

AZƏRBAYCAN RESPUBLİKASI

Əlyazması hüququnda

MİLLİ TERMİNOLOJİ İNFORMASIYA SİSTEMİNİN SİNTEZİ ÜÇÜN METOD VƏ ALQORİTMLƏRİN İŞLƏNMƏSİ

İxtisas: 3338.01 – Sistemli analiz, idarəetmə və
informasiyanın işlənməsi (informasiya texnologiyaları)

Elm sahəsi: Texnika elmləri

İddiaçı: **Əfruz Muxtar qızı Qurbanova**

Fəlsəfə doktoru elmi dərəcəsi almaq üçün
təqdim edilmiş dissertasiyanın

AVTOREFERATI

Bakı – 2024

Dissertasiya işi Azərbaycan Respublikası Elm və Təhsil Nazirliyinin
İnformasiya Texnologiyaları İnstitutunda yerinə yetirilmişdir.

Elmi rəhbər:

AMEA-nın həqiqi üzvü,
texnika elmləri doktoru, prof.
Rasim Məhəmməd oğlu Əliquliyev

Rəsmi opponentlər:

texnika elmləri doktoru, prof.

Nadir Bafadin oğlu Ağayev

texnika elmləri doktoru, prof.

Ramin Rza oğlu Rzayev

texnika üzrə fəlsəfə doktoru

Elviz Ənvər oğlu İsmayilov

Azərbaycan Respublikasının Prezidenti yanında Ali Attestasiya
Komissiyasının Azərbaycan Respublikası Elm və Təhsil Nazirliyinin
İnformasiya Texnologiyaları İnstitutunun nəzdində fəaliyyət göstərən
ED1.35 Dissertasiya Şurası

Dissertasiya şurasının sədri:

AMEA-nın həqiqi üzvü,
texnika elmləri doktoru, prof.

Rasim Məhəmməd oğlu Əliquliyev

Dissertasiya şurasının elmi katibi:

texnika üzrə fəlsəfə doktoru, dosent

Fərqanə Cabbar qızı Abdullayeva

Elmi seminarın sədri:

texnika elmləri doktoru, dosent

Zərifə Qasım qızı Cəbrayilova

İŞİN ÜMUMİ XARAKTERİSTİKASI

Mövzunun aktuallığı. Hazırda informasiya-kommunikasiya texnologiyaları (İKT) insanların həyat tərzinə çevrilməklə, bütün fəaliyyət sahələrinə nüfuz etməkdədir. Digər tərəfdən, müasir dövr qloballaşmanın yüksək tempi ilə xarakterizə olunur. Qloballaşma prosesinin və mədəniyyətlərarası kommunikasiyanın sürətlənməsi təbii dillərdə də müəyyən dəyişikliklərə səbəb olur. Belə bir dövrdə dillərin qorunması, inkişafı və tətbiqi imkanlarının genişləndirilməsi istiqamətində zəruri tədbirlərin görülməsi vacibdir. Qloballaşma şəraitində müxtəlif təhlükələrlə üz-üzə qalmış milli dillərin qorunması və inkişafı istiqamətində e-dövlət platformasında milli dillərin yeri və rolunun müəyyənləşdirilməsi aktual məsələlərdəndir. Dilin sosial funksiyasının inkişafında milli terminologiyanın sistemləşdirilməsi və inkişafı xüsusi əhəmiyyət daşıyır.

