləri çoxluğu da nəzərə alınır. Yazılan rəyin terrorla əlaqəli olma ehtimalını hesablamaq üçün bu rəyə daxil

olan hər bir sözün və onun sinonimləri çoxluğunun əvvəlcədən yaradılmış V_{terror} lüğət bazasında işlənmə tezliyinə baxılır. Bunun əsasında şübhəli bilinən rəyin terrorla əlaqəli olma ehtimalı hesablanır. Bunun üçün Bayes metoduna əsasən aşağıdakı düsturlardan istifadə olunması təklif olunur [150, s.5-10]:

$$P(V_{terror}|T) = \frac{P(T|V_{terror})P(V_{terror})}{P(T)} \tag{2.2.14}$$

$$P(T|\,V_{terror}) = P(w_1, w_2,..., w_n|V_{terror}) = \prod_{i=1}^{n} P(w_i|\,V_{terror}) \tag{2.2.15}$$

$$P(w_i|\,V_{terror}) = \frac{P(w_i \cap V_{terror})}{P(V_{terror})} \tag{2.2.16}$$

burada $P(V_{terror}|T)$ - şübhəli bilinən rəyin V_{terror} sinfindən olma ehtimalını göstərir.

Qeyd edək ki, (2.2.15)-də sözlərdən birinin lüğət bazasında olmaması halı nəticəyə birbaşa təsir edir (yəni ən azı bir söz V_{terror} lüğət bazasında olmadıqda istifadəçi rəyinin terrorla əlaqəli ehtimalı 0-a bərabər olur). Bu halda mətnin terrorla əlaqəli olma ehtimalını hesablamaq üçün (2.2.15) düsturunun əvəzinə aşağıdakı düsturdan istifadə edilməsi daha məqsədə uyğundur:

$$P(T|\,V_{terror}) = \frac{1}{n}\sum_{i=1}^{n} P(w_i|\,V_{terror}) \tag{2.2.17}$$

Beləliklə, istifadəçi rəyləri əsasında terrorizmlə əlaqəli şübhəli bilinən istifadəçilər müəyyən olunur və onların fəaliyyəti müvafiq orqanlar tərəfindən nəzarətə götürülür.

III FƏSİL. E-DÖVLƏTDƏ GİZLİ SOSİAL ŞƏBƏKƏLƏRİN AŞKARLANMASI ÜÇÜN METOD VƏ ALQORİTM

İnternetin sürətli inkişafı vebdə sosial şəbəkələrin daha da geniş yayılmasına təkan vermişdir. İnsanlar arasında sosial media alətlərinin geniş şəkildə istifadəsi dövlət orqanlarının bu saytlara qoşularaq onların üstünlüklərindən faydalanmağa vadar edir. E-dövlət xidmətləri ilə bağlı bir sıra tədqiqatlar e-dövlətdə fiziki, hüquqi şəxslər, təşkilatlar və dövlət arasında olan qarşılıqlı münasibətlərin şəbəkəsini öyrənir. E-dövlət saytları ictimai saytlar kimi qəbul olunur və bu saytlara daxil olan vətəndaşlar birbaşa dövlətin qərar-qəbuletmə prosesində iştirak edirlər.

Sosial media dövlətin fəaliyyətinə təsir edən əsas amillərdən biridir. Özəl sektor, ictimai təşkilatlar və insanlar tərəfindən sosial media saytları və alətlərinin geniş istifadəsi dövləti vətəndaşlarla münasibətlərin yenidən qurulması, onların qərarqəbuletmə prosesində iştirak səviyyəsinin artırılmasında bundan necə faydalanmaq haqqında düşünməyə vadar edir. Bu e-dövlət saytının ictimai saytlar kimi qəbul olunması və ondan istifadə edən vətəndaşların sayının artırılmasında vacibdir. Sosial şəbəkələr insanların qarşılıqlı fəaliyyət göstərdiyi saytlardır və bu cür saytlar dövləti vətəndaşlarla daha da yaxınlaşdıra bilər.

E-dövlət mühitində idarəetmənin yaxşılaşdırılması, dövlətin təhlükəsizliyinin təmin olunması və dövlətə qarşı təbliğatın vaxtında aşkarlanması günümüzün aktual məsələlərindən biridir. E-dövlətdə dövlət əleyhinə fəaliyyət göstərən gizli sosial şəbəkələrin aşkarlanması e-dövlətin təhlükəsizliyinin təmin olunmasının əsas amillərindən biri hesab olunur [2]. Məlumdur ki, istifadəçilər e-dövlətdə hər hansı məlumata yazdıqları şərhlər vasitəsilə münasibət bildirirlər. Bu şərhlərin arxasında kriminal qrupların, dövlətə qarşı təbliğatların olub-olmamasını müəyyən etmək üçün onlar analiz olunmalıdır. Bunu nəzərə alaraq e-dövlətdə idarəetmənin yaxşılaşdırılması, dövlətə qarşı təbliğatın qarşısının alınması və təhlükəsizliyinin təmin olunması məqsədilə gizli sosial şəbəkələrin aşkarlanması üçün metod təklif olunmuşdur. Təklif edilmiş metodda opinion və text mining texnologiyalarından istifadə olunmuşdur.

3.1. E-dövlətdə gizli sosial şəbəkələrin aşkarlanması üçün metod

Veb hazırda informasiyanın yayılmasında populyar vasitələrdən birinə çevrilmişdir. Veb-də verilənlərin sürətlə artması, müxtəlifliyi və dinamikası nəticəsində informasiya istifadəçiləri bir sıra problemlərlə qarşılaşa bilərlər. Bu problemlərə müvafiq informasiyanın tapılması, veb-də mövcud informasiyadan biliyin aşkarlanması, fərdi istifadəçilərin davranışının öyrənilməsi və s. daxildir [95].

Veb-də sosial şəbəkələrin çıxarılması sosial şəbəkə analizində mühüm məsələlərdən biridir. Sosial şəbəkələr müxtəlif mənbələrdən çıxarıla bilər. Bu mənbələrə veb səhifələr, e-poçt, bloqlar, onlayn sosial şəbəkə saytları və s. daxil ola bilər [153, s.17]. İnformasiya mənbəyindən asılı olaraq müxtəlif sosial şəbəkə çıxarılma metodları aşağıdakı kimi qruplaşdırıla bilər:

- Veb-də sosial şəbəkə çıxarılma metodları;
- E-poçt mühitində sosial şəbəkə çıxarılma metodları;
- Bloqlarda sosial şəbəkə çıxarılma metodları;
- Onlayn sosial şəbəkə saytlarında sosial şəbəkə çıxarılma metodları;
- Viki mühitdə sosial şəbəkə çıxarılma metodları;
- Çoxmənbəli verilənlərə əsaslanan sosial şəbəkə çıxarılma metodları;
- E-dövlətdə sosial şəbəkə çıxarılma metodları.

Veb-də sosial şəbəkə çıxarılma metodları: Bir sıra tədqiqatlarda veb-də sosial şəbəkələrin çıxarılması üçün axtarış mexanizmindən istifadə olunmuşdur.[86]-da ilk dəfə olaraq sosial şəbəkələrin çıxarılmasında avtomatlaşdırılmış interaktiv alət (vasitə) işlənib hazırlanmışdır. Bu interaktiv alət Altavista adlandırılmışdır. Burada, Altavistadan hər hansı sənəddə (məsələn, şəxsi ana səhifələr, məqalələrdə həmmüəlliflərin siyahıları, istinadlar və s.) adların birgə görünməsi ideyasından istifadə edərək, sosial şəbəkələrin çıxarılmasında istifadə olunur. [159]-da aparılan tədqiqatda konfrans iştirakçılarının ad, e-poçt, iş yerləri və s. kimi atributlardan istifadə edərək, sosial şəbəkələrin çıxarılması üçün sistem təklif olunmuşdur. Burada, hər hansı iki iştirakçı arasında əlaqələr Jakkard ölçüsündən istifadə edərək, [86]-da olduğu kimi axtarış mexanizminə sorğu göndərərək toplanan veb məlumatlar əsasında müəyyən olunur. [112]-də veb-də sosial şəbəkələrin aşkarlanmasına

əsaslanaraq, ünsiyyət və qarşlılıqlı əlaqələrin asanlaşdırılması məqsədilə akademik cəmiyyət üçün nəzərdə tutulmuş şəbəkə sistemi təsvir olunmuşdur. Bu sistem Polyphonet adlanır, polifoniya və şəbəkə terminlərinin birləşməsindən yaranmışdır.

***E-poçt mühitində sosial şəbəkə çıxarılma metodları*:** E-poçt insanların sosial şəbəkə saytlarına çıxış imkanı əldə etdiyi və əlaqə qurmaq üçün istifadə etdikləri əsas vasitələrdən biridir. Özünəməxsus xüsusiyyətlərinə görə e-poçt sosial şəbəkələrdə tədqiqat üçün çox aktual sahə hesab olunur. E-poçt verilənləri bir sıra üstünlüklərinə görə sosial şəbəkələrin analizində güclü məlumat mənbəyinə çevrilmişdir. Bunlara e-poçtdan istifadənin hər yerdə mümkün olması, uzun ömürlülük, e-poçtdan istfadənin qarşılıqlı olması, yəni həm göndərən, həm də qəbul edən tərəfindən istifadə olunması və s. daxildir. Üstünlükləri ilə yanaşı e-poçtdan istifadədə bəzi problemlər də mövcuddur. Burada sosial şəbəkələrin çıxarılması kimi məsələlər asan deyil. Belə ki, eyni şəxsin bir neçə e-poçt ünvanının olması; spamlar; e-poçt məzmununa görə sosial münasibətlərin təsnifatı və s. kimi məsələlər bir sıra çətinliklər yaradır.

[30]-da sosial şəbəkələrin çıxarılması üçün EmailNet sistemi təklif olunmuşdur. EmailNet sistemində əvvəlcə həm fərdi e-poçt istifadəçilərindən, həm də təşkilati mail serverlərindən gələn məktublar (poçtlar) çıxarılır. Daha sonra spamları xaric etmək üçün filtirlərdən istifadə olunur. Burada həmçinin e-poçtları sosial əlaqələrin bir neçə sinfinə bölmək üçün mətn klasterləşmə texnologiyasından istifadə olunmuşdur. E-poçt və internetdən sosial şəbəkə və əlaqə məlumatlarının çıxarılması [52]-də müzakirə olunmuşdur. Bu məqalədə verilmiş istifadəçinin e-poçtuna əsaslanaraq onun sosial şəbəkəsi və bu şəbəkəyə daxil olan üzvlərin əlaqə məlumatlarınınn çıxarılması üçün sistem təklif olunmuşdur. Bu sistem, rekursiv olaraq həyata keçirilir və "dostumun dostu" prinsipi əsasında geniş şəbəkə qurulur.

Bloqlarda sosial şəbəkə çıxarılma metodları: Bloq məkanı şəxslərin öz fikirlərini ifadə etdiyi, cəmiyyətdə baş verən hadisələri, faktları müzakirə etdiyi sosial media saytlarının şəbəkəsidir. Bu mühit sosial informasiyanın zəngin mənbələrindən biridir. Son illər bloqların miqyasının, müxtəlifliyinin və populyarlığının kəskin artması onu sosial şəbəkələrin avtomatik çıxarılması üçün yetkin sahəyə çevirmişdir.

[115]-də bloq məkanında şəxslərin adlarının və onlar arsındakı münasibətlərin müəyyən edilməsilə sosial şəbəkələrinin çıxarılması üçün SONEX [SOcial Network Extraction] adlı sistem təklif olunmuşdur. Bu sistem bloq məkanında ismarıclardan şəxs cütlüklərinin klasterləşdirilməsi və çıxarılması yolu ilə işləyir. Klasterləşmə addımında oxşar məzmunlu şəxslər birlikdə qruplaşdırılır, hər klaster analiz olunur və klaster daxilində cütlərin məzmunu yoxlanıldıqdan sonra onlar eyni qrupa daxil edilir.

Bloqlar informasiya diffuziyasında mühüm vasitəyə çevrilmişdir və gizli sosial strukturlar informasiya axınının səviyyəsi və əhatə dairəsində əhəmiyyətli təsirə malikdir. [152]-də sosial strukturun məlumat axınına təsirini ölçmək üçün yeni yanaşma təklif olunmuşdur. Bu məqalədə aparılan eksperimentlər sosial strukturların "maraq" mövzularının yayılma şəbəkəsinə təsir göstərdiyini və sosial şəbəkələrin informasiyanın yayılmasındakı təsirinin informasiyanın özünün xarakteristikası ilə bağlı olduğunu göstərir.

Onlayn sosial şəbəkə saytlarında sosial şəbəkə çıxarılma metodları: Onlayn sosial şəbəkələr son illərdə populyarlıq qazanmışdır və fərdi məlumatların böyük hissəsi bu saytlarda mövcuddur. Onlayn sosial şəbəkə saytlarında yarı-strukturlaşdırılmış şəxsi məlumatları ehtiva edən istifadəçi profillərinin böyük həcmi yerləşir. Bu saytlarda mövcud olan dinamik verilənlərdən (məlumatlardan) sosial şəbəkələrin çıxarılması istiqamətində bir sıra tədqiqatlar aparılmışdır. [41]-də Facebook istifadəçiləri arasında olan münasibətlər haqqında müvafiq informasiyanın çıxarılmasının mümkünlüyü müzakirə olunur. Onlayn sosial şəbəkə servislərində sosial şəbəkə çıxarılma metodlarının əksəriyyətində aşkar əlaqələrdən istifadə olunduğu halda, onların çox az qismində ismarıc mövzuları daxilindəki münasibətlər analiz olunmuşdur. [142]-də ismarıc mövzuları daxilində gizlənmiş sosial münasibətlərin çıxarılması yolu ilə sosial şəbəkələr aşkar olunur. Burada eyni mövzuya ismarıc yazan istifadəçilər arasında ən çox birlikdə görünən istifadəçilər çoxluğu müəyyən olunaraq ismarıc mövzuları daxilində gizlənmiş münasibətləri aşkar etməyə çalışırlar. [108]-də müəllimsiz sifət tanıma (unsupervised face

recognition.) metodlarının köməyilə Facebookda fotoalbomlardan sosial şəbəkələrin qurulması üzrə tədqiqat aparılmışdır.

***Viki mühitdə sosial şəbəkə çıxarılma metodları*:** Viki-mühitdə sosial şəbəkələrin analizi və istifadəçilər arasındakı münasibətlər müxtəlif tədqiqatçılar tərəfindən öyrənilmişdir. [82]-də viki istifadəçilər arasında sosial münasibətləri aşkarlayan və viki mühitdə vandalların, təcrübəsiz istifadəçilərin müəyyən edilməsinə imkan verən model təklif olunmuşdur. Viki mühitdə sosial əlaqələrin, təsirlərin ölçülməsi üçün digər yanaşma [114]-də təklif olunmuşdur. Təklif olunan metodda üç kriteriyadan istifadə olunur: 1) viki istifadəçilərin sayı; 2) İnteraktivlik, yəni xüsusi zaman intervalında viki istifadəçilər tərəfindən redaktələrin sayı; 3) İntensivlik, yeni redaktələr nəticəsində viki-səhifələrdə dəyişikliklərin sayı. [19]-da mətn klasterləşmə metodlarının köməyilə informasiya münaqişəsinə səbəb olan viki istifadəçilərin gizli sosial şəbəkələrinin çıxarılmasında istifadə oluna bilən metod təklif olunmuşdur. Bu məqalədə informasiya münaqişəsinə səbəb olan struktursuz məqalələrin klasterləşdirilməsilə viki istifadəçilərin sosial şəbəkəsi qurulur. Burada münaqişəli məqalələrin klaterləşdirilməsi üçün hibrid çəkili qeyri-səlis c-means metodundan istifadə olunur.

***Çoxmənbəli verilənlərə əsaslanan sosial şəbəkə çıxarılma metodları*:** Dünyanın ən böyük verilənlər bazası olaraq qəbul edilən Veb sosial şəbəkələrin çıxarılması məqsədilə aparılan müxtəlif tədqiqatlarda istifadə olunmuşdur. Tədqiqatların bir çoxunda sosial şəbəkələrin çıxarılması üçün tək mənbədən istifadə olunmuşdur. Çox az sayda aparılan tədqiqatlarda sosial şəbəkələrin çıxarılmasında bir neçə mənbədən istifadə olunmuşdur. [15]-də veb-də müxtəlif mənbələrdən əldə olunmuş informasiyalar əsasında elmi tədqiqatçıların şəbəkəsinin müəyyən olunması üçün metod təklif olunmuşdur. [14]-də veb-də əlçatan informasiya mənbələri əsasında elmi təşkilatların sosial şəbəkəsinin sintezi üçün yanaşma təklif olunmuşdur. [163]-də çatlardan və e-poçtdan sosial şəbəkələrin çıxarılması üçün sistem təklif olunmuşdur. Burada iki əsas komponentdən istifadə olunur: oflayn verilənlərin toplanması və digəri onlayn verilənlərin emalı. Əvvəlcə e-poçtdan və çatdan əlaqəli kommunikasiya verilənləri toplanır. Bu şəkildə çıxarılan verilənlər filtirlənir və

relyasion verilənlər bazasında saxlanlır. Oflayn verilənlərin toplama modulu tərəfindən toplanan, emal olunan və yadda saxlanılan verilənlər sosial şəbəkələrin qurulması və vizuallaşdırılması üçün onlayn emaletmə modulu tərəfindən istifadə olunur. [87]-də çoxölçülü sosial şəbəkələrin çıxarılması üçün ümumi model təklif olunmuşdur. Bu modeldə istifadəçilərin davranış dinamikası haqqında məlumat toplanılır. Təklif olunan modeldə sosial şəbəkələrin analizi üçün üç ölçüdən, yəni əlaqələr, zaman və qruplar-dan istifadə olunur. Burada əlaqə dedikdə sistem istifadəçiləri arasında olan bütün əlaqələr (e-poçt və telefon xətti ilə birbaşa və ya birbaşa olmayan əlaqələr) nəzərdə tutulur. Qrup ölçüsü, klasterləşmə prosesində əldə oluna biləcək bütün sosial qrupları əhatə edir. Zaman ölçüsü, müəyyən zaman anında mövcud olan əlaqələrə və ya müəyyən zaman periodu ərzində çıxarılan əlaqələrə, yəni müəyyən zaman pəncərəsi daxilindəki insan fəaliyyətinə əsaslanır.

***E-dövlət mühitində gizli sosial şəbəkələrin aşkarlanması*:** E-dövlət mühitində gizli sosial şəbəkələrin aşkarlanması dövlətin təhlükəsizliyinin təmin olunması, idarəetmənin yaxşılaşdırılmasının əsas amillərindən biridir. Hazırda müxtəlif cinayət təşkilatları əlaqə yaratmaq, işə cəlbetmə üçün bəzi sosial şəbəkə saytlarından istifadə edirlər və belə təşkilatlar həmişə münasibətlərini gizli saxlayırlar. E-dövlət mühitində sosial şəbəkə məlumatlarını analiz etməklə gizli əlaqələri izləmək mümkündür. Bu mühitdə sosial şəbəkə analizi kriminal şəbəkələrin aşkarlanmasında istifadə oluna bilər.

Məlumdur ki, e-dövlət mühitində hər hansı məlumata istifadəçilər şərhlər yazmaqla münasibət bildirə bilərlər. Belə ki, hər hansı qanun layihəsi qəbul olunmamışdan əvvəl onu e-dövlət sisteminə yerləşdirərək e-vətəndaşların rəyini öyrənmək, bunun vasitəsilə qanunun müsbət və mənfi cəhətlərini müəyyən etmək mümkündür. Həmçinin qəbul olunmuş qanun layihəsinin effektiv olmamasının və ya işləməməsinin səbəblərini əks-əlaqə mexanizmi qurmaqla vətəndaşların rəylərinin analizi vasitəsilə müəyyən etmək mümkündür. İstifadəçilərin bir-birinə yazdığı şərhlər vasitəsilə bu istifadəçilər arasında olan münasibətləri modelləşdirmək olar. Burada, müxtəlif yanaşmalar ola bilər, məsələn, hər hansı istifadəçinin yazdığı şərhə digər istifadəçilərin verdiyi cavab, bu şərhə cavab verən istifadəçilərin sayı və s.

Bunu nəzərə alaraq təklif olunan yanaşmada isə istifadəçilərin e-dövlət mühitində yazdığı şərhlər vasitəsilə gizli sosial şəbəkələrin aşkarlanması nəzərdə tutulmuşdur [18]. Təklif olunan yanaşma aşağıdakı hissələrdən ibarətdir:

1. Verilənlərin toplanması və ilkin emalı;
2. Təsnifat;
3. Sosial şəbəkənin qurulması;
4. Sosial şəbəkənin analizi.

Təklif olunan yanaşmanın hər bir mərhələsi aşağıda ətraflı izah edilir.

1) Verilənlərin toplanması və ilkin emalı: Bu mərhələdə ilkin olaraq, e-dövlət mühitində hər hansı məlumata yazılan şərhlər çoxluğu toplanılır. Tutaq ki, e-dövlət mühitində n sayda məlumat (veb səhifə, informasiya) $T_i, (i = 1,2,..., n)$ yerləşdirilmişdir. i-ci məlumata yazılan şərhlər çoxluğu aşağıdakı kimi işarə olunur:

$$C_i = \{c_i^j\}, \quad i = 1,2,..., n, \; j = 1,2,..., m \tag{3.1.1}$$

burada c_i^j – i-ci məlumata j-ci istifadəçinin yazdığı şərhlər çoxluğu, m istifadəçilərin sayıdır.

Şərhlər toplandıqdan sonra onlar üzərində ilkin emal həyata keçirilir. İlkin emal zamanı durğu işarələri və probellər aradan qaldırılır.