Ölkəmizdə İnternetin cəmiyyətin müxtəlif sferalarına sürətlə nüfuz etməsi, insanların müxtəlif bilik sahələrinə aid məlumatlardan sərbəst istifadə etməsi nəticəsində milli terminoloji mühiti nəzarətdə saxlamaq çox çətin məsələyə çevrilib. Son illər dilin alınma sözlərlə zənginləşməsi də təbii prosesdir. Bu prosesin qarşısına sədd çəkmək, eyni zamanda, dilin alınma sözlərlə lüzumsuz yerə yüklənməsinə də yol vermək olmaz. Real vəziyyətin yaranmasına səbəb terminoloji lüğətlərin hamı üçün istifadəsinin əlçatan olan vahid məlumat bazası və ya terminoloji reyestrinin olmamasıdır. Müxtəlif ölkələrdə yaşayan azərbaycanlıların müvafiq xarici dillərin təsiri altında olması, Azərbaycanda gənclərin ingilis və rus dillərində təhsilinin genişlənməsi, İnternetin insanların gündəlik fəaliyyətlərinin bir hissəsinə çevrilməsi nəticəsində fərqli terminoloji mühitlər formalaşmışdır. Terminoloji bazaların vahid, ümumi bir mühitə gətirilməsinə, Azərbaycan dili ekosisteminin tərkib hissəsi kimi hamı üçün əlçatan milli terminoloji informasiya sistemə ehtiyac yaranmışdır. Əvvəlki dövrlərlə müqayisədə son zamanlar İKT-nin geniş tətbiqi nəticəsində informasiyanın, o cümlədən elmi-innovativ biliklərin sürətlə inkişafı bir tərəfdən terminləşmə prosesini daha da dinamik edir, digər tərəfdən isə məkandan, zamandan asılı olmadan informasiyanın çox asanlıqla əldə olunması, yayılması və populyarlaşması de-terminləşmə prosesini sürətləndirir. Yuxarıda

deyilənlər terminoloji informatika (computational terminology) sahəsində tədqiqatları şərtləndirir. Terminoloji informatika təbii dilin emalı sahələrində - mətnin intellektual analizi, informasiya axtarışı, informasiyanın çıxarılması, mətnin təhlili, ontologiyanın qurulması və s. mühüm aspekti əhatə edir. Beləliklə, ölkəmizdə formalaşmaqda olan e-dövlət mühitində Azərbaycan dili ekosisteminin tərkib hissəsi kimi hər bir kəs üçün açıq, əlçatan Milli terminoloji informasiya sisteminin (MTİS) yaradılması zamanın tələbinə çevrilmişdir. Dissertasiya işində bunlar nəzərə alınaraq, MTİS-in formalaşmasının elmi-nəzəri və texnoloji problemləri araşdırılmış, sistemin sintezi və intellektuallaşdırılması üçün yeni yanaşma, metod və alqoritmlər təklif edilmişdir.

İşin məqsədi beynəlxalq təcrübə və standartlar əsasında e-dövlət platformasında Azərbaycan dili ekosisteminin tərkib hissəsi kimi Milli terminoloji informasiya sisteminin sintezi üçün metod və alqoritmlərin işlənməsidir.

Qarşıya qoyulmuş məqsədə nail olmaq üçün dissertasiya işi çərçivəsində aşağıdakı məsələlər müəyyən edilmişdir:

- Terminoloji informatika sahəsində mövcud vəziyyətin analiz olunması və elmi-nəzəri və texnoloji problemlərinin müəyyən olunması;
- MTİS-in formalaşması sahəsində mövcud yanaşmaların müqayisəli analizi və elmi-texnoloji problemlərinin müəyyən olunması;
- MTİS-in konseptual modeli və arxitektur-texnoloji prinsiplərinin işlənməsi;
- MTİS-də OLAP (ing. online analytical processing – onlayn analitik emal) əsasında qərarların qəbulunu dəstəkləyən sistemin (QQDS) işlənməsi;
- MTİS çərçivəsində dinamik proseslərin modelləşdirilməsi, analizi və qiymətləndirilməsi üçün model, metod və alqoritmin işlənməsi;
- MTİS əsasında fərdlərin nitqinin populyarlıq səviyyəsinin qiymətləndirilməsi üçün metodun işlənməsi;
- MTİS-in reallaşdırılması üçün təkliflərin işlənməsi.

Tədqiqatın obyektı və predmeti. Tədqiqatın obyektı MTİS-in sintezidir. Tədqiqatın predmeti MTİS-in sintezi proseslərinin idarə edilməsinin model və metodlarıdır.

Tətbiq olunan tədqiqat metodları. Dissertasiya işində nəzərdən keçirilən məsələlərin həlli zamanı təbii dilin emalı, OLAP, data mining, text mining, qərarların qəbulu nəzəriyyəsi, populyasiya modeli, statistik analiz metodlarından istifadə olunmuşdur.