***2) Təsnifat*:** Növbəti mərhələdə hər bir məlumata yazılan şərhlər çoxluğu 3 sinifdə qruplaşdırılır: Pozitiv, neqativ və neytral. C_i^+ – ilə pozitiv, C_i^- – ilə neqativ və C_i^0 – neytral sinif işarə olunur:

$$C_i = C_i^+ \cup C_i^- \cup C_i^0, \; i = 1,..., n \tag{3.1.2}$$

Yazılan şərhləri bu üç sinifdə qruplaşdırmaq üçün mətnlərin analizində ən qabaqcıl texnologiyalardan biri olan sentiment analizdən istifadə olunması təklif olunur.

Sentiment analiz, yazılı halda verilən ifadənin (sənəd, şərh, e-poçt və s.) əks etdirdiyi duyğunun kompüterin köməyi ilə avtomatik olaraq aşkar olunması prosesidir. Bu işlər ədəbiyyatda opinion mining olaraq da adlandırılır. Aşkar edilməsi hədəflənən duyğu müəllifin (istifadəçinin) ruh halı, mövzu haqqında düşüncəsi,

yaratmaq istədiyi təsir və s. ola bilər. Sentiment analiz üzrə aparılan tədqiqatlarda əsasən üç "Lüğətlərə əsaslanan yanaşmalar", "Maşın təlimi alqoritmləri" və "Qaydalara əsaslanan yanaşmalar"dan istifadə olunur [85, s.291].

Sentiment analiz müxtəlif səviyyələrdə həyata keçirilə bilər:

Söz səviyyəsi: Bu səviyyədə sözün və multi-sözlərin sentiment polyarlığı, yəni pozitiv və ya neqativ fikri ifadə etdiyi müəyyən olunur.

Fraza səviyyəsi:. Cümlə bir neçə frazadan ibarət ola bilər, hər fraza müxtəlif mənaları ifadə edə bilər (Məsələn, Mən bəyəndim, amma hər kəs bəyənmədi). Məhz buna görə də bu səviyyədə hər bir frazaya ayrı-ayrılıqda baxılır.

Cümlə səviyyəsi: Cümlələrin duyğu analizi söz və fraza səviyyəsinə əsaslanır. Belə ki, cümlədə müxtəlif polyarlığı (yəni həm pozitiv, həm də neqativ fikri) ifadə edən söz və ya frazalar varsa, onda cümlənin ümumi polyarlıq dərəcəsi təyin olunur.

Aspekt səviyyəsi: Eyni sahəyə aid olan bir cümlə bir neçə aspektdən ibarət ola bilər və hər aspektin müxtəlif polyarlığı ola bilər. Məsələn: İşıqlandırma yaxşıdır, amma səs effektlərini bəyənmədim.

Sənəd səviyyəsi: Bu səviyyədə duyğu analizində sənədin ümumi polyarlığı təyin olunur. Bir çox hallarda, sənədin polyarlığı sözlərin və ya cümlələrin hesablanmış polyarlıq dərəcəsindən istifadə olunaraq hesablanır. Tədqiqatlar göstərir ki, sənədin polyarlığının təyin olunmasında ilk və son cümlələrin ortada işlənən cümlələrə nisbətən təsiri daha böyükdür.

Şərhlər adətən bir neçə cümlədən ibarət olur, ancaq fikir cümlə səviyyəsində ifadə olunduğundan şərhi təşkil edən hər bir cümlənin polyarlığı müəyyən olunduqdan sonra ümumi şərhin polyarlığı təyin olunur [12, s.5-15].

Tutaq ki, sentiment analiz üzrə tətbiq olunan yanaşmaların köməyi ilə sözlərin polyarlıq dərəcəsini (yəni, pozitiv və neqativ balını) göstərən V cədvəli qurulmuşdur. Təklif olunan yanaşmada şərhin hansı sinfə (pozitiv, neqativ və neytral) daxil olduğunu müəyyən etmək üçün əvvəlcə şərhlər cümlələrə parçalanır. Şərhlər cümlələrə parçalandıqdan sonra hər cümlə ümumişlək sözlərdən təmizlənir. Təmizləndikdən sonra hər bir cümlənin polyarlıq dərəcəsi müəyyən olunur. Polyarlıq həmin cümlədəki sözlərin polyarlıq dərəcəsinə görə hesablanır.

Hər bir şərhin polyarlıq dərəcəsini hesablamaq üçün aşağıdakı düsturdan istifadə olunması təklif olunur:

$$score(c_i^j) = sign\left(\sum_{s \in c_i^j} score(s)\right) \tag{3.1.3}$$

burada $score\,(c_i^j)$ - i-ci məlumata j-ci istifadəçinin yazdığı şərhin polyarlıq dərəcəsi, $score\,(s)$ - şərhə daxil olan s cümləsinin polyarlıq dərəcəsidir.

$sign\,(x)$ işarə funksiyası aşağıdakı kimi təyin olunur:

$$sign(x) = \begin{cases} 1, & \text{əgər} \quad x > 0 \\ 0, & \text{əgər} \quad x = 0 \\ -1, & \text{əgər} \quad x < 0 \end{cases} \tag{3.1.4}$$

Şərhə daxil olan cümlələrin polyarlıq dərəcələrinin cəmi sıfırdan böyük olarsa, şərh pozitiv sinfə, sıfra bərabər olarsa, neytral sinfə, sıfırdan kiçik olarsa, neqativ sinfə daxil olur. Cümlənin polyarlıq dərəcəsi $score\,(s)$ aşağıdakı düstur vasitəsilə hesablanır:

$$score\,(s) = \sum_{t \in s} score\,(t) \tag{3.1.5}$$

Cümləyə daxil olan hər bir sözün polyarlıq dərəcəsini hesablamaq üçün Pointwise Mutual Information (PMI) metodundan istifadə olunur [73]. Hər bir sözün polyarlıq dərəcəsi aşağıdakı düstur vasitəsilə hesablanır:

$$score\,(t) = \sum_{x \in V} PMI(t, x)(score^{+}(x) - score^{-}(x)) \tag{3.1.6}$$

$$PMI(t, x) = \log_2\left(\frac{n_{tx}}{n_t \cdot n_x}\right) \tag{3.1.7}$$

burada $score\,(t)$ - cümləyə daxil olan t sözünün polyarlıq dərəcəsi, $score^{+}(x)$ - cədvəldə rast gəlinən x sözünün pozitiv balını, $score^{-}(x)$ - neqativ balını və n_t, n_x, n_{tx} uyğun olaraq t, x, t & x sözlərinin rast gəlindiyi səhifələrin sayını göstərir. Beləliklə, i-ci məlumata yazılan şərhlər 3 (pozitiv, neytral və neqativ) sinifdə qruplaşdırılır.

$$C_i^- = \left\{ c_i^j \,\middle|\, score(c_i^j) = -1 \right\}$$

$$C_i^+ = \left\{ c_i^j \,\middle|\, score(c_i^j) = 1 \right\} \qquad (3.1.8)$$

$$C_i^0 = \left\{ c_i^j \,\middle|\, score(c_i^j) = 0 \right\}$$

3) ***Sosial şəbəkənin qurulması***: Bu mərhələdə sosial şəbəkənin aktorları və onlar arasında olan münasibətlər müəyyən olunur. İlkin olaraq neqativ sinfin ətrafında toplanan istifadəçilər qrupu müəyyən olunur. Hesab edirik ki, hər bir istifadəçi identifikasiya oluna bilir (ya qeydiyyatdan keçməklə, ya da IP ünvan vasitəsilə). Heç olmasa bir məlumata mənfi şərh yazan istifadəçilər qrupu aşağıdakı kimi təyin olunur:

$$U_\Sigma^- = \bigcup_{i=1}^{n} U_i^- \qquad (3.1.9)$$

burada U_i^- - i-ci məlumata mənfi şərh yazan istifadəçilərdir.

Bütün məlumatlara mənfi şərh yazan istifadəçilər qrupu isə bu şəkildə müəyyən olunur:

$$U_\Pi^- = \bigcap_{i=1}^{n} U_i^- \qquad (3.1.10)$$

U_Π^-- istifadəçilər qrupu quracağımız (aşkarlayacağımız) sosial şəbəkənin nüvəsini (əsas aktorlarını) təşkil edir.

Sosial şəbəkənin aktorları arasında münasibətlərin müəyyən olunmasında iki tip yanaşmadan istifadə olunur:

Birinci yanaşmada neqativ sinfə daxil olan istifadəçilərin birgə göründüyü məlumatlar və bu məlumatlara yazılan şərhlərin sayı əsasında sosial şəbəkənin aktorları arasında münasibətlərin müəyyən olunması nəzərdə tutulur. Bunun üçün aşağıdakı düsturdan istifadə olunması təklif olunur:

$$w_1^{j_1 j_2} = \frac{n^{j_1 j_2}}{n^{j_1} + n^{j_2}} \qquad (3.1.11)$$

burada $n^{j_1 j_2}$ – j_1 və j_2 istifadəçilərinin mənfi şərh yazdığı məlumatların sayı, n^{j_1} – j_1 -ci istifadəçinin mənfi şərh yazdığı məlumatların sayı, n^{j_2} – j_2-ci istifadəçinin mənfi şərh yazdığı məlumatların sayıdır.

$n^{j_1 j_2}$, n^{j_1} və n^{j_2} aşağıdakı düsturlar vasitəsilə təyin olunur:

$$n^{j_1 j_2} = \sum_{i=1}^{n} I(c_i^{j_1}) \cdot I(c_i^{j_2}) \tag{3.1.12}$$

$$n^{j_1} = \sum_{i=1}^{n} I(c_i^{j_1}) \tag{3.1.13}$$

$$n^{j_2} = \sum_{i=1}^{n} I(c_i^{j_2}) \tag{3.1.14}$$

burada $I(c_i^j)$ funksiyası aşağıdakı şəkildə təyin olunur:

$$I(c_i^j) = \begin{cases} 1, & \text{əgər } c_i^j \neq \varnothing \\ 0, & \text{əks halda} \end{cases} \tag{3.1.15}$$

j-ci istifadəçi i-ci məlumata heç olmasa 1 dəfə şərh yazarsa, $I(c_i^j)$ funksiyası 1, əks halda, yəni j-ci istifadəçi i-ci məlumata şərh yazmayıbsa, 0 qiyməti alır.

Burada həmçinin istifadəçilərin eyni bir məlumata yazdığı mənfi şərhlərin sayı da nəzərə alına bilər. Bu halda istifadəçilər arasındakı əlaqənin çəkisi aşağıdakı düsturun köməyilə təyin olunur:

$$\tilde{w}_1^{j_1 j_2} = \frac{\sum_{i=1}^{n} \left(m_i^{j_1} + m_i^{j_2} \right) * I(c_i^{j_1}) \cdot I(c_i^{j_2})}{M^{j_1} + M^{j_2}} \tag{3.1.16}$$

$$M^j = \sum_{i=1}^{n} m_i^j \tag{3.1.17}$$

burada $m_i^{j_1}$ – j_1- ci istifadəçinin i-ci məlumata yazdığı şərhlərin sayı, $m_i^{j_2}$ – j_2-ci istifadəçinin i-ci məlumata yazdığı şərhlərin sayıdır. $\sum_{i=1}^{n} \left(m_i^{j_1} + m_i^{j_2} \right) * I(c_i^{j_1}) \cdot I(c_i^{j_2})$ – j_1 və j_2 istifadəçilərinin birgə göründükləri məlumatlara yazdıqları şərhlərin ümumi sayı, M^j – j-ci istifadəçi tərəfindən yazılan şərhlərin ümumi sayıdır.

Təklif olunan **ikinci yanaşma**da neqativ sinfə daxil olan istifadəçilərin yazdığı şərhlərin semantik yaxınlığı əsasında sosial şəbəkənin aktorları arasında münasibətlərin müəyyən olunması nəzərdə tutulur. Aydındır ki, sosial şəbəkələrin analizində yaxınlıq ölçülərinin böyük rolu vardır. Yaxınlıq ölçüsü vasitəsilə istifadəçilər arasında əlaqənin gücü haqqında mühakimə yürütmək mümkündür. Bu məqalədə şərhlər arasındakı yaxınlığı hesablamaq üçün Jakkard ölçüsündən istifadə olunması təklif olunur:

$$w_2^{j_1 j_2} = sim(c^{j_1}, c^{j_2}) = \frac{\left| c^{j_1} \cap c^{j_2} \right|}{\left| c^{j_1} \cup c^{j_2} \right|} \qquad (3.1.18)$$

burada $sim(c^{j_1}, c^{j_2})$ – j_1 və j_2 istifadəçilərinin yazdığı şərhlər arasında semantik yaxınlıqdır.

Beləliklə, gizli sosial şəbəkənin aktorları arasında münasibətləri aşkarlamaq üçün yuxarıda təklif etdiyimiz üsulların xətti kombinasıyasından istifadə olunur:

$$w^{j_1 j_2} = \alpha \cdot \widehat{w}_1^{j_1 j_2} + (1-\alpha) \cdot w_2^{j_1 j_2} \qquad (3.1.19)$$

burada $\alpha \left(0 \le \alpha \le 1\right)$ çəki parametridir.

Qurulmuş gizli sosial şəbəkədə istifadəçilər arasında olan münasibətləri daha dəqiq təyin etmək üçün əlavə informasiya mənbələrindən istifadə oluna bilər. Bunlara e-poçt, tvitlər, çatlar və s.daxildir. Onları hansı əlaqələrin bağladığını müəyyən etməklə qurduğumuz sosial şəbəkənin adekvatlığını, istifadəçilər arasında olan münasibətlərin gücünü artırmaq olar.

3.2 Aşkarlanmış sosial şəbəkələrin analizi üçün alqoritm

Dövlətin mühüm vəzifələrindən biri e-dövlət mühitində dövlət əleyhinə təbliğatın yayılmasının qarşısının alınması, fəaliyyət göstərən gizli sosial şəbəkələrin aşkarlanması və idarəetmənin yaxşılaşdırılmasından ibarətdir. Yuxarıda qeyd etdiyimiz kimi e-dövlətin idarə edilməsində sosial şəbəkələr mühüm rol oynayır. İnternetdə isə müxtəlif sosial şəbəkələr fəaliyyət göstərir. Bu sosial şəbəkələrin bir qismi xoş məramlı, digər qismi xoş məramlı olmaya bilər. Xoş məramlı dedikdə elə qruplar nəzərdə tutulur ki, onlar e-dövlət sistemini izləyərək, mövcud problemləri

aşkar edərək, onun üzərində bir sıra işlər həyata keçirirlər. Xoş məramlı olmayan dedikdə isə dövlətə qarşı təbliğat aparmaq, onu zəiflətmək, xalq arasında narazılıqların artmasına yönəlmiş və s. məqsədli qruplar nəzərdə tutulur. Əlavə olaraq, sosial şəbəkələr aşkar və qeyri-aşkar fəaliyyət göstərə bilərlər. Bu baxımdan dövlətin təhlükəsizliyinin təmin olunması üçün sosial şəbəkələrin analizi mühüm əhəmiyyət kəsb edir.

Hazırda sosial şəbəkələrin əhəmiyyəti artmaqdadır, onlar hüquq-mühafizə və xüsusi xidmət orqanları ilə birlikdə kriminal qrupların aşkarlanması, onlar arasında olan gizli əlaqələri və ya detalları öyrənmək üçün istifadə olunur. E-dövlət mühitində münasibətlərin mənzərəsini təsvir etmək üçün faydalı vasitələrdən biri sosial şəbəkə analizidir [145]. Sosial şəbəkə analizi sosial şəbəkələrdə qrupları, gizli strukturları müəyyən etmək üçün istifadə oluna biləcək güclü metodların məcmusudur. Eyni zamanda sosial şəbəkə analizi hakimiyyət, nüfuz və etimad anlayışlarının araşdırılması üçün ilkin mənbə kimi istifadə oluna biləcək tədbirlər çoxluğundan ibarətdir [46].

Bölmə 3.1-də e-dövlət mühitində vətəndaşların yazdıqları şərhlər əsasında gizli sosial şəbəkələrin aşkar olunması üçün təklif olunmuş metod ətraflı şəkildə izah edilmişdir. Bu bölmədə isə növbəti mərhələ olaraq qurulmuş sosial şəbəkənin analizi həyata keçirilir. Analiz dedikdə şəbəkədə əsas aktorların, onların əhəmiyyətlilik dərəcəsinin müəyyən olunması və s. nəzərdə tutulur. Qurulmuş sosial şəbəkədə əsas aktorların müəyyən olunması üçün nüvənin nə dərəcədə kompakt olduğunu göstərmək lazımdır. Bunun üçün qurulmuş sosial şəbəkədə istifadəçilər və onlar arasındakı əlaqələrin sayından istifadə olunması təklif olunur. İlk növbədə qurulmuş sosial şəbəkənin istifadəçilərinin sayı aşağıdakı düstur vasitəsilə müəyyən olunur:

$$N_{\Sigma}^{-} = \left| U_{\Sigma}^{-} \right| \tag{3.2.1}$$

Daha sonra sosial şəbəkənin nüvəsinə daxil olan istifadəçilərin sayı təyin olunur:

$$N_{\Pi}^{-} = \left| U_{\Pi}^{-} \right| \tag{3.2.2}$$

Sosial şəbəkəyə daxil olan istifadəçilərin sayı təyin olunduqdan sonra onlar arasındakı əlaqələrin sayını müəyyən etmək üçün aşağıdakı düsturdan istifadə olunması təklif olunur:

$$M_{\Sigma}^{-} = \sum_{j_1, j_2 \in U_{\Sigma}^{-}} I_1(w^{j_1 j_2}) \tag{3.2.3}$$

$$M_{\Pi}^{-} = \sum_{j_1, j_2 \in U_{\Pi}^{-}} I_1(w^{j_1 j_2}) \tag{3.2.4}$$

burada M_{Σ}^{-} – sosial şəbəkəyə daxil olan istifadəçilər arasında əlaqələrin ümumi sayı, M_{Π}^{-} –sosial şəbəkənin nüvəsinə daxil olan istifadəçilər arasında əlaqələrin sayıdır. $I_1(w^{j_1 j_2})$ funksiyası aşağıdakı şəkildə təyin olunur:

$$I_1(x) = \begin{cases} 1, & \text{əgər} \quad x > 0 \\ 0, & \text{əgər} \quad x = 0 \end{cases} \tag{3.2.5}$$

Daha sonra bütün şəbəkənin sıxlıq əmsalı aşağıdakı düsturun köməyilə təyin olunur:

$$\sigma_{\Sigma}^{-} = \frac{M_{\Sigma}^{-}}{\frac{N_{\Sigma}^{-}(N_{\Sigma}^{-}-1)}{2}} \tag{3.2.6}$$

burada $\frac{N_{\Sigma}^{-}(N_{\Sigma}^{-}-1)}{2}$ – sosial şəbəkənin aktorları arasında bütün mümkün əlaqələrin sayıdır.

Eyni qayda ilə nüvədə sıxlıq əmsalı təyin olunur:

$$\sigma_{\Pi}^{-} = \frac{M_{\Pi}^{-}}{\frac{N_{\Pi}^{-}(N_{\Pi}^{-}-1)}{2}} \tag{3.2.7}$$

burada $\frac{N_{\Pi}^{-}(N_{\Pi}^{-}-1)}{2}$ – nüvənin aktorları arasındakı bütün mümkün əlaqələrin sayıdır.

(3.2.6) və (3.2.7) düsturlarının köməyilə nüvənin bütün sosial şəbəkədəki payı (çəkisi) aşağıdakı düstur vasitəsilə təyin olunur:

$$\sigma^{-} = \frac{\sigma_{\Pi}^{-}}{\sigma_{\Sigma}^{-}} \tag{3.2.8}$$

burada σ^{-} – nüvənin bütün sosial şəbəkədəki payı (çəkisi)-dir. Bunun əsasında nüvənin kompaktlığı müəyyən olunur.

Nüvənin kompaktlığı müəyyən olunduqdan sonra nüvəyə daxil olan aktorları ranqlaşdırmaq üçün aktorlar arasındakı münasibətlərin çəkisindən və əlaqələrin sayından istifadə olunur. Bunun üçün [123]-də təsvir olunan metod vasitəsilə aşağıdakı düsturlardan istifadə olunması təklif olunur:

$$c_j^{wa} = k_j^{(1-\alpha)} \cdot s_j^{\alpha}, \quad 0 \le \alpha \le 1 \tag{3.2.9}$$

$$k_j = \sum_{l \in U_{\Sigma}^{-}} I_1(w^{jl}) \tag{3.2.10}$$

$$s_j = \sum_{l \in U_{\Sigma}^{-}} w^{jl} \tag{3.2.11}$$

burada $c_j^{w\alpha}$ – mərkəzilik dərəcəsinin ölçüsü, k_j – nüvənin j-ci aktoru ilə şəbəkəyə daxil olan digər aktorlar arasında əlaqələrin sayları cəmi, s_j – isə uyğun olaraq əlaqələrin çəkiləri cəmi və α – çəki parametridir. Qeyd edək ki, ranqlaşdırma (3.2.9) -a əsasən aktorların əhəmiyyətlilik dərəcəsinə görə azalma sıra ilə (yəni, ən əhəmiyyətli aktordan ən az əhəmiyyətli aktora doğru) aparılır.

Beləliklə, e-dövlət mühitində vətəndaşların rəyləri əsasında dövlət əleyhinə təbliğat törətməkdə şübhəli bilinən gizli sosial şəbəkələr, bu şəbəkəyə daxil olan əsas aktorlar və onların əhəmiyyətlilik dərəcəsi müəyyən olundu.

Təklif olunan yanaşma konseptual xarakterlidir və onun realizasiyası üçün kompleks məsələlər həll olunmalıdır. İlk növbədə şərhlərin müxtəlif dillərdə yazılması ehtimalı var. Bu halda çox dilli mühit üçün mətnlərin analizi texnologiyalarının işlənməsi zərurəti meydana çıxır. Digər tərəfdən, təklif olunmuş yanaşmada sosial şəbəkə yalnız bir informasiya mənbəyindən istifadə olunmaqla sintez olunur. Lakin tədqiqatlar göstərir ki, bu cür sintez olunan sosial şəbəkələr real vəziyyəti əks etdirmir. Ona görə də əlavə informasiya mənbələrinə ehtiyac yaranır. Bu halda isə yeni problem – müxtəlif mənbələrdən alınmış informasiyanın etibarlılıq

dərəcəsinin müəyyən olunması meydana çıxır. Burada, həllini gözləyən bir neçə problem var:

- ✓ İnformasiyanın toplanması;
- ✓ İnformasiyanın müxtəlif dillərdə olması;
- ✓ Bir informasiya mənbəyinin kifayət etməməsi;
- ✓ Bu informasiya mənbələrinin etibarlılığı.