Müdafiyyə çıxarılan əsas müddəalar. Dissertasiya işinin elmi yeniliyə malik olan mühüm nəticələri aşağıdakılardır:

- MTİS-in konseptual və arxitektur-texnoloji modeli;
- MTİS-də OLAP və Verilənlər Saxlancısı (ing. Data Warehouse) texnologiyalarına əsaslanan qərarların qəbulunu dəstəkləyən sistem;
- MTİS-in qiymətləndirilməsi üçün polikubik OLAP model;
- MTİS çərçivəsində dinamik proseslərin modelləşdirilməsi üçün populyasiya artım modelinə əsaslanan metod;
- MTİS çərçivəsində yeni terminlərin uyğun fənlər üzrə avtomatik paylanması üçün k-means klasterləşmə alqoritmi və Naive Bayes təsnifatlandırma modelinə əsaslanan metod;
- MTİS əsasında fərdlərin nitqlərinin populyarlıq səviyyəsinin qiymətləndirilməsi üçün statistik tezlik modelinə əsaslanan metod.

Tədqiqatın elmi yeniliyi. Tədqiqatın elmi yeniliyi aşağıdakı nəticələrlə təyin edilir:

- MTİS-in konseptual və arxitektur-texnoloji modeli işlənmişdir;
- MTİS-də toplanan verilənlər və metaverilənlərin operativ analizi üçün OLAP texnologiyalarına əsaslanan qərarların qəbulunu dəstəkləyən sistem və polikubik OLAP-model işlənmişdir;
- MTİS çərçivəsində dinamik proseslərin modelləşdirilməsi üçün Maltusun populyasiya artımı modelinə əsaslanan metod işlənmişdir;
- MTİS çərçivəsində yeni terminlərin uyğun fənlər üzrə avtomatik paylanması üçün k-means klasterləşmə alqoritminə və Naive Bayes təsnifatlandırmaya əsaslanan metod işlənmişdir;
- MTİS əsasında fərdlərin nitqlərinin populyarlıq səviyyəsinin

qiymətləndirilməsi üçün statistik tezlik modelinə əsaslanan metod işlənmişdir.

İşin praktiki əhəmiyyəti. Dissertasiyada təklif olunan yanaşmalar MTİS-in intellektuallaşdırılması, sistem çərçivəsində terminoloji proseslərin qiymətləndirilməsi və monitorinqi, dilin təhlükəsizliyi, xarici təsirlərdən qorunması və yeni terminlərin milli xüsusiyyətlərə uyğunlaşdırılması, termin yaradıcılığının səmərəliliyinin yüksəldilməsi, terminoloji lüğətləri hazırlamış alim və mütəxəssislərin müəlliflik hüquqlarının qorunması prosesində, e-terminoloji xidmətlərin intellektuallaşdırılmasında istifadə oluna bilər.

İşin aprobeasiyası. Əsas elmi-nəzəri və praktiki nəticələr aşağıda adı çəkilən konfranslarda məruzə edilmiş və müzakirə olunmuşdur: PCI'2010. The Third International Conference Problems of Cybernetics and Informatics (Baku, 2010); «TurkLang 2018» VI International Conference on Computer Processing of Turkic Languages, (Tashkent, 2018); Международная конференция «Развитие информатизации и государственной системы научно-технической информации», РИНТИ-2019 (Минск, 2019); The 4th International Symposium on Computer Science, Digital Economy and Intelligent Systems (CSDEIS2022), (Wuhan, China, 2022); 2023 IEEE 17th International Conference on Application of Information and Communication Technologies (AICT), (Baku, 2023); 2nd International Conference on Information Technologies and Their Application (ITTA 2024), (Baku, 2024).

Elmi nəşrlər. Dissertasiya mövzusu üzrə 20 elmi iş çap olunmuşdur. Onlardan 13 məqalə resenziya olunan jurnallarda, 7 tezis konfrans materiallarında nəşr edilmişdir. Bu elmi işlərdən 1 məqalə “Web of Science”, 3 məqalə “Scopus” bazasında indeksləşən jurnallarda çap edilmişdir.