Beləliklə, sonda belə qərara gəlmək olur ki, təklif olunmuş yanaşma çərçivə yanaşmadır və onun ayrı-ayrı detallarının dəqiqləşdirilməsi gələcək tədqiqatların işidir.

IV FƏSİL. E-DÖVLƏTDƏ ƏKS-ƏLAQƏ MEXANİZMLƏRI ÜÇÜN METOD VƏ ALQORİTM

E-dövlət xidmətləri ictimaiyyətin maraqlarına əsaslanaraq onların ehtiyaclarını maksimum dərəcədə nəzərə almaqla, xidmətlərin səmərəliliyini artırmalı, xidmət xərclərini azaltmalı və keyfiyyətini artırmalıdır. E-dövlət mühitində insanların maraq dairəsində olan məsələlərin vaxtında müəyyən olunması dövlət orqanlarına xidmətlərin keyfiyyətini yaxşılaşdırmağa, vətəndaş məmnuniyyətini artırmağa kömək edə bilər. O cümlədən, təklif olunan e-xidmətlərin tələbyönümlü prinsipə əsaslanması ilə vətəndaşların real ehtiyacları təmin oluna bilər. Bu məqsədə nail olmaq üçün dövlət qurumları vətəndaşların informasiya ehtiyacları barədə məlumatlarla bağlı interaktiv olmalı və ictimaiyyətin real ehtiyacını öyrənməlidir. Bu mühitdə hotspot xidmətlərin təyin olunması, bu xidmətlərə hansı regionların tələbatı olduğunu müəyyən etməklə vətəndaş məmnunluğu təmin oluna bilər. Bundan əlavə, vətəndaşların yazdığı şərhlər əsasında qısa zaman müddətində hansı mövzular ətrafında fikir mübadiləsi aparıldığı müəyyən oluna bilər. Bunun üçün mövzu modelləşdirmə və klasterləşmə metodlarından istifadə oluna bilər. Lakin burada ölçü, həm də semantik cəhətdən yaxın sənədlərin müxtəlif klasterlərdə toplanması problemləri meydana gələ bilər. Məqsədimiz bu problemlərin aradan qaldırılmasından ibarətdir.

4.1. E-vətəndaşların şərhlərinin analizi üçün metod

Hazırki dövrdə onlayn forumlar, bloqlar, mikrobloqlardan istifadə edən vətəndaşların sayı günbəgün artmaqdadır. Bu mənbələrdən istifadə etməklə vətəndaşlar müəyyən məsələlər barəsində fikirlərini ifadə edə bilirlər. Bu cür platformalardan biri də e-dövlətdir. E-dövlət inkişaf etmiş ölkələrdə e-demokratiya, xidmət və resursların inteqrasiyası kimi xarakterizə olunur. E-dövlət portalı ictimai xidmətlərin göstərilməsi və vətəndaş-dövlət qarşılıqlı əlaqəsinin qurulması üçün ən vacib kanallardan biri hesab olunur. Bu portalda dövlət qurumları tərəfindən təqdim olunan çoxsaylı zəngin informasiya resursları mövcuddur. Bu resurslardan istifadə etməklə cəmiyyətin narahat olduğu əsas məsələləri operativ və vaxtında müəyyən

etmək mümkündür. Belə ki, istifadəçilərin bu resurslara sərf etdiyi vaxt, müraciətlərin sayı, o cümlədən müəyyən xidmət haqqında yazılan rəylər və s. vasitəsilə insanların narahat olduğu məsələlər, həmçinin onların maraq dairəsində olan informasiyalar aşkarlana bilər. Ədəbiyyatda insanların daha çox maraqlandığı informasiyalar hotspot informasiya adlandırılır. Qeyd edək ki, hotspot təkcə informasiya yox, müəyyən hadisələr, forumlar, şəhərlər, xəstəliklər, biznes və s. də ola bilər. Məsələn, xəstəliklərin hotspotlarını aşkarlamaqla insanların ən çox əziyyət çəkdiyi xəstəliklər, hadisələrin hotspotu vasitəsilə müəyyən hadisələrin (yanğınlar və s.) ən çox baş verdiyi sahələr və onların səbəbləri müəyyənləşdirilə bilər. Bundan başqa yuxarıda qeyd etdiyimiz kimi hotspot informasiya anlayışı da mövcuddur ki, bu da son illərdə populyarlıq qazanmışdır [162].

Onlayn mühitdə hotspot informasiyanın aşkarlanmasının bir sıra üstünlükləri mövcuddur və aşağıdakı kimi qeyd edilir [164, s.11]:

- İstifadəçinin hər hansı zaman müddətində müəyyən əmtəə və xidmətlər haqqında doğru qərar qəbul etməsinə kömək edə bilər;
- Hökumətin müvafiq bölmələrinə (qurumları/strukturlarına) kömək məqsədilə cəmiyyətdə ictimai marağın istiqamətinin müəyyənləşdirilməsi və rəhbərliyə yönəldilməsi, internet istifadəçilərinin fikirlərinin "sağlam" istiqamətə yönəldilməsini təmin edə bilər;
- İnsanların ən çox narahat olduğu problemlərin müəyyən edilməsində, xidmətlərin yaxşılaşdırılmasında, vətəndaş məmnuniyyətinin təmin olunmasında dövlət orqanlarına hər bir periodda kömək edə bilər.

Hazırda hotspot informasiyanın aşkar olunması ilə əlaqəli bir sıra problemlər mövcuddur. Belə ki, veb səhifələrin sayı artdıqca hotspotları çıxarmaq, onları klasterləşdirmək çətin məsələlərdən birinə çevrilir. Bu isə tədqiqatların dərinləşdirilməsini tələb edir.

İnternetdə hotspot informasiyanın təyin olunması aktual mövzulardan biri hesab olunur. Bu istiqamət üzrə bir sıra işlər həyata keçirilmişdir. Bu işlərdən bəziləri aşağıda nəzərdən keçirilmişdir.

Müəyyən müddət ərzində internetdə hotspot hadisələrin avtomatik təyini üçün [172]-də sistem təklif olunmuşdur. [155]-də Twitter platforması analiz edilmiş və xalq arasında müəyyən hadisələrin baş vermə səbəbləri öyrənilmişdir. [44]-də onlayn rəylərin istehlakçılara məhsul haqqında vacib məlumatlar verdiyi qeyd olunmuş və statistik təhlilə əsaslanan yanaşma təklif olunmuşdur. [104]-də aparılan araşdırmalarda onlayn-hotspot forumların proqnozlaşdırılması üçün iki maşın təlim üsulu, k-means və SVM təklif olunmuşdur. Bu məqalədə k-means hər zaman pəncərəsi daxilində bütün forumlar üçün klasterləşmə mənzərəsini əldə etmək üçün istifadə olunmuş və hotspot forumlar olaraq klasterləşmənin mərkəzinə daha yaxın olan forumlar qəbul olunmuşdur. SVM isə əvvəlki zaman pəncərəsində olan məlumatlardan istifadə olunaraq cari vaxt pəncərəsi üçün hotspot forumların proqnozlaşdırılması üçün istifadə olunmuşdur. [155]-də hotspot forumların proqnozlaşdırılması üçün semantikaya əsaslanan yanaşma təklif olunmuşdur. Bu məqalədə mətnin emosional polyarlığı hesablandıqdan sonra forumların klaster analizi üçün k-means və SVM təsnifat metodları birləşdirilmişdir. [98]-də aparılan araşdırmalar nəticəsində şəhərin hotspot nöqtələrini müəyyən etmək üçün metod təklif olunmuşdur. Burada şəhərin hotspot nöqtəsi olaraq hadisələrin sıx şəkildə baş verdiyi sahə qəbul olunmuşdur. Bu məqsədlə taksi trayektoriyası məlumatlarından istifadə olunmuş və məlumatların sahəyə əsaslanan klaster təhlili metodu (data field-based cluster analysis technique) zaman çəkisi artırılaraq təkmilləşdirilmişdir. Belə ki, burada trayektoriya nöqtələri çox olan sahə hotspot, nazik trayektoriyalı sahə isə daha aşağı potensiala malik sahə kimi təyin olunmuşdur. [64]-də mobil telefonların məkan paylanmasını aşkar etmək üçün araşdırmalar aparılmışdır. Burada sıxlıq (ing. kernel) metodu [56, s.474] tətbiq edilərək paylanmalar hesablanmış, məkan obyektlərinin çəkili qonşu siyahısı qurulmuş və məkan asılılıqlarını hesablamaq üçün maraqların dəyişkənliyi ilə bağlı müxtəlif avtokorelyasiya testləri həyata keçirilmişdir. Həmçinin mobil qüllələrin bir neçə xüsusiyyətindən istifadə etməklə onlar arasında hotspotlar müəyyən olunmuşdur. [146]-da Twitter məlumatlarından istifadə edərək hotspotların aşkarlanması üçün yanaşma təklif olunmuşdur. Burada verilənlər üzərində müxtəlif klasterləşmə alqoritmlərinin (iyerarxik aglomerativ

klasterləşmə və DBSCAN) üstünlükləri və mənfi cəhətləri müzakirə olunmuş, həmçinin müəyyən olunmuş sosial hotspotlara istifadəçilərin münasibətini müəyyən etmək məqsədilə sentiment analizdən istifadə olunaraq yanaşma təklif olunmuşdur.

Göründüyü kimi hotspot məlumatların aşkarlanması məqsədilə bir sıra işlər həyata keçirilmişdir. Bu istiqamətdə aparılan araşdırmalar diqqət mərkəzindədir və tədqiqatların genişləndirilməsinə ehtiyac vardır.

E-dövlət xidmətlərindən vətəndaş məmnuniyyəti: Dövlət orqanları veb texnologiyaların əhəmiyyətini qəbul edir və vətəndaşların bu xidmətlərdən yararlanması üçün bir çox işlər həyata keçirilir. E-dövlət də bu məqsədlə həyata keçirilmiş layihələrdən biridir. E-dövlət dünya ölkələrində dövlətin ən proqressiv İT-xidmətlərdən və onlayn-proqramlardan istifadəsi üçün düzgün metod hesab olunur [80].

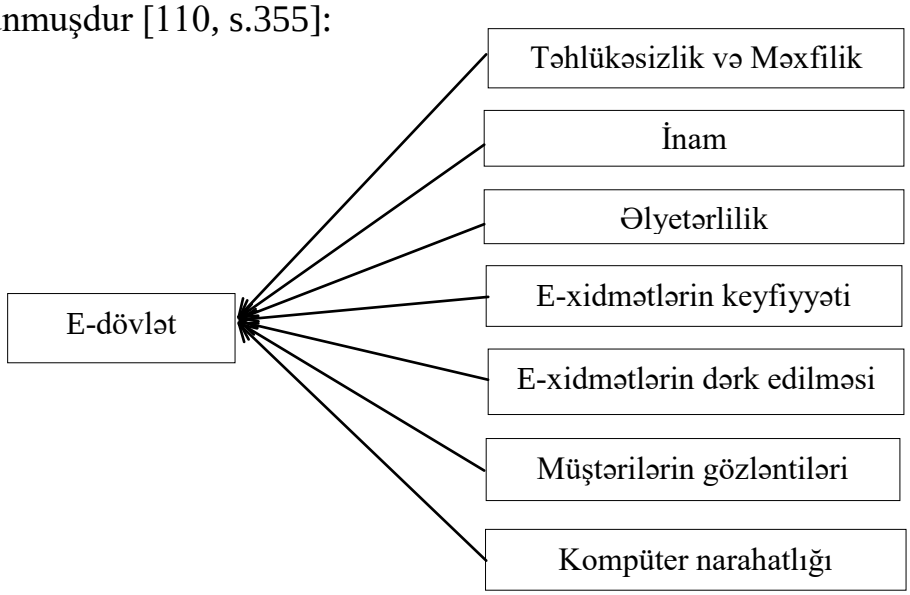
Təhlükəsizlik, əlyetərlilik, etibarlılıq, məxfilik, vətəndaş xidmətlərinin keyfiyyəti kimi məsələlər istənilən e-dövlət paradiqmasında əhəmiyyətlidir. E-dövlətin bütün dünyada, xüsusilə Asiyada sürətli təkamülü onun bir sıra əlverişli üstünlükləri ilə bağlıdır. Bunlara aşağı xərclə yaxşılaşdırılmış infrastruktur, daha yaxşı hökumət göstəriciləri, elastiklik, şəffaflıq, məsuliyyət, onlayn xidmətlərin geniş miqyası və s. daxildir. E-vasitələrlə məlumatların çatdırılması e-dövlət strategiyasının əsas komponentlərindən biri hesab olunur. Bununla belə, istifadəçiləri təkcə informasiya ilə təmin etmək kifayət deyil, dövlət xidmətlərindən vətəndaşların məmnunluğunu təmin etmək əsas məsələlərdən biri hesab olunur [34].

Vətəndaş məmnuniyyəti dedikdə, "bu və ya digər vəziyyətə təsir edən bir sıra amillərə qarşı hisslərin və münasibətlərin məcmusu" nəzərdə tutulur. İnformasiya sistemində vətəndaş məmnuniyyətinin ölçülməsi sistemin zəif və güclü tərəflərini istifadəçi baxımından qiymətləndirməyə imkan verir. Bundan əlavə, istifadəçilərin məmnunluğu dedikdə, sistemin texniki keyfiyyətindən daha çox, istifadəçilərin informasiya sistemini necə qiymətləndirməsi nəzərdə tutulur.

E-dövlət xidmətlərinin səviyyəsini qiymətləndirməklə dövlət vətəndaşlar qarşısında olan borcunu effektiv şəkildə həyata keçirə bilər [29, s.39]. E-dövlətin effektivliyinin ölçülməsində istifadəçi məmnuniyyətinin parametr olaraq təyin

edilməsinin bir sıra səbəbləri mövcuddur. Birincisi, sosial medianın istifadəçilər arasında perspektivli və populyar olmasıdır. Belə ki, sosial media kontekstində sosiallaşdırma və insanları birləşdirmə xarakteristikası istifadəçi məmnuniyyəti vasitəsilə daha yaxşı təyin oluna bilər. İkincisi, istifadəçi məmnuniyyəti e-dövlətin qəbul edilməsi və həyata keçirilməsi məqsədini daşıyır. Belə ki, istifadəçi məmnuniyyətinin ölçülməsi e-dövlət layihələrinin məqsədini özündə əks etdirir. Üçüncüsü, sistem istifadəçiləri tərəfindən yaxşı istifadə edilmədikdə, ümumiyyətlə, İT uğurlu hesab olunmur. İstifadəçi məmnuniyyətinin ölçülməsi vasitəsilə sistemin istifadəçilər arasında necə qəbul olunduğunu müəyyən etmək mümkündür. Dördüncüsü, vətəndaş məmnunluğunun təmin olunması vasitəsilə e-dövlətdən istifadənin artması dövlət proseslərinin və əməliyyatlarının xərclərinin azaldılmasına səbəb ola bilər. Beşincisi, informasiya sisteminin effektivliyinin ölçülməsində vətəndaş məmnuniyyətinin seçilməsi vasitəsilə e-dövlət layihələrinin tam effektiv qiymətləndirilməsi həyata keçirilə bilər [45].

E-dövlət portalında vətəndaş məmnuniyyətinin əsas göstəriciləri şəkil 4.1.1-də təsvir olunmuşdur [110, s.355]:



Şəkil 4.1.1. E-dövlətdə vətəndaş məmnuniyyətinin əsas göstəriciləri

E-dövlət portalının vətəndaşların informasiya ehtiyaclarını təmin etmək imkanı istifadəçilərin portalı necə qəbul etməsi, yenidən nəzərdən keçirməsi və onların məmnuniyyət dərəcəsinin müəyyənləşdirməsinə təsir göstərir. Beləliklə, vətəndaşların məmnuniyyətini ölçmək portalın müvəffəqiyyətsizliyini və ya bir

istifadəçi baxımından uğursuzluğunu müəyyənləşdirməyə kömək edə və portalın təkmilləşdirilməsinə, portal istifadəçilərinin sayının artmasına səbəb ola bilər. Araşdırmaların əksəriyyəti e-dövlət portallarının texniki dizaynı, xidmət keyfiyyəti və ya təklif olunan qiymətləndirmə metodlarına yönəlmişdir. E-dövlət portalının müvəffəqiyyətini istifadəçi perspektivindən araşdırmaq üçün tədqiqatların aparılmasına ehtiyac var.

E-dövlətdə hotspot informasiyadan istifadə etməklə vətəndaş məmnuniyyətinin təyin olunması: E-dövlət son illərdə istifadəsi rahat və asan olması səbəbindən tez bir zamanda sosial medianın populyar və ümumi forması olaraq ortaya çıxmışdır. Bu platformadan istifadə edən insanlar günbəgün artmaqdadır. Burada insanların maraq dairəsini, o cümlədən regionların e-xidmətlər üzrə ümumi mənzərəsini, onların xidmətlərdən məmnuniyyətini müəyyənləşdirmək mümkündür [24].

Məlumdur ki, e-dövlət portalı ictimai xidmətlərin çatdırılması və dövlət-vətəndaş münasibətlərində əsas kanallardan biridir. Bu portalda dövlət orqanları tərəfindən təqdim edilən çoxsaylı və zəngin informasiya resursları mövcuddur. Bu resurslardan isifadə etməklə insanların ən çox hansı problemlərdən əziyyət çəkdiyini, hansı xidmətlərə daha çox ehtiyacı olduğunu və regionlarda hansı məsələlərin aktual olduğunu müəyyən etmək mümkündür. Problemləri vaxtında müəyyən etməklə dövlət düzgün qərarlar qəbul edə və vətəndaşları razı sala bilər. Çünki dövlətin əsas məqsədi vətəndaşların məmnunluğunu təmin etmək, onların yaşayış səviyyəsini yüksəltmək və əsas narahatlıqlarını aradan qaldırmaqdır. E-dövlət portalında dövlət orqanları tərəfindən təqdim olunan xidmətlərin sayı günbəgün artmaqdadır. Bu xidmətlər arasında hotspot olanları müəyyən etməklə dövlət xidmətlərinin səmərəliliyini artırmış olarıq [164]. Bunları nəzərə alaraq, e-dövlət mühitində dövlət orqanları tərəfindən təqdim olunan xidmətlərdən vətəndaş məmnunluğunun müəyyən olunması, xidmətlər arasında hotspotların və bu hotspotlar ətrafında hansı regionların toplandığını müəyyən etmək üçün metod təklif edilmişdir. Qeyd edək ki, burada bir sıra məsələlər həll oluna bilər [77]:

- ✓ Dövlət orqanlarının ranqlaşdırılması (bəzi məlumatlar üzrə kliklərin sayı, sərf olunan orta zaman müddətinə görə);
- ✓ Hotspot məlumatların müəyyən edilməsi;
- ✓ Maraq dairəsinə görə regionların qruplaşdırılması;
- ✓ Hotspotlar və onların ətrafında toplanmış regionların müəyyən edilməsi;
- ✓ Şərhlərə əsasən vətəndaşlar və onların mənsub olduğu regionun qiymətləndirilməsi və s.

Təklif etdiyimiz metod haqqında ətraflı məlumat aşağıda verilmişdir.

Tutaq ki, e-dövlət portalında təklif olunan xidmətlərin sayı m və bu xidmətlərdən istifadə edən vətəndaşların sayı n-dir. Bunları uyğun olaraq $(EG_1, EG_2, ..., EG_m)$ və $(U_1, U_2, ..., U_n)$ ilə işarə edək. Qeyd edək ki, hər bir xidməti müəyyən zaman müddətində qiymətləndirmək mümkündür. Bunun üçün vətəndaşların xidmətə müraciətlərinin sayı və məmnunluq dərəcəsindən istifadə oluna bilər. Təklif etdyimiz metodda hər bir xidmət üzrə vətəndaşların məmnunluq dərəcəsini müəyyən etmək və hotspot xidmətləri tapmaq üçün müəyyən T – zaman periodu ərzində əvvəlcə hər bir istifadəçinin xidmətdən istifadə vektoru qurulur. Bunun üçün (4.1.1) bərabərliyindən istifadə olunur:

$$U_i = \{ u_{i1}, u_{i2}, ..., u_{im} \}, \quad i = 1,2,...,n \tag{4.1.1}$$

burada u_{ij} $(i = 1,2,...,n, \ \ j = 1,2,...,m)$ – i-ci istifadəçinin j-ci xidmətdən istifadəsini ifadə edir. Qeyd edək ki, burada hər bir u_{ij} elementi iki ölçülü vektordur:

$$u_{ij} = (u_{ij,1}, u_{ij,2}) \tag{4.1.2}$$

burada $u_{ij,1}$ – i-ci istifadəçinin j-ci xidmətə müraciətini, $u_{ij,2}$ – isə xidmətdən məmnunluq dərəcəsini ifadə edir. Qeyd edək ki, burada iki variant mümkündür: 1) Xidmətdən istifadə sayı nəzərə alınmır; 2) Xidmətdən istifadə sayı nəzərə alınır.

Birinci variantda xidmətdən istifadə sayı nəzərə alınmır. Yəni, istifadəçinin xidmətə müraciətlərinin sayı nəzərə alınmır, yalnız onun xidmətə müraciət edib-etməməsinə baxılır. Əgər vətəndaş xidmətə müraciət edibsə, bu 1 ilə, etməyibsə, 0 ilə qiymətləndirilir:

$$u_{ij,1} = \begin{cases} 1, & \text{əgər } i\text{-ci istifadəçi } j\text{-ci xidmətdən istifadə edibsə,} \\ 0, & \text{əks halda.} \end{cases} \quad (4.1.3)$$

Vətəndaşların xidmətlərdən məmnunluq dərəcəsini qiymətləndirmək üçün [1,5] şkalasından istifadə olunması təklif olunur:

$$u_{ij,2} = \begin{cases} 1 & \text{çox pis,} \\ 2 & \text{pis,} \\ 3 & \text{orta,} \\ 4 & \text{yaxşı,} \\ 5 & \text{çox yaxşı.} \end{cases} \quad (4.1.4)$$

Qeyd edək ki, əgər istifadəçi xidmətə müraciət etməyibsə, onda xidmətdən istifadə vektoru $u_{ij} = (0,0)$ qəbul olunur.

(4.1.2)-ni nəzərə alsaq, onda (4.1.1) matrisini aşağıdakı şəkildə yaza bilərik:

$$U = \begin{Vmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{n1} & \dots & u_{nm} \end{Vmatrix} = \begin{Vmatrix} (u_{11,1}, u_{11,2}) & \dots & (u_{1m,1}, u_{1m,2}) \\ \vdots & \ddots & \vdots \\ (u_{n1,1}, u_{n1,2}) & \dots & (u_{nm,1}, u_{nm,2}) \end{Vmatrix} \quad (4.1.5)$$

Hər bir istifadəçinin xidmətlər üzrə vektoru qurulduqdan sonra bu vektorların sətirlər üzrə uyğun elementlərini cəmləyə bilərik. Bu bizə hər bir xidmətə vətəndaşların nə dərəcədə tələbatı olduğunu və bu xidmətlərdən məmnunluğunu müəyyən etməyə kömək edə bilər. Bu halda hər bir e-xidmət $EG_j\ (j = 1,2,...,m)$, iki ölçülü vektor şəklində ifadə oluna bilər:

$$EG_j = \sum_{i=1}^{n} u_{ij} = \sum_{i=1}^{n} (u_{ij,1}, u_{ij,2}) = (u_{j,1}, u_{j,2}) \quad j = 1,2,...,m \quad (4.1.6)$$

burada vektorun birinci elementi $u_{j,1}$ – j-ci xidmətdən istifadə sayı, ikinci elementi $u_{j,2}$ – j-ci xidmətdən məmnunluq dərəcəsini ifadə edir. (4.1.6)-dan istifadə etməklə, j-ci xidmətdən orta məmnunluq dərəcəsini hesablaya bilərik. Bunun üçün aşağıdakı düsturdan istifadə olunması təklif olunur:

$$EG_j^{savg} = \frac{u_{j,2}}{u_{j,1}} \tag{4.1.7}$$

burada EG_j^{savg} - j-ci xidmətdən orta məmnunluq dərəcəsini ifadə edir. EG_j^{savg}-a görə ranqlaşdırma (azalan sıra ilə) aparsaq, xidmətlərin məmnunluq reytinqini almış olarıq.

İkinci variantda xidmətdən istifadə sayı nəzərə alınır. Yəni, ola bilər ki, eyni vətəndaş xidmətdən bir neçə dəfə istifadə etsin və hər dəfə xidməti müxtəlif məmnunluq balları ilə qiymətləndirsin. Bu halda e-xidmətlərin orta məmnunluq reytinqini aşağıdakı qayda ilə aparmaq olar.

İstifadəçinin xidmətə müraciətlərinin sayını nəzərə alsaq, xidmətdən istifadə vektorunu aşağıdakı şəkildə yaza bilərik:

$$u_{ij} = \left(u_{ij,1}, u_{ij,2}\right) = \left(N_{ij}, \sum_{k=1}^{N_{ij}} u_{ij,2}^k\right) = \left(N_{ij}, u_{ij,2}^{\Sigma}\right) \tag{4.1.8}$$

burada N_{ij} - i-ci istifadəçinin j-ci xidmətə müraciətlərinin sayı, $u_{ij,2}^k$ - k-cı dəfə müraciət etdikdə verdiyi məmnunluq balı, $u_{ij,2}^{\Sigma}$ – isə i-ci istifadəçinin j-ci xidmətə verdiyi ümumi məmnunluq balıdır. (4.1.8)-dən istifadə etməklə həmçinin i-ci istifadəçinin j-ci xidmətdən orta məmnunluq dərəcəsini tapa bilərik:

$$u_{ij}^{savg} = \frac{u_{ij,2}^{\Sigma}}{N_{ij}} \tag{4.1.9}$$

Bu halda (4.1.5) matrisi aşağıdakı şəkildə ifadə olunur:

$$U = \begin{Vmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{n1} & \dots & u_{nm} \end{Vmatrix} = \begin{Vmatrix} (N_{11}, u_{11}^{savg}) & \dots & (N_{1m}, u_{1m}^{savg}) \\ \vdots & \ddots & \vdots \\ (N_{n1}, u_{n1}^{savg}) & \dots & (N_{nm}, u_{nm}^{savg}) \end{Vmatrix} \tag{4.1.10}$$

Bunları nəzərə alsaq, (4.1.6)-nı aşağıdakı şəkildə ifadə edə bilərik:

$$EG_j = \sum_{i=1}^{n} u_{ij} = \left(\sum_{i=1}^{n} N_{ij}, \sum_{i=1}^{n} u_{ij,2}^{\Sigma}\right) = \left(N_j, u_{j,2}^{\Sigma}\right) \tag{4.1.11}$$

(4.1.11)-dən istifadə etməklə müraciətlərin sayını nəzərə almaqla, hər bir xidmətdən orta məmnunluq dərəcəsini tapa bilərik:

$$EG_j^{savg} = \frac{u_{j,2}^{\Sigma}}{N_j} \tag{4.1.12}$$

EG_j^{savg}-ə görə ranqlaşdırma aparsaq, xidmətlərin məmnunluq reytinqini almış olarıq.

Burada həmçinin hər bir istifadəçinin bütün xidmətlərdən orta məmnunluq dərəcəsini tapa bilərik:

$$u_i^{savg} = \frac{1}{m}\sum_{j=1}^{m} u_{ij}^{savg} \tag{4.1.13}$$

burada u_i^{savg} - i-ci istifadəçinin bütün xidmətlərdən orta məmnunluq dərəcəsini ifadə edir.

Bu halda bütün istifadəçilərin e-xidmətlərdən orta məmnunluq dərəcəsi aşağıdakı şəkildə ifadə olunur:

$$U^{savg} = \frac{1}{n}\sum_{i=1}^{n} u_i^{savg} \tag{4.1.14}$$

burada U^{savg} -istifadəçilərin e-xidmətlərdən orta məmnunluq dərəcəsini müəyyən edir. (4.1.14) vasitəsilə e-sistem ümumi qiymətləndirilir. Burada e-sistem dedikdə e-dövlət platforması nəzərdə tutulur.

Burada həmçinin hər bir xidmət üzrə regional qiymətləndirmə apara bilərik. Yəni, bu xidmətlərə daha çox hansı regionların müraciət etdiyi, o cümlədən razı qalıb-qalmadığını təyin edə bilərik. Bunun üçün u_{ij}^{savg}-dan istifadə oluna bilər. Belə ki, u_{ij}^{savg}-ə görə ranqlaşdırma aparsaq, istifadəçilərin j-ci xidmətdən məmnunluq reytinqini almış olarıq. Ranqlaşdırma aparıldıqdan sonra reytinq cədvəlini (4.1.4)-dən istifadə etməklə üç hissəyə bölə bilərik: U_j^- (məmnun deyil), U_j^0 (məmnundur), U_j^+ (çox məmnundur). Burada U_j^- - j -ci xidmətdən məmnunluq balları [1,2) intervalına

düşən, U_j^0 - məmnunluq balları [2,4) intervalına düşən, U_j^+ - [4,5) intervalına düşən istifadəçilər qrupudur.

İstifadəçi qrupları təyin olunduqdan sonra xidmətlər üzrə bu qrupların kəsişməsinə baxa bilərik:

$$U^- = \bigcap U_j^-$$

$$U^0 = \bigcap U_j^0 \qquad (4.1.15)$$

$$U^+ = \bigcap U_j^+$$

burada U^- - bütün xidmətlərdən narazı, U^0 - razı , U^+ - çox razı olan istifadəçilər qrupudur. Burada hər bir xidmətdən istifadə edən vətəndaşı regiona bağlaya bilərik. Belə ki, hər bir istifadəçinin arxasında IP ünvan olduğunu nəzərə alsaq, (4.1.15) düsturu vasitəsilə bütün xidmətlərdən narazı və razı olan vətəndaşların hansı regiona aid olduğu avtomatik olaraq müəyyən olunur. Ola bilsin ki, bu vətəndaşlar bir regiona toplansın və ya regionlar üzrə paylanmış olsunlar.

Həmçinin regionların ümumi maraq dairəsini təyin edə bilərik. Yəni, hər bir regionun istifadə etdiyi xidmətləri və bu xidmətlərin reytinqini müəyyənləşdirməklə, regionlar üçün hotspot xidmətləri təyin etmiş olarıq. Bunun üçün $(U_1, U_2, ..., U_n)$ istifadəçiləri yuxarıda qeyd olunduğu kimi IP vasitəsilə regionlara bölürük. (4.1.8)-dən istifadə etməklə hər bir istifadəçinin (regionun) istifadə etdiyi xidmətlər müəyyən olunur. Eyni regiondan olan istifadəçilərin müraciət etdiyi xidmətləri ranqlaşdırsaq (azalma sırası ilə), xidmətlərin istifadə reytinqini alarıq. Bu reytinqə əsasən eyni regiondan olan istifadəçilərin ən çox müraciət etdiyi xidmətlər müəyyən olunur. Buradan isə avtomatik olaraq regionların e-xidmət maraqları təyin olunmuş olur.

Göründüyü kimi təklif olunmuş metodda regionların xidmətlərdən razı qalıb-qalmadığı, o cümlədən hansı xidmətlərə tələbatı olduğu müəyyən olunur. Bu gələcəkdə regionlarda vətəndaş məmnuniyyətinin artırılması, dövlətə inamın yüksəldilməsində əsas rol oynaya bilər. Belə ki, bu problemləri bilməklə dövlət vaxtında düzgün qərarlar verməklə vətəndaşların məmnunluğunu təmin edə bilər.

Eksperiment

Təklif olunan metodu qiymətləndirmək üçün hesablamalar *Matlab 2018* proqramında yerinə yetirilmişdir.

Tutaq ki, müəyyən zaman intervalında e-xidmətlərə olan müraciətlərin sayı və bu xidmətlərdən məmnunluq dərəcəsini ifadə edən ballar cədvəl 4.1.1-də təsvir olunmuşdur:

Cədvəl 4.1.1

İstifadəçilərin xidmətlərə müraciət sayı və verdiyi ümumi bal

İstifadəçilər	***E-xidmətlər***				
	EG_1	EG_2	EG_3	EG_4	EG_5
U_1	(41, 82)	(33, 44)	(22,80)	(38,76)	(17,34)
U_2	(46,153)	(1, 4)	(19,69)	(13,43)	(42,112)
U_3	(6,16)	(43,129)	(39,104)	(25,100)	(29,106)
U_4	(46,107)	(47,156)	(40,106)	(35,128)	(28,56)
U_5	(32,64)	(34,124)	(9,30)	(45,45)	(46,214)
U_6	(4,13)	(38,152)	(24,64)	(48,80)	(14,46)
U_7	(14,56)	(37,86)	(22,58)	(27,90)	(38,126)
U_8	(27,108)	(20,40)	(32,96)	(7,32)	(38,139)
U_9	(48,192)	(33,88)	(36,72)	(7,11)	(19,76)
U_{10}	(49,130)	(8,32)	(38,126)	(13,43)	(28,46)
U_{11}	(8,24)	(36,96)	(14,46)	(42,98)	(3,11)
U_{12}	(49,98)	(1,3)	(34,90)	(12,32)	(2,7)
U_{13}	(48,160)	(14,28)	(33,132)	(41,123)	(27,117)
U_{14}	(24,88)	(2,7)	(8,26)	(12,32)	(39,78)
U_{15}	(40,146)	(4,5)	(6,20)	(47,125)	(47,141)
U_{16}	(7,21)	(41,41)	(25,91)	(17,39)	(6,8)
U_{17}	(21,56)	(35,105)	(48,112)	(10,40)	(29,106)
U_{18}	(46,168)	(16,32)	(17,45)	(12,32)	(23,69)
U_{19}	(40,53)	(48,144)	(29,87)	(31,103)	(0,0)
U_{20}	(48,112)	(1,2)	(11,25)	(24,56)	(17,45)

Cədvəl 4.1.1-də verilən hər bir vektorun 1-ci ədədi hər bir istifadəçinin xidmətə olan ümumi müraciət sayı, 2-ci ədədi isə ona verdiyi ümumi məmnunluq balıdır. Qeyd edək ki, bu xidmətlərə müraciət edən istifadəçilərin mənsub olduğu regionlar cədvəl 4.1.2-də təsvir olunmuşdur:

Cədvəl 4.1.2

Regionlar üzrə istifadəçilərin paylanması

Regionlar	**İstifadəçilər**
R_1	$U_2, U_5, U_7, U_{15}, U_{14}$
R_2	U_1, U_{20}, U_9, U_{19}
R_3	U_8, U_{11}, U_{12}
R_4	U_3, U_4, U_{13}, U_{10}
R_5	$U_{16}, U_{17}, U_6, U_{18}$

Cədvəl 4.1.1-dən istifadə etməklə (4.1.9) düsturu əsasında hər bir istifadəçinin hər bir xidmətə verdiyi orta məmnunluq balları hesablanmışdır. Hesablamaların nəticələri cədvəl 4.1.3-də verilmişdir.

Cədvəl 4.1.3

Hər bir istifadəçinin xidmətlərdən orta məmnunluq balı

İstifadəçilər	EG_1	EG_2	EG_3	EG_4	EG_5
U_1	2.0000	1.3333	3.6364	2.0000	2.0000
U_2	3.3261	4.0000	3.6316	3.3077	2.6667
U_3	2.6667	3.0000	2.6667	4.0000	3.6552
U_4	2.3261	3.3191	2.6500	3.6571	2.0000
U_5	2.0000	3.6471	3.3333	1.0000	4.6522
U_6	3.2500	4.0000	2.6667	1.6667	3.2857
U_7	4.0000	2.3243	2.6364	3.3333	3.3158
U_8	4.0000	2.0000	3.0000	4.5714	3.6579
U_9	4.0000	2.6667	2.0000	1.5714	4.0000
U_{10}	2.6531	4.0000	3.3158	3.3077	1.6429
U_{11}	3.0000	2.6667	3.2857	2.3333	3.6667
U_{12}	2.0000	3.0000	2.6471	2.6667	3.5000
U_{13}	3.3333	2.0000	4.0000	3.0000	4.3333

Cədvəl 4.1.3-ün ardı

U_{14}	3.6667	3.5000	3.2500	2.6667	2.0000
U_{15}	3.6500	1.2500	3.3333	2.6596	3.0000
U_{16}	3.0000	1.0000	3.6400	2.2941	1.3333
U_{17}	2.6667	3.0000	2.3333	4.0000	3.6552
U_{18}	3.6522	2.0000	2.6471	2.6667	3.0000
U_{19}	1.3250	3.0000	3.0000	2.3333	0
U_{20}	2.3333	2.0000	2.2727	5.0000	2.6471

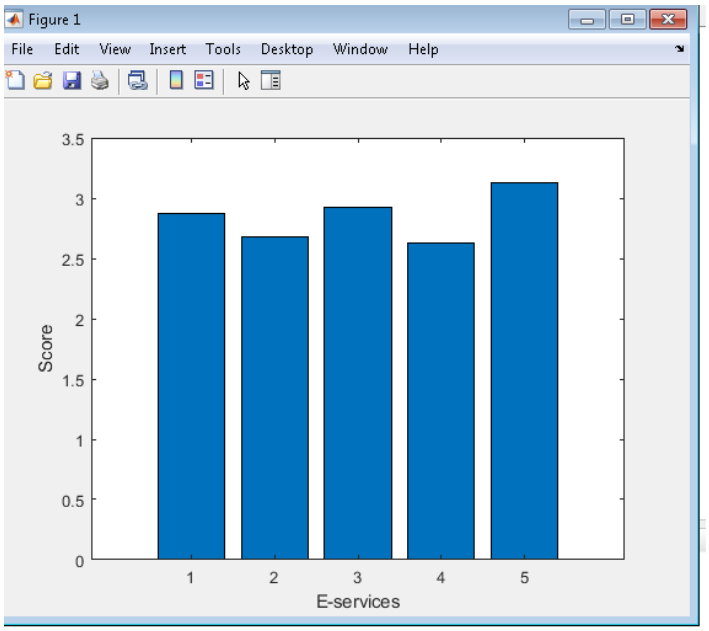
Daha sonra xidmətlərin məmnunluq reytinqini tapmaq üçün əvvəlcə (4.1.12) düsturu əsasında xidmətlərdən bütün istifadəçilərin orta məmnunluq balları hesablanmışdır. Bunun üçün hər bir xidmətə olan ümumi müraciət sayı və ona verilən ümumi məmnunluq balından istifadə olunmuşdur. Hesablamaların nəticələri cədvəl 4.1.4-də verilmişdir.

Cədvəl 4.1.4

Bütün istifadəçilər tərəfindən hər bir xidmətə verilən orta məmnunluq balları

E-xidmətlər	***Müraciət sayı***	***Ümumi bal***	***Orta məmnunluq balı***
EG_1	644	1847	2.8680
EG_2	492	1318	2.6789
EG_3	506	1479	2.9229
EG_4	506	1328	2.6245
EG_5	492	1537	3.1240

Cədvəl 4.1.4-dən göründüyü kimi xidmətlərə verilən orta məmnunluq balları bir-birinə yaxındır. Bu qiymətlərdən istifadə etməklə xidmətlərin məmnunluq reytinqini tapa bilərik (şəkil 4.1.2):

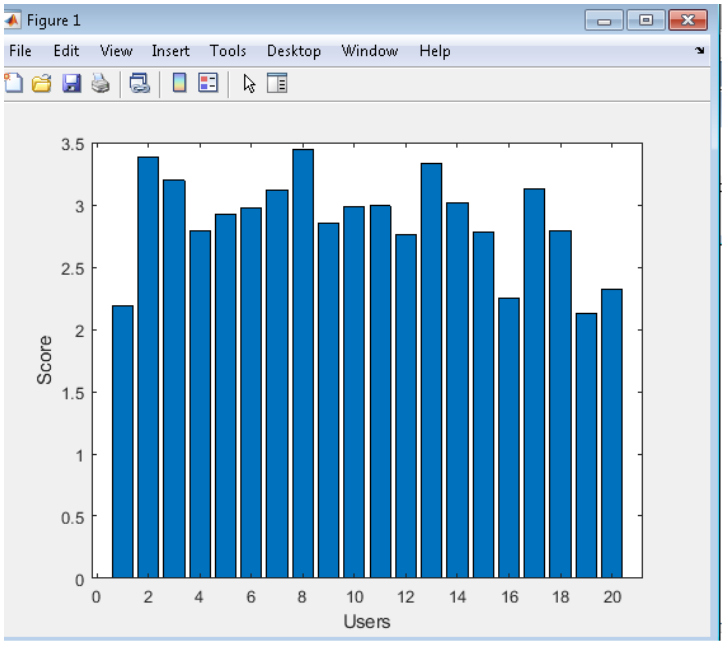


Şəkil 4.1.2. E-xidmətlərdən məmnunluq reytinqi

Daha sonra hər bir istifadəçinin müraciət etdiyi bütün xidmətlərdən ümumi məmnunluq dərəcəsi hesablanmışdır:

u_i^{savg} = [2.1939 3.3864 3.1977 2.7905 2.9265 2.9738 3.1220 3.4459 2.8476 2.9839 2.9905 2.7628 3.3333 3.0167 2.7786 2.2535 3.1310 2.7932 2.1295 2.3173]

Hesablamaların nəticələri şəkil 4.1.3-də təsvir olunmuşdur:



Şəkil 4.1.3. İstifadəçilərin reytinqi

Daha sonra bu hesablamalardan istifadə edərək e-sistem ümumi şəkildə qiymətləndirilmişdir:

$$U^{savg} = \frac{1}{20} * (2.1939 + 3.3864 + ... + 2.3173) = 2.8687$$

Sistemdən olan ümumi məmnunluq dərəcəsi $(U^{savg} = 2.8687) [2,4)$ intervalına düşdüyündən sistemin fəaliyyətini "qənaətbəxş" hesab etmək olar.

Bütün xidmətlərdən ancaq razı və narazı olan istifadəçi qruplarını tapmaq üçün cədvəl 4.1.3-dən istifadə etməklə ranqlaşdırma aparılmışdır. Bu ranqlaşdırma əsasında onların verdiyi bala uyğun olaraq məmnun olmayan, məmnun olan və çox məmnun olan istifadəçi qrupları təyin olunmuşdur (cədvəl 4.1.5).

Bütün xidmətlər üzrə bu qrupların kəsişməsinə baxdıqda, cədvəl 4.1.5-dən göründüyü kimi bütün xidmətlərdən məmnun olmayan və çox məmnun olan istifadəçilər olmadığı halda, məmnun olan istifadəçilər mövcuddur. Belə ki, $\{u_4, u_{11}, u_{12}, u_{14}, u_{18}\}$ istifadəçiləri bütün xidmətlərdən çox məmnun olan istifadəçilərdir. Cədvəl 4.1.2-də bu istifadəçilərin mənsub olduğu regionlara baxsaq, $\{R_1, R_3, R_4, R_5\}$ regionlarının bütün xidmətlərdən məmnun olduğunu görmüş olarıq.

Regionların maraq dairələrini müəyyən etmək üçün isə cədvəl 4.1.1-dən istifadə olunmuşdur. Belə ki, eyni regiondan olan istifadəçilərin ən çox müraciət etdiyi xidmətləri təyin etdikdən sonra ranqlaşdırma vasitəsilə regionlar üçün hotspot xidmətlər təyin olunmuşdur. Qeyd edək ki, cədvəldə M.sayı ilə müraciət sayı işarə olunmuşdur. Müraciət sayı dedikdə istifadəçilərin hər bir xidmətə olan müraciət sayından maksimum olanı götürülmüşdür. Bu hesablamaların nəticələri cədvəl 4.1.6-da təsvir olunmuşdur.