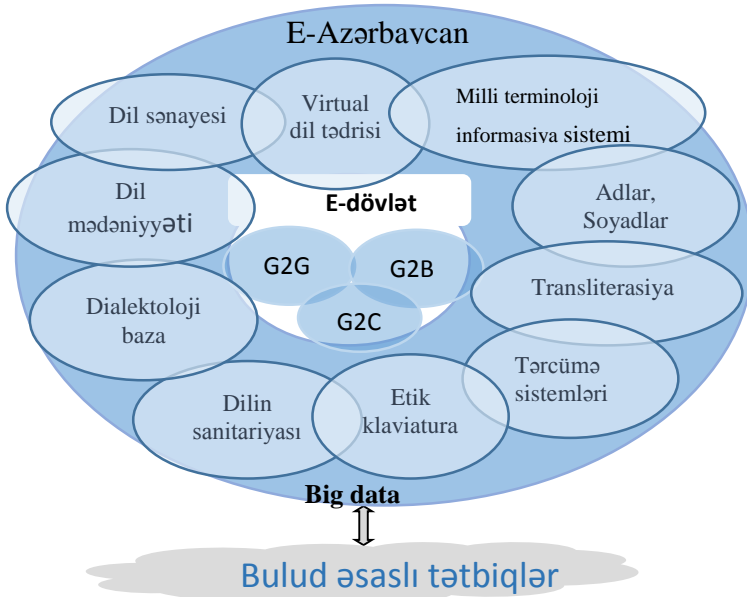
Dissertasiya işinin strukturu və həcmi. Dissertasiya işi giriş, 4 fəsil, nəticə, 170 adda ədəbiyyat siyahısı və iki əlavə, 7 cədvəl və 25 şəkildən ibarətdir. İşin ümumi həcmi 121 vərəq məşinlə yazılmış mətndir.

DİSSERTASIYA İŞİNİN MƏZMUNU

Girişdə dissertasiya işinin aktuallığı əsaslandırılmış, tədqiqatın məqsədi və həll olunacaq məsələlər müəyyən edilmişdir. Əldə edilmiş nəticələrin elmi yeniliyi və praktiki əhəmiyyəti göstərilmişdir.

Birinci fəsilə (“**Terminoloji informatikanın elmi-nəzəri problemlərinin tədqiqi və təhlili**”) dilin qorunması və inkişafı, müasir dövrdə terminologiyanın rolu və təşəkkül tarixi, bu sahədə xarici təcrübə və beynəlxalq standartlar təhlil olunmuş, terminoloji informatikanın əsas elmi-texnoloji problemləri analiz olunmuşdur.

E-dövlət mühitində dilin qorunması ilə bağlı təşəbbüslərin yerinə yetirilməsi istiqamətində e-dövlət platformasında Azərbaycan dili ekosisteminin konseptual modeli təklif edilmişdir (Şəkil 1). Azərbaycan dili ekosisteminin tərkib hissəsi kimi Milli terminoloji informasiya sisteminin işlənməsi, sintezi ilə əlaqəli həlli vacib olan əsas məsələlər müəyyənləşdirilmişdir [1, 2, 4, 5, 6, 12, 16, 18].