Cədvəl 4.1.5

Xidmətlər üzrə istifadəçi qruplarının məmnunluğu

	EG_1	EG_2	EG_3	EG_4	EG_5
Məmnun olmayanlar	U_{19}	U_1 U_{15} U_{16}	-	U_5 U_6 U_9	U_{10} U_{16} U_{19}
Məmnun olanlar	U_1 U_2 U_3 U_4 U_5 U_6 U_{10} U_{11} U_{12} U_{13} U_{14} U_{15} U_{16} U_{17} U_{18} U_{20}	U_2 U_3 U_4 U_5 U_7 U_8 U_9 U_{11} U_{12} U_{13} U_{14} U_{17} U_{18} U_{20}	U_1 U_2 U_3 U_4 U_5 U_6 U_7 U_8 U_9 U_{10} U_{11} U_{12} U_{14} U_{15} U_{16} U_{17} U_{18} U_{19} U_{20}	U_1 U_2 U_4 U_7 U_{10} U_{11} U_{12} U_{14} U_{15} U_{16} U_{18} U_{19}	U_1 U_2 U_3 U_4 U_6 U_7 U_8 U_9 U_{11} U_{12} U_{14} U_{15} U_{17} U_{18} U_{20}
Çox məmnun olanlar	U_7 U_8 U_9	U_2 U_6 U_{10}	U_{13}	U_3 U_8 U_{17} U_{20}	U_5 U_9 U_{13}

Cədvəl 4.1.6

Regionlar üzrə hotspot xidmətlər

E-xidmətlər	*Region 1*		*Region 2*		*Region 3*		*Region 4*		*Region 5*	
	M.sayı	*Ranq*	*M.sayı*	*Ranq*	*M.sayı*	*Ranq*	*M.sayı*	*Ranq*	*M.sayı*	*Ranq*
EG_1	46	***3***	48	***1***	49	***1***	49	***1***	46	***3***
EG_2	37	***4***	48	***2***	36	***4***	47	***2***	41	***4***
EG_3	22	***5***	36	***4***	34	***5***	40	***4***	48	***1***
EG_4	47	***2***	38	***3***	42	***2***	41	***3***	48	***2***
EG_5	47	***1***	19	***5***	38	***3***	29	***5***	29	***5***

Beləliklə, xidmətlərin məmnunluq reytinqi, istifadəçilərin bütün xidmətlərdən ümumi məmnunluq reytinqi, bütün xidmətlərdən ancaq məmnun olan və olmayan regionlar, bu regionların maraq dairələri və e-sistem ümumi şəkildə qiymətləndirilmişdir.

Eksperiment nəticəsində 5 xidmət üzrə 5 regiondan olan 20 istifadəçinin bu xidmətlərdən məmnunluq dərəcəsi, həmçinin bu regionların ümumi maraq dairələri müəyyən olunmuşdur. Hesablamalar nəticəsində e-xdimətlərin məmnunluq reytinqinin 1-1-nə yaxın olduğu aşkar olunmuşdur. Lakin istifadəçilərin bütün xidmətlər üzrə məmnunluq reytinqində fərqliliklərin olduğu müəyyən olunmuşdur. Sistemin ümumi fəaliyyəti, yəni sistemdən olan ümumi məmnunluq dərəcəsi "qənaətbəxş" kimi qiymətləndirilmişdir. Bütün xidmətlərdən məmnun olmayan və çox məmnun istifadəçilərin olmadığı aşkar olunmuşdur. Lakin məmnun olan istifadəçilərin olduğu və bu istifadəçilərin 4 region üzrə paylandığı təyin olunmuşdur. Sonda regionlar üzrə hotspot xidmətlər müəyyən olunmuşdur.

4.2. E-vətəndaşları maraqlandıran aktual mövzuların müəyyən olunması üçün alqoritm

Məlum olduğu kimi, e-dövlət hazırda istifadə olunan mühüm sosial platformalardan biridir. Bu platformada vətəndaş məmnuniyyətinin təmin olunması üçün təklif olunan e-xidmətlər tələbyönümlü prinsipə əsaslanmalıdır. Bu məqsədə

nail olmaq üçün dövlət qurumları vətəndaşların informasiya ehtiyacları barədə məlumatlarla bağlı interaktiv olmalı və ictimaiyyətin real ehtiyacını öyrənməlidir. Bunun üçün vətəndaş (istifadəçi) şərhlərindən istifadə oluna bilər. Lakin bu cür istifadəçi şərhləri geniş mətn məlumatlarının yaranmasına gətirib çıxara bilər ki, bu da onların mürəkkəb, struktursuz olması baxımından analizində müəyyən çətinliklər yaradır. Mətnlərin analizi, korrelyasiyaların aşkarlanması üçün text mining kimi qabaqcıl hesablama və analitik vasitələrdən istifadə etmək nəzərdə tutulur [74]. Vətəndaşların hansı mövzular ətrafında fikir mübadiləsi apardığını aşkar etmək üçün son zamanlar mövzu modelləşdirmə metodları populyarlıq qazanmışdır. Mövzu modeləşdirmə alqoritmləri mətnlərdə sözləri analiz edərək mövzuları aşkarlayan statistik metodlardır [71]. Maşın təlimi və text mining-də istifadə olunan bu metodlar sənədlər çoxluğunda gizli mövzuların müəyyən olunmasında uğurla tətbiq olunur. Bu alqoritmlər xüsusilə klassifikasiya (təsnifat) problemində sənədlərin siniflərə "yumşaq" (qeyri-səlis) təsnifatı kimi şərh edilir, yəni sənəd yalnız bir sinfə yox, bir neçə sinfə müxtəlif mənsubiyyət dərəcəsi ilə daxildir. Bundan başqa, bu modellərin nəticələri sənədi yalnız bir qrupa daxil etmək üçün də istifadə oluna bilər.

Yuxarıda qeyd etdiyimiz kimi mövzu modelləşdirmə metodlarına diqqət getdikcə artmaqdadır. Bununla əlaqədar olaraq, bu istiqamətdə aparılan tədqiqatlar da artmışdır. Belə ki, mövzu modelləşdirmə metodlarının yaxşılaşdırılması üçün bir sıra metodlar, yanaşmalar təklif olunmuşdur.

Məsələn, [43]-də aparılan tədqiqatlarda mətnləri təhlil etmək və Vikipediyada obyektlərin təsvirini və bu obyektlər arasında əlaqələri təhlil etmək üçün mövzu modelləşdirməyə əsaslanan metod təklif olunmuşdur. [113]-də fərdlər arasında qrupları və onlara məxsus uyğun sənədlər (mətnlər, məqalələr) arasından mövzuları aşkar etmək üçün model təklif olunmuşdur. [176]-da mövzu modelləşdirmə alqoritmlərindən ən çox istifadə olunan Gizli Dirixle Paylanmasına (GDP) əsaslanan model təklif olunmuşdur. [133]-də müəllif-mövzu modeli işlənmişdir. Bu tədqiqatda müəlliflər və onların uyğun mövzu paylanmalarını müəyyən edən model təklif olunmuşdur. Bu işdə aparılan eksperiment nəticəsində test sənədlərində sadəcə az sayda sözdən istifadə olunduqda GDP-nin yaxşı nəticə verdiyi qeyd olunmuşdur.

[130]-da aparılan tədqiqatlarda GDP modeli genişləndirilmiş və onun mikro-bloq mühitində tətbiqi öyrənilmişdir. [126]-da GDP vasitəsilə qısa mətnlərin modelləşdirilməsi problemi tədqiq olunmuşdur. [175]-də sənədlərin klasterləşməsi məqsədilə Ehtimallı Gizli Semantik Analiz və GDP modelləri araşdırılmışdır. Bu tədqiqatda mövzu modelləşmədən sənədlərin xüsusiyyətlərini qiymətləndirmək məqsədilə mövzuların sayının müəyyən olunmasında istifadə olunur. Oxşar şəkildə, [174]-də Korrelyasiyalı Mövzu Modelləri (Correlated Topic Models), İyerarxik GDP (Hierarchical Latent Dirichlet Allocation) və İyerarxik Dirixle Prosesi (Hierarchical Dirichlet Process) kimi modellərdən sənədlərin klasterləşməsində istifadə olunmuşdur. Burada bir neçə fərqli sahədən elmi məqalələr toplanmış və eyni sahədən olan məqalələrin birlikdə klasterləşib - klasterləşməməsi araşdırılmışdır. [128]-də mövzu və birləşdirmə (ing. fusion) modelindən istifadə edərək klasterləşmənin keyfiyyətini artırmaq məqsədilə metod təklif olunmuşdur. Təklif olunan metodda mövzu modelləşdirmə alqoritmi əvvəlcə sənədlər kolleksiyasına bir neçə dəfə tətbiq olunur və hər iterasiyada hər bir sənəd üçün xüsusi mövzular müəyyən olunur. Daha sonra hər bir iterasiyada əldə olunmuş xüsusi mövzular kombinasiya olunur, sənəd vahid vektor şəklində təsvir olunur və sonda sənədlər klasterləşdirilir. [120]-də təklif olunan yanaşmada isə əvvəlcə klasterləşmə metodu sənədlər çoxluğuna tətbiq olunur, daha sonra hər bir klasterdən mövzular çıxarılır. Metod dörd mərhələdən ibarətdir. Birinci mərhələdə klasterləşmə metodu vasitəsilə oxşar sənədlər qruplaşdırılır. İkinci mərhələdə klaster mövzularını müəyyən etmək üçün GDP alqoritmi hər bir klasterə tətbiq olunur. Üçüncü mərhələdə çıxarılmış mövzu terminləri içərisindən qlobal tezlikli terminlər çıxarılır və onlara semantik yaxın olan sözlər WordNet-dən istifadə olunaraq müəyyən olunur. Sonuncu mərhələdə bu terminlərin və onlara yaxın olan sözlərin daxil olduğu cümlələr seçilir və ümumiləşdirmə aparılır.

Mövzu modelləşdirmə metodları və tətbiq sahələri: Son zamanlarda e-sənədlərin sayının sürətli artımı onların idarə olunması, axtarışı və s. üçün yeni üsul və vasitələrin istifadəsini tələb edir. Bu məqsədlə yaradılmış modellərdən biri də mövzu modelləşdirmə alqoritmləridir. Mövzu modelləşdirmə alqoritmləri sənədlər

üçün generativ bir modeldir [103]. Belə ki, burada hər bir sənədə mövzular qarışığı kimi baxılır və hər bir mövzu sözlərin onlar üzərində ehtimal paylanması vasitəsilə təyin olunur. Bir sıra modellər, məsələn, Gizli Semantik Analiz, Ehtimallı Gizli Semantik Analiz, Gizli Dirixle Paylanması və s. kimi mövzu modelləri sənədlərdən mövzuların aşkar olunması sahəsində (təsnifat dəqiqliyinin yaxşılaşdırılmasında) uğurla tətbiq olunmuşdur [35, 121]. Bu modellər haqqında aşağıda məlumat verilmişdir.

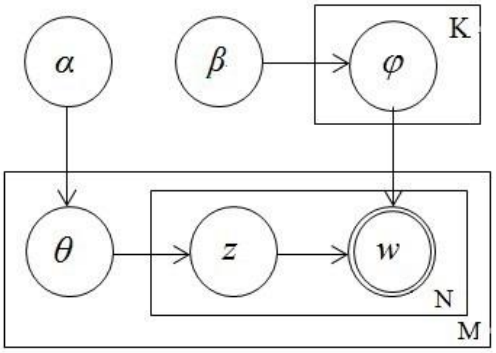
Gizli Semantik Analiz: Gizli Semantik Analiz (GSA) Təbii Dil Emalında istifadə olunan üsullardan biridir. Keçmişdə GSA Gizli Semantik İndeksləmə adlandırılmışdır, lakin sonralar informasiya axtarışı məqsədi üçün təkmilləşdirilmişdir. Bu metod vasitəsilə sorğuya uyğun sənədlər bir neçə sənəd çoxluğundan seçilir. Bu alqoritmdə mənaca yaxın sözlərin oxşar mətn hissələrində işlənməsi ehtimalı fərz olunur. Burada böyük mətn hissələrindən sözlərin sayına əsaslanan matris qurulur, matrisin sətirləri sözləri, sütunları abzasları təmsil edir. Sinqulyar ayrılış (Singular value decomposition) adlı riyazi metoddan istifadə edərək sütunlar arasında yaxınlığı saxlayaraq sətirlərin sayının azaldılması həyata keçirilir. Daha sonra abzaslar hər iki sütunun yaratdığı vektorlar arasında bucağın kosinusunun hesablanması ilə müqayisə olunur. Hesablama nəticəsində alınan qiymətlər 1-ə yaxındırsa, abzaslar oxşar, əks halda fərqli hesab olunur. GSA-nın əsas tətbiq sahələri bunlardır: terminlər arasında əlaqələrin aşkarlanması, mətnlər çoxluğunda söz birləşmələrinin təhlili, sorğuya uyğun sənədlərin tapılması və s.

Ehtimallı Gizli Semantik Analiz: Ehtimallı Gizli Semantik Analiz (EGSA), GSA metodunun bəzi çatışmayan cəhətlərini aradan qaldırmaq üçün yaradılmış yanaşmadır. Jan Puzicha və Thomas Hofmann onu 1999-cu ildə təqdim etmişdir [107]. EGSA gizli sinif modelinə əsaslanaraq sənədlərin indeksləşdirilməsini avtomatlaşdırmaq üçün yaradılmış metoddur. EGSA-ın əsas məqsədi lüğətdən və ya tezaurusdan istifadə etmədən sözün müxtəlif kontekstləri arasında fərq qoymaq və onları fərqləndirməkdir. İki mühüm təsirə malikdir. Birincisi, bu metod çoxmənalı sözləri fərqləndirməyə imkan verir. İkincisi, ümumi məzmunlu sözləri qruplaşdırmaqla tipik oxşarlıqları müəyyən edir. EGSA təsvirlərin axtarışı (Image

retrieval), sorğuların avtomatik tövsiyə olunması (Automatic question recommendation) və s. kimi sahələrdə uğurla tətbiq olunur [70].

Gizli Dirixle Paylanması: Gizli Dirixle Paylanmasının (GDP) yaradılmasında əsas məqsəd EGSA və GSA metodlarını yaxşılaşdırmaqdan ibarətdir. GDP statistik (Bayes) mövzu modelinə əsaslanan və çox geniş istifadə edilən text mining alqoritmidir. GDP mövzu modelləşdirmədə standart alətlərdən biri hesab olunur və mövzu modelləşdirmənin ən ümumi alqoritmlərindən biridir [13, 35]. Qeyd edək ki, burada əvvəlcədən yalnız mövzuların sayı bəlli olur və iki əsas prinsip gözlənilir: 1) Hər bir sənəd bir neçə gizli mövzunun qarışığıdır; 2) Hər bir mövzu bir neçə sözün qarışığıdır.

Sənəd və mövzu arasında əlaqə Dirixle paylanmasına, mövzu və söz arasında əlaqə polinomial paylanmaya uyğun olaraq müəyyən olunur. GDP-nin generativ prosesi şəkil 4.2.1-də təsvir olunmuşdur [137, s.200]:



Şəkil 4.2.1. GDP-nin generativ prosesi

burada M – sənədlərin sayı, K – gizli mövzuların sayı, N – sənəddəki sözlərin sayıdır. z – i -ci sənəddə j-ci sözün məxsus olduğu mövzu, w – sənəddə müşahidə olunan söz və α, β – parametrlərdir, α – sənədlər çoxluğunda gizli mövzuların nisbi gücünü, β – bütün gizli mövzuların ehtimal paylanmasını təyin edir. θ – cari sənəd üçün mövzular üzrə ehtimalların paylanması (topic probability distribution), φ – cari gizli mövzu üzrə sözlərin paylanmasıdır (word distribution). Düzbucaqlı təkrarlanan seçmə prosesi, dairə gizli dəyişənləri, ikili dairə müşahidə olunan dəyişənləri göstərir. Bu modelin hesablama düsturu aşağıda göstərilmişdir:

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha)\prod_{n=1}^{N} p(z_n|\theta)\, p(w_n|z_n, \beta) \qquad (4.2.1)$$

GDP-nin əsas addımları aşağıdakı mərhələlərlə həyata keçirilir:

1) i-ci gizli mövzu üçün Dirixle paylanmasına uyğun olaraq φ polinomial paylanma hesablanır;

2) Puasson paylanmasına uyğun olaraq sənəddəki sözlərin sayı (N) əldə olunur;

3) Hər bir mətn üçün θ mövzu ehtimal paylanması hesablanır;

4) Sənədlər çoxluğunda hər bir sənədə daxil olan söz üçün aşağıdakılar nəzərə alınır:

a) θ mövzular üzrə ehtimal paylanmasından təsadüfi olaraq z gizli mövzu seçilir;

b) z mövzusunun polinomial paylanmasından təsadüfi olaraq söz seçilir.

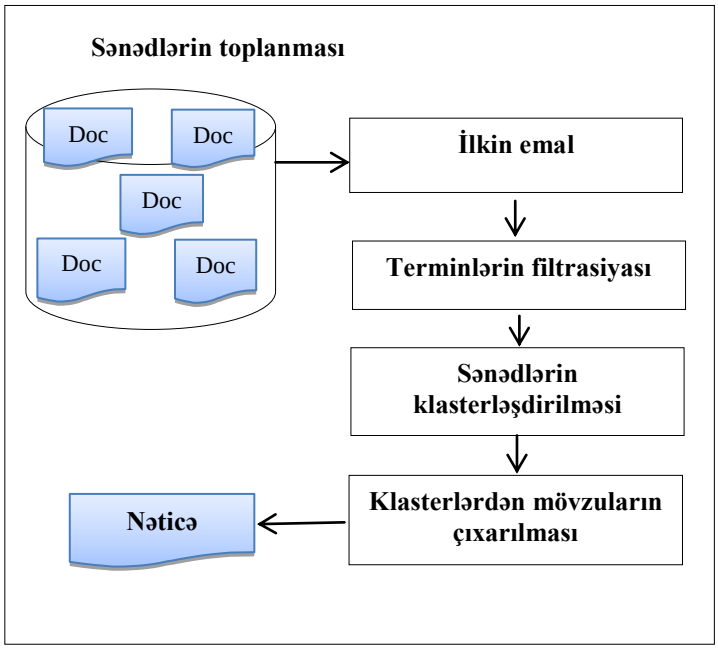
GDP-nin tətbiq sahələri bunlardır: emosiyanın mövzusunun təyini və ya mövzuların sentimental analizi (Emotion topic), inşaların avtomatik qiymətləndirməsi (Automatic essay grading), fişinqlə mübarizə (Anti-Phishing) və s.

E-dövlətdə mövzu modelləşdirmənin tətbiqi: E-dövlət xidmətlərinin əlçatanlığını və effektivliyini artırmaq üçün mütəmadi olaraq istifadəçi yönümlü qiymətləndirmə aparmaq lazımdır. Bu cür qiymətləndirmə nəticəsində e-dövlət saytlarının resurslarının və xidmətlərinin keyfiyyətini artırmaq olar. Burada bir neçə məsələyə toxunmaq olar [79]:

- Vətəndaşların daha çox hansı xidmətlərə tələbatı var?
- Vətəndaşların tələblərini əks etdirən xidmətlərin prioritetlərini necə tapmalıyıq?
- E-dövlət xidmətlərinin istifadəsi barədə insanlar nə düşünürlər? və s.

Məlum olduğu kimi, e-dövlət mühitində vətəndaşlar hər hansı xidmətə şərhlər yazmaqla münasibət bildirə bilərlər. Bu şərhləri analiz etməklə, onların narahat olduğu əsas məsələləri müəyyən etmək olar [75, 77]. Məlumdur ki, şərhlərin sayı artdıqca onları analiz etmək də çətinləşir. Vətəndaşların narahat olduğu əsas məqamları tez bir şəkildə müəyyən etmək üçün text mining metodlarının tətbiqi

vacibdir. Bunları nəzərə alaraq, e-dövlət mühitində vətəndaş şərhlərinin analizi vasitəsilə onların hansı mövzulardan narahat olduğunu və ya maraqlandığını müəyyən etmək üçün yanaşma təklif etmişik. Təklif etdiyimiz yanaşmada GDP və k-Means alqoritmlərindən istifadə olunur. Təklif olunan yanaşmanın əsas addımları şəkil 4.2.2-də təsvir olunmuşdur [78]:



Şəkil 4.2.2. Təklif olunan yanaşmanın sxemi

Bu addımların hər biri aşağıda ətraflı şəkildə şərh olunmuşdur:

Addım 1. Əvvəlcə e-dövlət mühitində istifadəçi şərhləri toplanılır. Sadəlik üçün bu şərhlərə sənəd kimi baxılır və aşağıdakı kimi işarə olunur:

$$D = \{d_1, d_2, ..., d_n\} \tag{4.2.2}$$

burada n - sənədlərin (şərhlərin) sayıdır.

Addım 2. Toplanan şərhlər ilkin emal olunur. İlkin emal zamanı sənədlərdən ümumişlək sözlər, rəqəmlər və durğu işarələri təmizlənir. Hər bir söz müxtəlif formalarda şəkilçilər qəbul etdiyi üçün onlar ilkin variantına (kökünə) qaytarılır.

Addım 3. Şərhlərdən terminlər çıxarılır. Daha sonra sənədlər çoxluğu "Term Frequency - Inverse Document Frequency (TF-IDF)" sxeminin köməyilə vektor kimi təsvir olunur [93].

$$d_i = \{w_{i1}, w_{i2}, ..., w_{im}\}, \quad i = 1,2,...,n \tag{4.2.3}$$

burada w_{ij} – i -ci sənəddə j -ci terminin TF-IDF çəkisidir. TF-IDF sxeminə görə w_{ij} çəkisi belə hesablanır:

$$w_{ij} = tf_{ij} \times \log(\frac{n}{n_j}) \qquad (4.2.4)$$

burada tf_{ij} – i -ci sənəddə j -ci terminin işlənmə tezliyi, n_j – j -ci terminin rast gəlindiyi sənədlərin sayıdır.

Sənədlər arasındakı məsafəni hesablamaq üçün Evklid məsafəsindən istifadə olunur. Onda $d_i = \{w_{i1}, w_{i2}, ..., w_{im}\}$ və $d_l = \{w_{l1}, w_{l2}, ..., w_{lm}\}$ vektorları arasındakı yaxınlıq belə hesablanır:

$$dist(d_i, d_l) = \sqrt{\sum_{j=1}^{m} (w_{ij} - w_{lj})^2} \quad i, l = 1, ..., n \qquad (4.2.5)$$

Məlumdur ki, sənədlər çoxluğunda rast gəlinən terminlərin sayı həddən artıq çox olur və bu say bir sənəddə rast gəlinən terminlərin sayından çox-çox böyük olur. Onda sənədlərin TF-IDF sxemi ilə təsvir olunan vektorlarının elementlərinin çox hissəsi "0"-lar olacaq. Başqa sözlə vektorlar seyrək olacaq. Bu isə sənədlərin klasterləşməsində iki mühüm problem yaradır:

- "Lənətə gəlmiş" ölçü problemi;
- Klasterləşmənin keyfiyyəti.

Bu problemləri aradan qaldırmaq üçün vektordan seyrək terminlər əvvəlcədən təmizlənir. Seyrək terminlər təmizləndikdən sonra yuxarıda göstərilən problemlərə təsir edən digər bir amil ortaya çıxır. Bu da sənədlər çoxluğunda sinonim sözlərin olmasıdır. Belə ki, sənədlər çoxluğunda sinonim sözlər olduqda klasterləşmə zamanı oxşar məzmunlu sənədlər müxtəlif klasterlərə düşə bilər. Bu isə klasterləşmənin keyfiyyətinin aşağı düşməsinə gətirib çıxarır. Bu kimi halları aradan qaldırmaq üçün sənədlər çoxluğunda semantik yaxın sözlərin tapılması, onlardan birinin saxlanılması və digərlərinin kənarlaşdırılması təklif olunur. Sözlərin bir-birinə semantik yaxınlığını tapmaq üçün hər bir terminin genişlənmiş sinonimləri çoxluğundan istifadə olunması təklif olunur. Bunun üçün WordNet şəbəkəsindən istifadə etməklə hər bir terminin sinonimlər çoxluğu tapılır və $t_i \rightarrow synset(t_i)$ ilə işarə olunur. Qeyd

edək ki, WordNet sözlər arasındakı semantik münasibətləri müəyyən etməyə imkan verən şəbəkədir. Məsələn, bu şəbəkənin köməyilə sinonimləri, hipernimləri, hiponimləri və s. asanlıqla tapmaq mümkündür [17, 23].

Hər bir terminin genişlənmiş sinonimləri çoxluğu tapıldıqdan sonra sözlər arasında semantik yaxınlıq aşağıdakı metrikadan istifadə etməklə hesablanır [151]:

$$sim(t_g, t_s) = \frac{2 * \left| synset\ (t_g) \cap synset\ (t_s) \right|}{\left| synset\ (t_g) \cup synset\ (t_s) \right|} \geq \alpha, \quad g, s = 1,2,..., m \qquad (4.2.6)$$

burada $\left|synset\ (t)\right|$ – t sözünün sinonimlərinin sayı, $0 \leq \alpha \leq 1$ – idarəolunan parametrdir. Əgər sözlər arasında yaxınlıq α ədəddindən böyükdürsə, bu sözlər bir termin kimi qəbul olunur. Belə ki, bu sözlərdən yalnız biri saxlanılır, digərləri kənarlaşdırılır. Beləliklə, d vektorunun ölçüsünü azaltmış oluruq. Bu halda d_i vektoru aşağıdakı vektora transformasiya olunur:

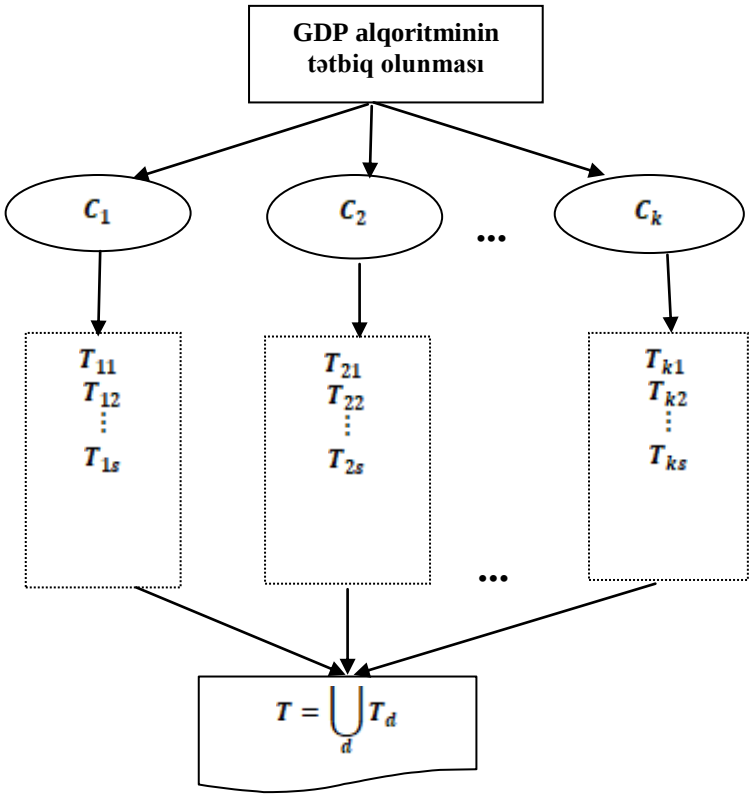
$$d_i \to d_i^* = \left\{ \overline{w}_{i1}, \overline{w}_{i2}, ..., \overline{w}_{im_0} \right\}, \quad m_0 \leq m \qquad (4.2.7)$$

burada $\overline{w}_{ij}$ – sinonim sözlər kənarlaşdırıldıqdan sonra j-ci sözün i-ci sənəddəki TF-IDF çəkisidir.

Addım 4. Sənədlər vektor şəklində təsvir olunduqdan sonra klasterləşdirilir. Sənədləri klasterlərə ayırmaq üçün bir sıra metodlar mövcuddur. Bu məqalədə sənədləri klasterləşdirmək üçün k-means metodundan istifadə olunması təklif olunur. k-means böyük verilənlərin analizində icra müddətinin az və tətbiqinin asanlığı səbəbindən populyar alqoritmlərdən biri hesab olunur. Bu alqoritmə görə klasterlər (4.2.8) düsturu vasitəsilə təyin olunur:

$$E = \sum_{i=1}^{k} \sum_{d \in C_i} \left| d - O_i \right|^2 \to \min \qquad (4.2.8)$$

burada k – klasterlərin sayı, C_i – i-ci klaster, O_i – i-ci klasterin mərkəzidir.



Şəkil 4.2.3. Klasterlərdən mövzuların çıxarılması

Addım 5. Sənədlər klasterlərə ayrıldıqdan sonra hər bir klaster üzrə mövzuları tapmaq mümkündür. Bunun üçün GDP-dən istifadə olunması təklif olunur. GDP haqqında əvvəlki bölmələrdə ətraflı məlumat verilmişdir. GDP vasitəsilə hər bir klaster üzrə sənədlərdən əsas mövzuların çıxarılması aşağıdakı şəkildə həyata keçirilir.

Tutaq ki, $\{C_1, C_2, ..., C_k\}$ klasterləri təyin olunmuşdur. Hər bir klasterə GDP alqoritmi tətbiq olunur və hər bir C_q klasteri üçün $T_q = \{T_{q1}, T_{q2}, ..., T_{qs}\}$ mövzuları təyin olunmuş olur. Burada s – mövzuların sayıdır. Proses sxematik olaraq şəkil 4.2.3-də təsvir olunmuşdur.

Beləliklə, vətəndaşların yazdığı şərhlərin əsas mövzusunu müəyyən etmiş oluruq.

Eksperiment

Eksperiment R proqramlaşdırma dilində aparılmışdır. Proqram kodu əlavədə (səh.131) verilmişdir. Eksperiment üçün BBC NEWS verilənlər çoxluğundan istifadə olunmuşdur. Bu verilənlər çoxluğu 2004-2005-ci illəri əhatə edən Biznes, Əyləncə,

Siyasət, İdman və Texnologiya adlanan beş aktual sahəyə uyğun BBC xəbər veb saytından toplanmış 2225 sənəddən ibarətdir [65]. Eksperimentdə Biznes, Əyləncə və İdman sahələrindən müxtəlif sayda sənədlər toplanmış və analiz olunmuşdur. Klasterləşmə nəticələrini qiymətləndirmək üçün "Təmizlik əmsalı" (Purity) meyarından istifadə olunmuşdur. Təmizlik əmsalı entropiya anlayışı ilə əlaqəli sadə və şəffaf qiymətləndirmə meyarıdır. Bu meyara görə C_p klasterinin təmizlik əmsalı aşağıdakı kimi təyin olunur [21]:

$$purity(C_p) = \frac{1}{|C_p|} \max_{p^+=1,...,k^+} \left| C_p \cap C_{p^+}^+ \right|, \quad p = 1,2,...,k, \quad p^+ = 1,2,..,k^+ \quad (4.2.9)$$

burada k^+ – siniflərin sayı, $C^+ = \left(C_1^+,...,C_{p+}^+\right)$ – siniflər çoxluğu, k – klasterlərin sayıdır. Qeyd edək ki, hər bir klasterdə müxtəlif siniflərdən sənədlər ola bilər. Təmizlik əmsalı klasterdə dominant sinfin ölçüsünün klasterin ölçüsünə nisbətini göstərir. Təmizlik əmsalı həmişə $\left[\frac{1}{k^+},1\right]$ intervalından qiymətlər alır. Bu əmsalın böyük qiyməti klasterin dominant sinfin "təmiz" alt hissəsi olduğunu göstərir. Bütün klasterlər çoxluğunun təmizlik əmsalı ayrı-ayrı klasterlərin təmizlik əmsallarının cəmi kimi qəbul olunur:

$$purity(C) = \sum_{p=1}^{k} \frac{|C_p|}{n} purity(C_p) = \frac{1}{n} \sum_{p=1}^{k} \max_{p^+=1,...,k^+} \left| C_p \cap C_{p^+}^+ \right| \quad (4.2.10)$$

burada n – bütün sənədlərin sayıdır. Təmizlik əmsalının böyük qiyməti klasterləşmənin keyfiyyətinin yüksək olduğunu göstərir.

İlkin emal

İlkin emal text mining-də əsas mərhələlərdən biri hesab olunur. Bunu nəzərə alaraq eksperiment zamanı toplanan sənədlər əvvəlcə ilkin emal olunmuşdur. İlkin emal zamanı sənədlər çoxluğundan durğu işarələri, rəqəmlər, simvollar, ümumişlək sözlər təmizlənmiş və sənədər çoxluğu TF-IDF sxeminin köməyilə vektor şəklində təsvir olunmuşdur. Daha sonra sənədlərdən seyrək terminlər təmizlənmişdir. İlkin emaldan əvvəl və sonra sənədlər çoxluğunda qalan sözlərin sayı cədvəl 4.2.1-də təsvir olunmuşdur.

Cədvəl 4.2.1
Sənədlər və sözlərin sayı

Sənədlərin sayı	Sözlərin sayı	
	İlkin emaldan əvvəl	İlkin emaldan sonra
100	8040	4421
300	18356	8851
500	25490	11766
800	33410	14750
1000	36346	15860

İlkin emaldan sonra sənədlər çoxluğuna təklif etdiyimiz metod tətbiq olunmuşdur. Sözlər arasında semantik yaxınlıq α –nın müxtəlif qiymətlərində $(0.1, 0.2, 0.3, 0.4, 0.5)$ hesablanmışdır. Metoddan sonra sənədlər çoxluğunda qalan terminlərin sayı cədvəl 4.2.2-də təsvir olunmuşdur. Cədvəldə sənədlərin sayı 100 olan varianta baxdıqda, buradan aydın görünür ki, $\alpha = 0.1$ qiymətində daha çox sayda semantik yaxın sözlər aşkarlanmış və vektorun ölçüsü seyrək terminlər təmizləndikdən sonra qalan sözlərə nisbətən xeyli dərəcədə (26.42%) azalmışdır. α –nın qiyməti artdıqca, daha az sayda sözlər atılmışdır. Belə ki, $\alpha = 0.5$ qiymətinə baxdıqda, vektorun ölçüsünün daha az (1.29%) azaldığını görərik. Sənədlərin sayı artdıqca, semantik yaxın olan sözlərin sayı da uyğun olaraq artmış və vektorun ölçüsü xeyli dərəcədə azalmışdır. Məsələn, sənədlərin sayı 800 olan varianta baxsaq, $\alpha = 0.1$ qiymətində vektorun ölçüsünün 31.67% azaldığını görərik.

Daha sonra sənədlər çoxluğuna k-means klasterləşmə metodu tətbiq olunmuş və klasterləşmənin dəqiqliyi cədvəl 4.2.3-də təsvir olunmuşdur. Qeyd edək ki, burada $\alpha = 0$ qiyməti seyrək terminlər təmizləndikdən sonra qalan terminlər çoxluğunu ifadə edir. Cədvəldən göründüyü kimi, terminlər çoxluğundan semantik yaxın sözlərin atılması klasterləşmənin keyfiyyətinə mənfi təsir göstərməmiş, əksinə təmizlik əmsalı kifayət qədər yüksək qiymət almışdır. Bu da klasterləşmənin keyfiyyətinin yüksək olduğunu göstərir. Cədvəldən də göründüyü kimi sənədlərin sayı artdıqca, təmizlik əmsalının qiyməti də kifayət qədər yüksək qiymət almışdır. Belə ki, $\alpha = 0.1$ vəziyyətinə baxsaq, sənədlərin sayı 100 olan halda təmizlik əmsalı 0.88 olduğu halda, sənədlərin sayı 1000 olduqda bu əmsal artaraq artaraq 0.97 olmuşdur.

Cədvəl 4.2.2
Emaldan sonra sənədlərdəki sözlərin sayı

$\alpha =$	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
Sənədlərin sayı	**Sözlərin sayı**					
	Seyrək terminlər təmizləndikdən sonra	**Semantik yaxın sözlər kənarlaşdırıldıqdan sonra**				
100	772	568 (26.4%)	691 (10.4%)	733 (5.05%)	756 (2.07%)	762 (1.29%)
300	878	633 (27.90%)	777 (11.50%)	838 (4.56%)	856 (2.51%)	865 (1.48%)
500	827	589 (28.77%)	725 (12.33%)	789 (4.59%)	809 (2.17%)	816 (1.33%)
800	821	561 (31.67%)	716 (12.79%)	778 (5.23%)	802 (2.31%)	810 (1.34%)
1000	801	561 (29.96%)	702 (12.36%)	765 (4.49%)	784 (2.12%)	791 (1.25%)

Cədvəl 4.2.3
α-nın müxtəlif qiymətlərində klasterləşmənin təmizlik (purity) əmsalı

$\alpha =$	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
Sənədlərin sayı	**Purity**					
100	0.95	0.88	0.81	0.82	0.82	0.88
300	0.996	0.98	0.98	0.98	0.98	0.98
500	0.996	0.93	0.94	0.99	0.99	0.99
800	0.98	0.81	0.89	0.99	0.99	0.99
1000	0.982	0.97	0.98	0.97	0.98	0.98

Sənədlər klasterləşdirildikdən sonra hər bir klasterdən mövzuları çıxarmaq üçün mövzu modelləşdirmə metodu tətbiq olunmuşdur. Bunun üçün hər bir klasterə Gizli Dirixle Paylanması tətbiq olunmuş və klaster mövzuları çıxarılmışdır. Cədvəl 4.2.4 və 4.2.5-də hər bir klaster üzrə çıxarılan top 10 söz təsvir olunmuşdur. Top 10 söz dedikdə, yəni hər bir mövzu üzrə sənədlər çoxluğunda ən çox işlənən sözlər nəzərdə tutulur.

Cədvəl 4.2.4
Hər bir klaster üzrə top 10 söz ($\alpha = 0$)

Cluster 1	Cluster 2	Cluster 3
music	ireland	cluster
award	england	growth
people	cluster	year
show	win	rate
year	wale	economi
won	side	bank
radio	game	econom
veto	tri	oil
years	nation	price
song	scotland	rise

Cədvəl 4.2.5
Hər bir klaster üzrə top 10 söz ($\alpha = 0.3$)

Cluster 1	Cluster 2	Cluster 3
film	england	cluster
star	play	team
year	cluster	year
award	year	rate
role	rugbi	economi
cluster	player	rise
includ	game	price
director	season	bank
play	cup	econom
bbc	week	month

Cədvəl 4.2.4-də $\alpha = 0$ halında, yəni semantik yaxın sözlər kənarlaşdırılmamışdan əvvəl sənədlər çoxluğundan çıxarılan top 10 söz təsvir olunmuşdur. Müqayisə üçün cədvəl 4.2.5-də $\alpha = 0.3$ halında, semantik yaxın sözlər kənarlaşdırıldıqdan sonra sənədlərdən çıxarılan top 10 söz təsvir olunmuşdur. Bu iki cədvələ diqqət yetirsək, semantik yaxın sözlərin sənədlər çoxluğundan kənarlaşdırılmasının nəticəyə təsir göstərmədiyini görərik. Yəni, GDP-in keyfiyyəti aşağı düşməmişdir. Belə ki, hər iki cədvəldə klaster 1-ə baxdıqda buradakı terminlərin əyləncə, klaster 2-də idman, klaster 3-də isə biznes ilə əlaqəli olduğunu görərik. Bu isə klaster 1-in Əyləncə, klaster 2-in İdman və klaster 3-ün Biznes sənədlərindən ibarət olduğunu göstərir. Göründüyü kimi, hər iki cədvəldə

klasterlərdən çıxarılan mövzular üst-üstə düşür. Bu isə təklif olunan metodun yaxşı işlədiyini göstərir. Qeyd edək ki, α-ın digər qiymətlərində (0.1, 0.2, 0.4, 0.5) də bu nəticə özünü doğruldur. Yəni, α-ın digər qiymətlərində də klasterlərdən çıxarılan mövzular cədvəl 4.2.5 ilə üst-üstə düşür, buna görə də müqayisə üçün yalnız $\alpha = 0.3$ qiyməti verilmişdir.

Göründüyü kimi mövzular dəqiqliklə çıxarılmış, vaxtda isə uduş əldə olunmuşdur. Belə ki, aşağıdakı cədvəldə klasterləşməyə və hər bir klasterdən mövzuların çıxarılmasına sərf olunan vaxt və onların müqayisəli analizi təsvir olunmuşdur (cədvəl 4.2.6).

Cədvəl 4.2.6

Sənədlərin klasterləşməsi və GDP-nin tətbiqinə sərf olunan vaxt

$\alpha =$	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
Sənədlərin sayı	**Sərf olunan zaman**					
100	11.6	9.3 (19.82%)	9.86 (15%)	10.32 (11.04%)	10.65 (8.19%)	11.06 (4.65%)
300	12.39	9.08 (26.71%)	9.74 (21.39%)	10.68 (13.80%)	11.26 (9.12%)	11.51 (7.10%)
500	18.7	11.35 (39.30%)	12.87 (31.18%)	13.28 (28.98%)	14.21 (24.01%)	14.98 (19.89%)
800	22.28	13.49 (39.45%)	14.59 (34.52%)	15.76 (29.26%)	16.57 (25.63%)	17.85 (20.33%)
1000	21.06	12.78 (39.31%)	14.09 (33.09%)	15.01 (28.72%)	16.01 (23.97%)	17.25 (18.09%)

Cədvəl 4.2.6-dan göründüyü kimi təklif olunmuş metod vasitəsilə zamanda xeyli uduş əldə olunmuşdur. Belə ki, sənədlərin sayı 100 olan variantda zamanda 19.82 % uduş əldə olunursa, sənədlərin sayı artdıqca (sənədlərin sayı 1000) $\alpha = 0.1$ halında bu faizin 39.31 %-ə qədər yüksəldiyini görə bilərik. Bu isə səmərəliliyin xeyli dərəcədə yaxşılaşması deməkdir.

Beləliklə, aparılan eksperiment və nəticələri onu göstərir ki, təklif olunmuş metod vasitəsilə böyük sənədlər çoxluğunun ölçüsünü xeyli dərəcədə azaldaraq, bu verilənlərin analizinə sərf olunan vaxta qənaət etmək, klasterləşmə və GDP alqoritminin keyfiyyətini yaxşılaşdırmaq olar.

DİSSERTASİYA İŞİNİN ƏSAS NƏTİCƏLƏRİ

Dissertasiya mövzusu üzrə aparılan tədqiqat prosesində qarşıya qoyulmuş məsələlər həll edilmiş və aşağıdakı nəticələr əldə edilmişdir:

1. E-dövlətin analizində sosial şəbəkə və mətnlərin intellektual analizi texnologiyalarının rolu araşdırılmış, o cümlədən e-dövlətə qoyulan tələblər və onun inkişafına yanaşmalar, e-dövlətin inkişafının və analizinin əsas problemləri müəyyən olunmuşdur [1, 4, 5, 25];
2. E-dövlətdə terrorizmlə əlaqəli məqalələrin aşkarlanması üçün hibrid təsnifatlandırma metodu işlənmişdir [3, 23];
3. E-dövlətdə terrorizmlə bağlı mətnlərin filtrasiyası üçün sentiment analiz texnologiyasına və Bayes klassifikatoruna əsaslanan metod işlənmişdir [8, 17];
4. E-dövlətdə fəaliyyət göstərən gizli sosial şəbəkələrin aşkarlanması və analizi üçün sentiment analiz texnologiyasına əsaslanan metod işlənmişdir [2, 18];
5. E-dövlət xidmətlərində vətəndaş məmnuniyyətinin avtomatik qiymətləndirilməsi üçün metod işlənmişdir [24, 77];
6. E-dövlətdə vətəndaşları (o cümlədən regionları) maraqlandıran aktual mövzuların müəyyən edilməsi üçün metod işlənmişdir [78, 79].

İSTİFADƏ EDİLMİŞ ƏDƏBİYYAT SİYAHISI

1. Alıquliyev, R.M. E-Dövlətin analizi texnologiyları: text mining və sosial şəbəkələr. Ekspress-informasiya. “İnformasiya Texnologiyaları seriyası” / R.M. Alıquliyev, G.Y. Niftəliyeva –Bakı: –2016. –78 s.
2. Alıquliyev, R.M., Niftəliyeva, G.Y. E-dövlət mühitində gizli sosial şəbəkələrin aşkarlanmasi üçün yanaşma // İnformasiya təhlükəsizliyinin multidissiplinar problemləri üzrə II respublika elmi-praktiki konfransı, – Bakı, –14 may 2015, – s.116-118.
3. Alıquliyev, R.M., Niftəliyeva, G.Y. E-dövlət mühitində terrorizmlə əlaqəli mətnlərin aşkarlanması metodu // İnformasiya təhlükəsizliyinin multidissiplinar problemləri üzrə II respublika elmi-praktiki konfransı, – Bakı, – 14 may 2015, – s.111-115.
4. Alıquliyev, R.M., Niftəliyeva, G.Y. E-dövlət sisteminin analizində Data mining texnologiyalarının tətbiq imkanları // “Big Data: imkanları, multidissiplinar problemləri və perspektivləri” I respublika elmi praktiki konfransı, –Bakı, –25 fevral 2016, – s. 81-84.
5. Alıquliyev, R.M., Niftəliyeva, G.Y. E-Dövlətin Big Data Mənbələri // “Big Data: imkanları, multidissiplinar problemləri və perspektivləri” I respublika elmi praktiki konfransı, – Bakı, 25 fevral 2016, – s. 78-80.
6. Əliquliyev, R.M. Sosial şəbəkələr / R.M. Əliquliyev, Y.N. İmamverdiyev, F.C. Abdullayeva – Bakı: “İnformasiya texnologiyaları”, – 2010. –278 s.
7. Əliquliyev, R.M., Yusifov, F.F. Elektron dövlətin formalaşmasının bəzi aktual elmi-nəzəri problemləri və həll perspektivləri // İnformasiya cəmiyyəti problemləri, –2014, № 2, – s. 3-13.
8. İskəndərli, G.Y. E-dövlətə kiber hücumlar və onlarla mübarizə üsulları haqqında // “İnformasiya təhlükəsizliyinin aktual multidissiplinar problemləri” üzrə IV respublika elmi-praktiki konfransı, – Bakı, – 14 dekabr 2018, – s.158-160.

9. Алыгулиев, Р.М. Роль технологии интеллектуального анализа текстов в обеспечении национальной безопасности // Проблемы Информационных Технологий, –2013. 1, –с.38-43.
10. Abdi, A., Idris, N., Alguliev, R.M., Aliguliyev, R.M. Automatic summarization assessment through a combination of semantic and syntactic information for intelligent educational systems // Information Processing & Management, – 2015. 51 (4), – p.340-358.
11. Abdi, A., Shamsuddin, S. M., Aliguliyev, R. M. QMOS: Query-based multi-documents opinion-oriented summarization // Information Processing and Management, –2018. 54 (2), – p. 318–338.
12. Akbaş, E. Aspect Based Opinion Mining on Turkish Tweets: / degree of master of science, –Turkey, 2012. –68 p.
13. Alghamdi R., Alfalqi, K. A Survey of Topic Modeling in Text Mining // International Journal of Advanced Computer Science and Applications, –2015. 6 (1), –p. 147-153.
14. Alguliev, R.M., Aliguliyev, R .M., Ganjaliyev, F.S. Building a social network of research institutes from information available on the web // International Journal of Networking and Virtual Organizations, –2012, 11 (1), –p. 62-76.
15. Alguliev, R.M., Aliguliyev, R.M., Ganjaliyev, F.S. Extracting a heterogeneous social network of academic researchers on the web based on information retrieved from multiple sources // American Journal of Operations Research, – 2011. 1 (2), – p. 33-38.
16. Alguliev, R.M., Aliguliyev, R.M., Mehdiyev, C.A. Sentence selection for generic document summarization using an adaptive differential evolution algorithm // Swarm and Evolutionary Computation, – 2011. 1 (4), – p.213-222.
17. Alguliyev, R. M., Aliguliyev, R. M. Niftaliyeva, G. Y. Filtration of Terrorism-Related Texts in the E-government Environment // International Journal of Cyber Warfare and Terrorism, –2018. 8 (4), p.35-48.

18. Alguliyev, R. M., Aliguliyev, R. M., Niftaliyeva, G. Y. A Method for Social Network Extraction From E-Government // International Journal of Information Systems in the Service Sector, –2019. 11 (3), p.37-55.
19. Alguliyev, R., Aliguliyev, R., Alakbarova, I. Extraction of hidden social networks from wiki-environment involved in information conflict // International Journal of Intelligent Systems and Applications, –2016. 8 (2), –p.20-27.
20. Alhomod, S. M. A. Best Practices in E government: A review of Some Innovative Models Proposed in Different Countries / S. M. A. Alhomod, M. M. Shafi, M. N. Kousarrizi [et al.] // International Journal of Electrical & Computer Sciences, –2012. 12 (1), –p. 1–6.
21. Aliguliyev, R. M. Performance evaluation of density-based clustering methods // Information Sciences, –2009. 179, –p.3583–3602.
22. Aliguliyev, R.M. A new sentence similarity measure and sentence based extractive technique for automatic text summarization // Expert Systems with Applications, –2009. 36 (4), –p.7764-7772.
23. Aliguliyev, R.M., Niftaliyeva, G.Y. Detecting terrorism-related articles on the e-government using text-mining techniques // Problems of Information Technology, –2015, 6 (2). –p. 36-46.
24. Aliguliyev, R.M., Niftaliyeva, G.Y. Hotspot Information of Public Opinion in E-Government // 10th IEEE International Conference on Application of Information and Communication Technologies (AICT2016), –Baku, –12-14 october, 2016, –p.645-646.
25. Aliguliyev, R.M., Niftaliyeva, G.Y. The current state, problems and perspectives of e-government analysis technologies // Problems of Information Technology, – 2017. 8 (1), –p. 53-63.
26. Almazan, R. S., Gil-Garcia, J. R. E-Government Portals in Mexico // Electronic Government: Concepts, Methodologies, Tools and Applications, –2008. –p. 367-376.

27. Alruily, M., Ayesh, A., Al-Marghilani, A. Using self organizing map to cluster arabic crime documents / International Multiconference on Computer Science and Information Technology, –Wisla, Poland, – 18-20 October, – 2010, – p.357–363.
28. Al-Shboul, M. Challenges and Factors Affecting the Implementation of E-Government in Jordan / M. Al-Shboul, O. Rababah, R. Ghnemat [et al.] // Journal of Software Engineering and Applications, –2014. 7, –p. 1111-1127.
29. Alshehri, A. Measuring User Satisfaction with e-Government Systems: An Empirical Study to Evaluate IS Effectiveness / A Doctoral Thesis, –2016. –290 p.
30. Alstyne, M. V., Zhang, J. EmailNet:A system for automatically mining social networks from organizational email communication // 2003 North American Association for Computational Social and Organizational Science, –2003,–p.1-4
31. Archmann, S., Iglesias, J. Perspectives on e-government in Europe // Information communication technologies and the virtual public sphere: impacts of network structures on civil society, –2011.– p. 195-206.
32. As-Saber, Sh.N., Srivastava, A., Hossain, Kh. Information Technology Law and E-government: A Developing Country Perspective // JOAAG, –2006. 1 (1), – p.84-101.
33. Attard, J. A systematic review of open government data initiatives / J. Attard, F. Orlandi, S. Scerri [et al.] // Government Information Quarterly, –2015. 32 (4), – p. 399-418.
34. Bernhard, I. Degree of Digitalization and Citizen Satisfaction: A Study of the Role of Local e-Government in Sweden / I. Bernhard, L. Norström, U. L. Snis [et al.] // The Electronic Journal of e-Government, –2018. 16 (1), –p.59-71.
35. Blei, D. M. Introduction to probabilistic topic models // Communications of the ACM, –2011. –p.1-16.
36. Borgatti, P. S., Halgin, S. D. On Network Theory // Organization Science, – 2011. 22 (5), –p. 1168-1181.

37. Bsoul, Q., Salim, J., Zakaria, L.Q. An intelligent document clustering approach to detect crime patterns // Procedia Technology, – 2013. 11, – p.1181-1187.
38. Cardeñosa, J., Gallardo, C., Moreno, J.M. Text Mining Techniques to Support e-Democracy Systems // In: CSREA EE 2009, –2009. –p. 401–405.
39. Carrara, S. Towards e-ECI's European participation by online Pan-European mobilization // Perspectives on European Politics and Society, –2012. 13 (3), – p. 252-369.
40. Carter, L., Bélanger, F. The Utilization of E-Government Services: Citizen Trust, Innovation and Acceptance Factors // Information Systems Journal, – 2005. 15 (1), –p. 5-25.
41. Catanese, S. S., Meo, P. P. D., Ferrara, E., Fiumara, G. Analyzing the Facebook Friendship Graph // 1st International Workshop on Mining the Future Internet-MIFI'10, – Berlin, Germany, –2010, –p 14-19.
42. Chan, C.M. From open data to open innovation strategies: Creating e-services using open government data // 46th Hawaii International Conference on System Sciences, –2013, –p. 1890–1899.
43. Chang, J., Boyd-Graber, J., Blei, D. M. Connections between the lines: augmenting social networks with text // 15th ACM SIGKDD International Conference on Knowledge Discovery and Data mining,–2009, – p. 169-178.
44. Chaovalit, P., Zhou, L. Movie review mining: a comparison between supervised and unsupervised classification approaches // 38th Hawaii International Conference on System Sciences, –2005, – p.1-9.
45. Chatfield, A., AlAnazi, J. Service quality, citizen satisfaction, and loyalty with self-service delivery options to transforming e-government services // 24th Australasian Conference on Information Systems, –2013, –p. 1-11.
46. Chen, Y.-C. eGovernment Network: The Role of Information Technology in Managing Networks // National Public Management Research Conference, – 2003, –p. 1-36.

47. Chen, Y.-Ch., Hsieh, T.-Ch. Big Data for Digital Government: Opportunities, Challenges, and Strategies // International Journal of Public Administration in the Digital Age, –2014. 1 (1), –p. 1-14.
48. Choi, D., Ko, B. Kim, H. Text analysis for detecting terrorism-related articles on the web // Journal of Network and Computer Applications, – 2014. 38, – p.16-21.
49. Chutimaskul, W. e-Government Analysis and Modeling // 3rd International Workshop on "Knowledge Management in e - Government", –2002,7, –p. 112-123.
50. Contractor, S. N., Wasserman, S., Faust, K. Testing Multitheoretical, Multilevel Networks: An Analytic Framework and Empirical Example // Academy of Management Review, –2006. 31 (3), –p. 681-703.
51. Conway, M. Terrorist 'use' of the internet and fighting back // Conference Cybersafety: Safety and Security in a Networked World: Balancing Cyber-Rights and Responsibilities, – 2005, – p.1-34.
52. Culotta, A., Bekkerman, R., McCallum, A. Extracting social networks and contact information from e-mail and the web // Conference on Email and Anti-Spam, –2004, – p.1-9.
53. Davies, R. eGovernment: Using technology to improve public services and democratic participation: [Electronic resource] / European Parliamentary Research Service, – 2015. URL:https://www.europarl.europa.eu/RegData/etudes/IDAN/2015/565890/EPRS_IDA(2015)565890_EN.pdf
54. Deloitte Consulting, Deloitte & Touche. At the dawn of egovernment: The citizen as customer / New York: Deloitte Research, –2000. –66 p.
55. DeRosa, M. Data Mining and Data Analysis for Counterterrorism: [Electronic resource] / Center for Strategic and International Studies, –Washington, –2004. URL: https://csis-website-prod.s3.amazonaws.com/s3fs-public/legacy_files/files/media/csis/pubs/040301_data_mining_report.pdf

56. Devroye, L. A probabilistic theory of pattern recognition / L. A Devroye, L. Gyorfi, G. Lugosi, Springer, –1996. –654 p.
57. Egovernment / H. El-Sersik, A. Abu Hamad, A.E. Alfalet, A. [et al.], –2010. – 55 p.
58. Elovici, Y., Shapira, B., Last, M., Zaafrany, O., Friedman, M., Schneider, M., Kandel, A. Detection of Access to Terror-Related Web Sites Using an Advanced Terror Detection System (ATDS) // Journal of the American Society for Information Science and Technology, –2010. 61 (2), – p. 405-418.
59. Escher, T. Governing from the Centre? Comparing the Nodality of Digital Governments / Escher, T., Margetts, H., Petricek [et al.] // Prepared for delivery at the 2006 Annual Meeting of the American Political Science, –2006. –32 p.
60. Fatudimu, I.T., Musa, A.G. Knowledge Discovery in Online Repositories: A Text Mining Approach // European Journal of Scientific Research, –2008, 22 (2), –p. 241–250.
61. Fred, C., Terrorism and Cyberspace // Network Security, –2002. 5, –pp. 17-19.
62. Furnell, S., Warren, M. Computer Hacking and Cyber Terrorism: The Real Threats in the New Millennium // Computers and Security, –1999. 18 (1), – p.28-34.
63. Gandomi, A., Haider, M. Beyond the hype: Big data concepts, methods, and analytics // International Journal of Information Management, –2015. 35 (2), – p.137–144.
64. Ghahramani, M., Hon, Ch. T. Mobile Phone Data Analysis: A Spatial Exploration Toward Hotspot Detection // IEEE Transactions on Automation Science and Engineering, –2018, 99, –p.1-12.
65. Greene, D., Cunningham, P. Practical solutions to the problem of diagonal dominance in kernel document clustering // International Conference on Machine Learning, –2006, –p. 377- 384.
66. Halchin, L. E. Electronic government: Government capability and terrorist resource // Government Information Quarterly, –2004. 21 (4), –p. 406–419.

67. Halpern, D., Katz, J. E. From e-government to Social Network Government: Towards a Transition Model // 4th Annual ACM Web Science Conference, –2012, –p. 119-127.
68. Haugen, S. E-government, cyber-crime and cyber-terrorism: a population at risk // Electronic Government, –2005. 2 (4), –p. 403-412.
69. Hiller, J. S., Belanger, F. Privacy Strategies for Electronic Government: [Electronic resource] / Rowman & Littlefiels publisher, –2001. URL: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=46B8532F7F08A688B8467432EC94FE52?doi=10.1.1.470.2867&rep=rep1&type=pdf
70. Hofmann, T. Unsupervised learning by probabilistic latent semantic analysis // Machine Learning, –2001. 42 (1), –p. 177-196.
71. Hong, L., Davison, B. D. Empirical Study of Topic Modeling in Twitter // First Workshop on Social Media Analytics, –2010, –p.80-88.
72. Hood, C. The Tools of Government in the Digital Age / C.Hood, H. Margetts, – London : Palgrave, –2006. –232 p.
73. Hu, Y-H., Chen, Y.-L., Chou, H.-L. Opinion mining from online hotel reviews – A text summarization approach // Information Processing and Management, –2017. 53 (2), –p. 436-449.
74. Huang, Ch.-Ch. User's Segmentation on Continued Knowledge Management System Use in the Public Sector // Journal of Organizational and End User Computing, –2020. 32 (1), – p. 19-40.
75. Iftikhar, R., Khan, M. S. Social Media Big Data Analytics for Demand Forecasting: Development and Case Implementation of an Innovative Framework // Global Information Management, –2020. 28 (1), –p.103-120.
76. Ingram, M. Internet privacy threatened following terrorist attacks on US: [Electronic resource] / – 24 september, 2001. URL: http://www.wsws.org/articles/2001/sep2001/isp-s24.shtml
77. Iskandarli, G.Y. Using Hotspot Information to Evaluate Citizen Satisfaction in E-Government: Hotspot Information // International Journal of Public Administration in the Digital Age, –2020. 7 (1), –p. 47-62.

78. Iskandarli, G.Y. Applying Clustering and Topic Modeling to Automatic Analysis of Citizens' Comments in E-Government // International Journal of Information Technology and Computer Science, – 2020. 12 (6), –p.1-10.
79. Iskandarli, G.Y. Detecting the Main Topics of Citizens' Comments in e-Government // 2nd International Symposium on Applied Sciences and Engineering, Atatürk University, – Erzurum, Turkey, –7-9 april, –2021, –p.539-543.
80. Jaeger, P. T., Thompson, K. M. E-government around the world: Lessons, challenges, and future directions // Government Information Quarterly, –2003. 20 (4), –p. 389–394.
81. Janssen, M., Charalabidis, Y., Zuiderwijk, A. Benefits, adoption barriers and myths of open data and open government // Information Systems Management, –2012. 29 (4), –p. 258-268.
82. Javanmardi, S., McDonald, D. W., Lopes, C.V. Vandalism Detection in Wikipedia: A High-Performing, Feature-Rich Model and Its Reduction through Lasso // 7th International Symposium on Wikis and Open Collaboration, –New York, ACM, –2011, –p. 82-90.
83. Jovanovska, M. B., Erman, N., Todorovski, L. Evaluating the Maturity Level of e-Government Back-Office with Social Network Analysis // 20th NISPAcee Annual Conference. Public Administration East and West: Twenty Years of Development, –Ohrid, Macedonia, –2012, –p.1-6.
84. Kadriu, A., Saveska, M., Abazi-Bexheti, L. E-Government Exploration using Social Network Analysis Methods // International Journal of Humanities and Management Sciences, –2013. 1 (2), –p. 151-154.
85. Kaushik, A., Kaushik, A., Naithani, S. A Study on Sentiment Analysis: Methods and Tools // International Journal of Science and Research, –2013. 4 (12), –p. 287-292.
86. Kautz, H., Selman, B., Shah, M. The Hidden Web // American Association for Artificial Intelligence Magazine, –1997. 18 (2), –p 27–35.

87. Kazienko, P. Multidimensional Social Network: Model and Analysis / Kazienko, P. Musial, K., Kukla, E. [et al.] // International Conference on Computer and Computational Intelligence, –Bangkok, Thailand, –2011, –p 378-387.
88. Keselj, V. N-gram based author profiles for authorship attribution / Keselj, V. Peng, F., Cercone N. [et al.] // Conference of the Pacific Association for Computational Linguistics, –Nova Scotia, Canada, –August 22-25, –2003, –p.255-264.
89. Khan, N. G., Bhagat V.B. Effective Data Mining Approach For Crime-Terrorpattern Detection Using Clustering Algorithm Technique // International Journal of Engineering Research & Technology, –2013. 2 (4), –p. 2043-2048.
90. Khan, G. F., Park, H. W. The e-government research domain: A triple helix network analysis of collaboration at the regional, country, and institutional levels // Government Information Quarterly, –2013. 30 (2), –p. 182-193.
91. Kim, D.-Y., Grant, G. E-government maturity model using the capability maturity model integration // Journal of Systems and Information Technology, –2010. 12 (3), –p. 230–244.
92. Kim, G.-H., Trimi, S., Chung, J.-H. Big-Data Applications in the Government Sector // Communications of the ACM, –2014. 57 (3), –p. 78-85.
93. Kim, S.W., Gil, J.M. Research paper classifcation systems based on TF-IDF and LDA schemes // Human-centric Computing and Information Sciences, –2019. 9 (30), –p. 1-21.
94. Koc-Menard, S. Trends in terrorist detection systems // Journal of Homeland Security & Emergency Management,–2009. 6 (1), –p. 16-21.
95. Kosala, R., Blockeel, H. Web mining research: A survey // SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery and Data Mining, ACM, –2000. 2 (1), – p 1-15.
96. Krebs, V. Visualizing Human Networks, Release, 1.0, Esther Dyson’s monthly report // Daphne Kis publisher, –February, –1996. –25 p.

97. Ku, C.-H., Leroy, G. A decision support system: automated crime report analysis and classification for e-government // Government Information Quarterly, – 2014. 31 (4), – p.534-544.
98. Qin, K. Hotspots Detection from Trajectory Data Based on Spatiotemporal Data Field Clustering / K.Qin, Q. Zhou, T. Wu [et al.] // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, – 2017. 42 (2), –p.1319-1325.
99. Qureshi, P. A. R., Memon, N ., Wiil, U . K. Detecting Terrorism Evidence in Text Documents // IEEE International Conference on Social Computing,–2010, – p.521-527.
100. Last, M., Markov, A., Kandel, A., Multi-lingual detection of terrorist content on the web // Lecture Notes in Computer Science, – 2006. 3917, – p.16-30.
101. Layne, K., Lee, J. Developing fully functional e-government: A four stage model // Government Information Quarterly, –2001. 18 (2), –p. 122−136.
102. Lee, G., Kwak, Y. H. An Open Government Maturity Model for social media-based public engagement // Government Information Quarterly, –2012. 29 (4), – p. 492-503.
103. Levy, K. E. C., Franklin, M. Driving Regulation: Using Topic Models to Examine Political Contention in the U.S. Trucking Industry // Social Science Computer Review, –2013. 32 (2), –p. 182-194.
104. Li, N., Wu, D. D. Using text mining and sentiment analysis for online forums hotspot detection and forecast // Decision Support Systems, –2010. 48 (2), –p. 354–368.
105. Li, Y. Sentence similarity based on semantic nets and corpus statistics / Li, Y McLean, D., Bandar, Z.A. [et al.] // IEEE Transactions on Knowledge and Data Engineering, – 2006, 18 (8), – p.1138-1150.
106. Linders, D. From e-government to we-government: Defining a typology for citizen coproduction in the age of social media // Government Information Quarterly, –2012. 29 (4), –p. 446–454.

107. Liu, S., Xia, C., Jiang, X., Efficient Probabilistic Latent Semantic Analysis with Sparsity Control // IEEE International Conference on Data Mining, –2010, –p. 905-910.
108. Luo, D., Huang, H. Link Prediction of Multimedia Social Network via Unsupervised Face Recognition // 17th ACM International Conference on Multimedia, – Beijing, China, –2009, –p. 805-808.
109. Maciel, C., Garcia, A.C. DemIL: an online interaction language between citizen and government // 15th International World Wide Web Conference, – Edinburgh, –2006, –p. 849–850.
110. Malik, B. H. Evaluating Citizen e-Satisfaction from e-Government Services: A Case of Pakistan / B.H. Malik, A.Gh. Mastoi, N. Gul [et al.] // European Scientific Journal, –2016. 12 (5), –p.346-370.
111. Marin, A., Wellman, B. Social network analysis: an introduction // Handbook of Social Network Analysis, –London, 2009. –p.11-25.
112. Matsuo, Y., Mori J., Hamasaki, M. POLYPHONET: An Advanced Social Network Extraction System from the Web // World Wide Web Conference,– 2007, –p. 262-278.
113. McCallum, A., Wang, X., Mohanty, N. Joint group and topic discovery from relations and text // Journal of Statistical Network Analysis: Models, Issues and New Directions, –2007. 4503, –p. 28–44.
114. Meishar-Tal, H., Tal-Elhasid, E. Measuring Collaboration in Educational Wikis – A Methodological // Emerging Technologies in Learning, –2008. 3,–p. 46–49.
115. Mesquita, F., Merhav, Y., Barbosa, D. Extracting Information Networks from the Blogosphere: State-of-the-Art and Challenges // The 4th International AAAI Conference on Weblogs and Social Media--Data Challenge, – Washington, – 2010, –p.1-32.
116. Miller, G.A. WordNet: a lexical database for English // Communications on the ACM, – 1995. 38 (11), – p.39-41.
117. Moon, M. J. The Evolution of E-Government among Municipalities: Rhetoric or Reality? // Public Administration Review,– 2002. 62 (4), –p. 424–433.

118. Mostafa, M. M., El-Masry, A. A. Citizens as consumers: Profiling e-government services' users in Egypt via data mining techniques // International Journal of Information Management, –2013. 33 (4), –p. 627–641.
119. Mutuku, L.N., Colaco, J. Increasing kenyan open data consumption: A design thinking approach // 6th International Conference on Theory and Practice of Electronic Governance, –2012, –p. 18–21.
120. Nagwani, N. K. Summarizing large text collection using topic modeling and clustering based on MapReduce framework // Journal of Big Data, –2015. 2 (6), –p.1-18.
121. Nikolenko, S. I., Koltcov, S., Koltsova, O. Topic modelling for qualitative studies // Journal of Information Science, –2017. 43 (1), –p.88-102.
122. Nooy, W. Exploratory social network analysis with Pajek / W. Nooy, A. Mrvar, V. Batagelj,–New York : Cambridge University Press, –2005. –423 p.
123. Opsahl, T., Agneessens, F., Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths // Social Networks, –2010. 32 (3), –p.245–251.
124. Patel, H., Jacobson, D. Factors Influencing Citizen Adoption of EGovernment: A Review and Critical Assessment // 16th European Conference on Information Systems, –Galway, –2008, –p. 176-188.
125. Petricek, V. The Web Structure of E-Government -Developing a Methodology for Quantitative Evaluation / V. Petricek, T. Escher, I. J. Cox [et al.] // International World Wide Web Conference Committee, –2006, –p. 669-678.
126. Phan, X.-H., Nguyen, L.-M., Horiguchi, S. Learning to classify short and sparse text & web with hidden topics from large-scale data collections // 17th International Conference on World Wide Web, –2008, –p. 91–100.
127. Potential and impacts of cloud computing services and social network websites / T. Leimbach, D. Hallinan, D. Bachlechner [et al.], –2014. –127 p.
128. Pourvali, M., Orlando, S., Omidvarborna, H. Topic Models and Fusion Methods: a Union to Improve Text Clustering and Cluster Labeling //

International Journal of Interactive Multimedia and Artificial Intelligence, –2018. 5 (4), –p. 28-34.

129. Provan, G. K., Fish, A., Sydow, J. Interorganizational networks at the Network level: A review of the Empirical literature on Whole Networks // Journal of Management, –2007. 33 (3), –p. 479-516.
130. Ramage, D., Hall, D., Nallapati, R., Manning, C. D. Labeled lda: a supervised topic model for credit attribution in multi-labeled corpora // Conference on Empirical Methods in Natural Language Processing, –2009, –p. 248–256.
131. Rao, G. K., Dey, S. Text Mining Based Decision Support System (TMbDSS) for E-governance: A Roadmap for India // Advances in Computing and Information Technology, –2011. 198, –p.270-281.
132. Rivinius, J. Majority of 2013 terrorist attacks occurred in just a few countries // National Consortium for the Study of Terrorism and Responses to Terrorism (START), –2014. –p. 1-2.
133. Rosen-Zvi, M., Griffiths, T., Steyvers, M., Smyth, P. The author-topic model for authors and documents // 20th Conference on Uncertainty in Artificial Intelligence, –2004, –p. 487–494.
134. Sarrayrih, M. A., Sriram, B. Major challenges in developing a successful e-government: A review on the Sultanate of Oman // Journal of King Saud University – Computer and Information Sciences, –2015. 27, –p. 230-235.
135. Schelin, S. H. E-Government: An overview // Public Information Technology: Policy and Management Issues, –2003. –p. 120−137.
136. Shahkooh, K. A., Saghafi, F., Abdollahi, A. A proposed model for e-Government maturity // 3rd International Conference on Information & Communication Technologies: from Theory to Applications, –2008, –p. 1–5.
137. Shao, M., Qin, L. Text Similarity Computing Based on LDA Topic Model and Word Co-occurrence // 2nd International Conference on Software Engineering, Knowledge Engineering and Information Engineering, –2014, –p.199-203.

138. Shapira, B., Last, M., Elovici, Y., Kandel, A., Zaafrany, O. Using data mining techniques for detecting terror-related activities on the web // Journal of Information Warfare, – 2003. 3 (1), –p.17-28.
139. Sharef, N.M., Martin, T. Evolving fuzzy grammar for crime texts categorization // Applied Soft Computing, – 2015. 28, – p.175-187.
140. Siau, K., Long, Y. Synthesizing e-government stage models–a meta-synthesis based on metaethnography approach // Industrial Management & Data Systems, –2005. 105 (4), –p. 443–458.
141. Solar, M.,Concha, G., Meijueiro, L. A model to assess open government data in public agencies // Lecture Notes in Computer Science,– 2012. –p. 210–221.
142. Song, M., Lee, T., Kim, J. Extraction and Visualization of Implicit Social Relations on Social Networking Services // The 24th AAAI Conference on Artificial Intelligence-AAAI'10, – Atlanta, Georgia, –2010, –p. 1425-1430.
143. Song, W. Using IM and SMS for emergency text communications / W. Song, J. Y. Kim, H. Schulzrinne [et al.] // 3rd International Conference on Principles, Systems and Applications of IP Telecommunications, –2009, –p.1–4.
144. Spalević, Ž., Ilić, M., Spalević, Ž. Electronic Government in the Fight Against Terrorism // International Scientific Conference on Ict and E-business Related Research, – 2016, – p. 518-525.
145. Stanforth, C. Using Actor-Network Theory to Analyze E-Government Implementation in Developing Countries // Information Technologies and International Development,– 2006. 3 (3), –p. 35-60.
146. Stojanovski, D. Social Networks VGI: Twitter Sentiment Analysis of Social Hotspots / D.Stojanovski, I. Chorbev, I. Dimitrovski [et al.] // European Handbook of Crowdsourced Geographic Information, –2016. –p. 223–235.
147. Strang, K. D. Multicultural face of organizations // In The New Faces of Organizations in the 21st Century, –2012. 5, –p. 1-21.
148. Strang, K.D., Sun, Z. Analyzing relationships in terrorism big data using Hadoop and statistics // Journal of Computer Information Systems, –2017. 57(1), –p.67-75.

149. Stylios, G. Public Opinion Mining for Governmental Decisions / G. Stylios, D. Christodoulakis, J. Besharat [et al.] // Electronic Journal of eGovernment, –2010. 8 (2), –p. 203-214.
150. Sun, T. Spam Filtering based on Naive Bayes Classification: [Electronic resource] / –2009. –44 p. URL: http://www.cs.ubbcluj.ro/~gabis/DocDiplome/Bayesian/000539771r.pdf
151. Takale, Sh.A., Nandgaonkar, S. S. Measuring Semantic Similarity between Words Using Web Documents // International Journal of Advanced Computer Science and Applications, –2010. 1 (4), –p.78-85.
152. Tang, J., Wang, T., Wang, J. Measuring the influence of social networks on information diffusion on blogosphere // 8th International Conference on Machine Learning and Cybernetics, –2009, – p 3492-3498.
153. Tasleem, A., Rashid, A., Asger, M. Social Network Extraction: A Review of Automatic Techniques // International Journal of Computer Applications, –2014. 95 (1), – p.16-23.
154. Terrorist Screening Center: [Electronic resource] / FBI. URL: https://www.fbi.gov/about-us/nsb/tsc/
155. Thelwall, M. Sentiment strength detection in short informal text / M.Thetwall, K. Buckley, G. Paltoglou [et al.] // Journal of the American Society for Information Science and Technology, –2010. 61 (12), –p.2544–2558.
156. Thelwall, M., Buckley, K., Paltoglo, G. Sentiment in Twitter Events // Journal of the American Society for Information Science and Technology, –2011. 62 (2), –p. 406–418.
157. Thomas, T. L. Al Qaeda and the Internet: The Danger of “Cyberplanning // Parameters,– 2003.– p. 112-123.
158. Tilahun, T., Sharma, D. P. Design and Development of EGovernance Model for Service Quality Enhancement // Journal of Data Analysis and Information Processing, –2015. 3 (3), –p. 55-62.

159. Tomobe, H., Matsuo, Y., Hasida, K . Social Network Extraction of Conference Participants // 12th International Conference on World Wide Web, –Budapest, Hungary, – May 2003, –p.1-3.
160. United Nations E-Government Survey 2012: E-Government for the People: [Electronic resource] / – New York, – 15 march, 2012. URL: https://www.un.org/en/development/desa/publications/connecting-governments-to-citizens.html
161. Vighne, S. An Approach to Detect Terror Related Activities on Net / Vighne, S. Trimbake, P., Musmade, A. [et al.] // International Journal of Advance Research and Innovative Ideas in Education, –2016. 2 (1), –p.401-406.
162. Wang, J. Detecting Hotspot Information Using Multi-Attribute Based Topic Model / J.Wang, L. Li, F. Tan [et al.] // PLoS ONE, –2015. 10 (10), –p.1-16.
163. Wang, K., Ting, I., Wu, H., Chang, P. A Dynamic and Task-Oriented Social Network Extraction System Based on Analyzing Personal Social Data // 2010 International Conference on Advances in Social Networks Analysis and Mining, – Denmark, –2010, –p 464-469.
164. Wang, S. Research on Cluster Analysis Method of E-government Public Hotspot Information Based on Web Log Analysis / Wang, S., Zhang, J., Yang, F. [et al.] // Computing and Information Technology, –2013. 22, –p. 11–19.
165. Weimann, G. WWW.terror.net: How Modern Terrorism Uses the Internet : [Electronic resource] / Washington DC: United States Institute of Peace, –2004. URL: https://www.usip.org/sites/default/files/sr116.pdf
166. Weimann, G. New Terrorism and New Media // Commons Lab, Research series, –2014. 2, –p.1-20.
167. What is NodeXL?: [Electronic resource] / Social Media Research Foundation. URL: https://www.smrfoundation.org/nodexl/
168. Wimmer, M., Codagnone, C., Janssen, M. Future e-government research: 13 research themes identified in the eGovRTD2020 project // 41st Hawaii International Conference on System Sciences, –Hawaii, USA, –7-10 January, – 2008, – p.1-11.

169. Winder, D. Anti-terror measures: How tech helps fight the counter-terrorism war: [Electronic resource] / – 9 december, 2014. URL: https://www.itpro.co.uk/security/23685/anti-terror-measures-how-tech-helps-fight-the-counter-terrorism-war
170. Windley, P. J. eGovernment Maturity: [Electronic resource] / URL: https://www.windley.com/docs/eGovernment%20Maturity.pdf
171. Wu, Z., Palmer, M. Verb semantics and lexical selection // 32nd Annual Meeting of the Association for Computational Linguistics, –New Mexico, USA, –27-30 June, –1994, –p.133-138.
172. Ya, B. T. Research on public opinion hotspot detection based on SVM // Science and Technology Management Research, –2009. 25 (2), – p. 64–69.
173. Yang, Z., Kankanhalli, A. Innovation in government services: The case of open data / International Working Conference on Transfer and Diffusion of IT, –Bangalore, India, –2013, –p. 644–651.
174. Yau, C.-K. Clustering scientific documents with topic modeling / C.-K. Yau, A.L. Porter, N.C. Newman, A. Suominen // Scientometrics, –2014. 100 (3), –p. 767-786.
175. Yue, L., Qiaozhu, M., Cheng Xiang, Z. Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA // Information Retrieval, –2011.14 (2), –p. 178–203.
176. Zhang, H., Giles, C. L., Foley, H. C., Yen, J. Probabilistic community discovery using hierarchical latent gaussian mixture model // 22nd National Conference on Artificial Intelligence, –2007, –p. 663–668.
177. Zhao, J.J., Zhao, S.Y. Opportunities and threats: security assessment of state e-government websites // Government Information Quarterly, –2010. 27 (1), –p. 49-56.
178. Zhao, L., Wu, L., Huang, X. Using query expansion in graph-based approach for query-focused multi-document summarization // Information Processing & Management, – 2009. 45 (1), –p.35-41.

179. Zuiderwijk, A. Socio-technical impediments of Open Data / A. Zuiderwijk, M. Janssen, S. Choenni [et al.] // Electronic journal of e-Government, –2012. 10 (2),–p. 156-172.

PROQRAM

“Topic modeling və clustering alqoritmi”

```
## sənədlərin daxil edilməsi
library(tm)
params<-
list(minDocfreq=1,removenumbers=TRUE,stopwords=TRUE,stemming=TRUE,weighting=weight
Tf)
setwd("C:/Users/user/Desktop/Eksperiment")
filenames<-list.files(getwd(),pattern="*.txt")
files<-lapply(filenames, readLines)

library(readtext)
docs<-Corpus(VectorSource(files))
corpus.copy <- docs

## sənədlərin təmizlənməsi
corpus <- tm_map(docs, stemDocument, language = "english")
docs <-tm_map(corpus,content_transformer(tolower))
docs <- tm_map(docs, removeWords, stopwords("english"))
docs <- tm_map(docs, removeNumbers)
docs<-tm_map(docs, removePunctuation)

## sənədlərin termdocument matrisi şəklində göstərilməsi
library(SnowballC)
dtm <- TermDocumentMatrix(docs)
dtm1 <- as.matrix(dtm)
dtmm<-TermDocumentMatrix(corpus.copy)
dtm2<-as.matrix(dtmm)
dtm_df<-as.data.frame(dtm1)
remove(dtmm)
rown<-rownames(dtm_df)
dtm_df<-cbind(terms=rown,dtm_df)
dtm_df.copy <- dtm_df
dtm_df1<-dtm_df[,1:2]

for(x in seq(nrow(dtm_df1))) {
  dtm_df1[["ss_typ1"]]<-c("ADJECTIVE")
}
for(x in seq(nrow(dtm_df1))) {
  dtm_df1[["ss_typ2"]]<-c("VERB")
}
for(x in seq(nrow(dtm_df1))) {
  dtm_df1[["ss_typ3"]]<-c("NOUN")
}
for(x in seq(nrow(dtm_df1))) {
  dtm_df1[["ss_typ4"]]<-c("ADVERB")
}
```

```
## Sinonimlərin tapılması
library(wordnet)
setDict("C:/Program Files (x86)/WordNet/2.1/dict")
Sys.setenv(WNHOME = "C:/Program Files (x86)/WordNet/2.1")

dtm_df1$synset1 <- mapply(synonyms, as.character(dtm_df1$terms),
as.character(dtm_df1$ss_typ1))
dtm_df1$synset2 <- mapply(synonyms, as.character(dtm_df1$terms),
as.character(dtm_df1$ss_typ2))
dtm_df1$synset3 <- mapply(synonyms, as.character(dtm_df1$terms),
as.character(dtm_df1$ss_typ3))
dtm_df1$synset4 <- mapply(synonyms, as.character(dtm_df1$terms),
as.character(dtm_df1$ss_typ4))

##Sətirlərə bölmə
library(dplyr)
paste_noNA <- function(x,sep=", ") {
  gsub(", " ,sep, toString(x[!is.na(x) & x!="" & x!="character(0)"] ) )
}
sep=", "
dtm_df1$x <- apply( dtm_df1[ , c(7:10) ] , 1 , paste_noNA , sep=sep)

library(stringr)
dtm_df1$total <- str_count(dtm_df1$x, '\\s+')+1
h<-dtm_df1 %>% filter(total != 1)
h$y <- do.call(rbind, strsplit(as.character(h$x), " "))
test<-cbind(h$x, h$y)
l<-test[,-1]

## Semantik yaxınlığın tapılması
library(dplyr)
row_cf <- function(x, y, df){
  (2*length(intersect(df[x,],df[y,]))) /(h$total[x]+h$total[y])
}

res <- expand.grid(1:nrow(l), 1:nrow(l)) %>%
  rename(row_1 = Var1, row_2 = Var2) %>%
  rowwise() %>%
  mutate(similarity = row_cf(row_1, row_2,l))
res
j<- as.data.frame(res)
p<-j %>% filter(similarity > 0.5, row_1!=row_2)


## sinonimlərin silinməsi
library(tidyverse)
pp<-p[!duplicated(t(apply(p,1,sort))), ]
a<-h[pp$row_1,]
aaa<-a$terms
aaa
```

```
library(tm)
stopwords = aaa
stopwords
stopwords("aaa")
dtm_df$terms <- gsub(paste0(aaa,collapse = "|"),"", dtm_df$terms)
library(dplyr)
dtm_df2<-dtm_df %>% filter(terms != "")

## Tf*IDf tapılması
library(tidyverse)
library(tidytext)
library(tm)
dtm_df3 = dtm_df2[,2:63]
dtm <- as.DocumentTermMatrix(dtm_df3, weighting = weightTfIdf)
dtm<-as.matrix(dtm)
final_df <- as.data.frame(t(dtm))
final_df1<-as.data.frame(t(dtm_df3))

##K-means tətbiq olunması
tic()
library(factoextra)
set.seed(123)
km.res <- kmeans(final_df, 3, nstart = 25)
print(km.res)
{aggregate(final_df, by=list(cluster=km.res$cluster), mean)
  dd <- cbind(final_df, cluster = km.res$cluster)
  head(dd)
  km.res$size
  km.res$centers}


## Klasterlərin tapılması
aggregate(final_df1, by=list(cluster=km.res$cluster), mean)
dd1 <- cbind(final_df1, cluster = km.res$cluster)
head(dd1)
km.res$size
km.res$centers
dd_dd1<-as.data.frame(dd1)
c1<-subset(dd_dd1,dd_dd1$cluster=="1")
c2<-subset(dd_dd1,dd_dd1$cluster=="2")
c3<-subset(dd_dd1,dd_dd1$cluster=="3")
toc()


##Topic modelingin tətbiq olunması
tic()
## 1-ci klaster
library(topicmodels)
ap_lda1 <- LDA(c1, k = 3, control = list(seed = 1234))
ap_lda1
library(tidytext)
```

```
ap_topics1 <- tidy(ap_lda1, matrix = "beta")
ap_topics1
library(ggplot2)
library(dplyr)

ap_top_terms1 <- ap_topics1 %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

ap_top_terms1 %>%
  mutate(term = reorder(term, beta)) %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
library(tidyr)

beta_spread1 <- ap_topics1 %>%
  mutate(topic = paste0("topic", topic)) %>%
  spread(topic, beta)%>%
  filter(topic1 > .001 | topic2 > .001 | topic3 > .001)

beta_spread1
ap_documents1 <- tidy(ap_lda1, matrix = "gamma")
ap_documents1

tidy(dtm_train) %>%
  filter(document == 3) %>%
  arrange(desc(count))
deprecated


## 2-ci klaster
ap_lda2 <- LDA(c2, k = 3, control = list(seed = 1234))
ap_lda2
library(tidytext)

ap_topics2 <- tidy(ap_lda2, matrix = "beta")
ap_topics2
library(ggplot2)
library(dplyr)

ap_top_terms2 <- ap_topics2 %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

ap_top_terms2 %>%
```

```
  mutate(term = reorder(term, beta)) %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
library(tidyr)

beta_spread2 <- ap_topics2 %>%
  mutate(topic = paste0("topic", topic)) %>%
  spread(topic, beta)%>%
  filter(topic1 > .001 | topic2 > .001|topic3 > .001 )

beta_spread2
ap_documents2 <- tidy(ap_lda2, matrix = "gamma")
ap_documents2


## 3-cü klaster
ap_lda3 <- LDA(c3, k = 3, control = list(seed = 1234))
ap_lda3
library(tidytext)

ap_topics3 <- tidy(ap_lda3, matrix = "beta")
ap_topics3
library(ggplot2)
library(dplyr)

ap_top_terms3 <- ap_topics3 %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

ap_top_terms3 %>%
  mutate(term = reorder(term, beta)) %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()

ap_top_terms3
library(tidyr)

beta_spread3 <- ap_topics3 %>%
  mutate(topic = paste0("topic", topic)) %>%
  spread(topic, beta) %>%
  filter(topic1 > .001 | topic2 > .001 |topic3 > .001 )

beta_spread3
ap_documents3 <- tidy(ap_lda3, matrix = "gamma")
ap_documents3
toc()
```