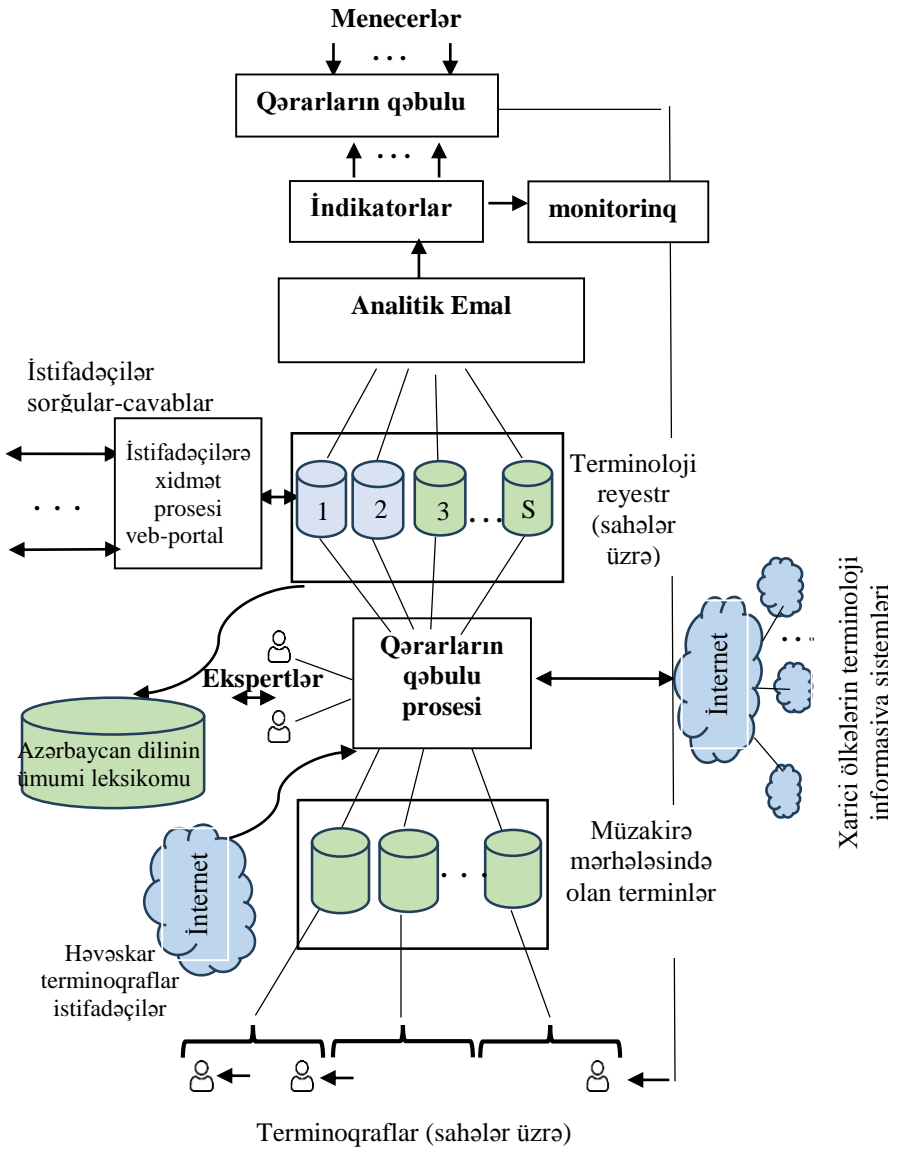
Şəkil 1. E-dövlət platformasında Azərbaycan dili ekosisteminin konseptual modeli

İkinci fəsilə (“Milli terminoloji informasiya sisteminin konseptual əsaslarının işlənməsi”) MTİS-in konseptual modeli və arxitektur-texnoloji prinsipləri işlənməmiş, milli terminoloji informasiya sisteminin qiymətləndirilməsi üçün OLAP və Verilənlər Saxlanıcı (VS) texnologiyalarına əsaslanan QQDS təklif olunmuş və verilənlər analiz olunmuşdur.

İkinci fəslin birinci bölməsində MTİS-in konseptual modeli təklif edilmişdir [3, 7, 10]. Təklif olunan konseptual modelin funksional alt sistemləri aşağıda ətraflı izah olunmuşdur.

Təklif olunan model. MTİS-nin əsas nüvəsini bütün terminoloji lüğətlərin vahid onlayn mühitdə etibarlı saxlanmasını təmin edən və hamı üçün əlçatan olan Terminoloji reyestr təşkil edir (Şəkil 2). Müxtəlif illərdə dərc olunmuş, həmçinin daşıyıcılarda mövcud olan terminoloji lüğətlər toplanılır, rəqəmsallaşdırılır və terminoloji reyestrə daxil edilir. Terminoloji veb-portal vasitəsilə terminoloji reyestrə toplanmış məlumatlar əsasında internet vasitəsilə istifadəçilərə e-terminoloji xidmətlər göstərilir. MTİS-in sintezi ilə termin yaradıcılığı prosesi əvvəlki dövrə nəzərən xeyli asanlaşacaq. Belə ki, hər bir alim və mütəxəssis yeni termin yaradarkən MTİS-in Terminoloji reyestrinə müraciət edir ki, bu söz əvvəllər termin kimi qeydə alınmışdır? Əgər varsa, onda yeni termin qəbul etməyə ehtiyac qalmır (ölkədaxili harmonikləşdirmə). Əgər terminoloji reyestrə bu termin yoxdursa, onda MTİS ilə inteqrasiya olunmuş digər ölkələrin terminoloji informasiya sistemlərində həmin termin axtarılır, təhlil olunur və daha səmərəli qərar qəbul olunur (beynəlxalq harmonikləşdirmə). Terminoloji ekspertiza altsistemi çərçivəsində terminoqraflar tərəfindən təqdim olunmuş lüğətlər və digər təkliflərin həm əənəvi, həm də onlayn mühitdə ictimai müzakirəsi və qərarların qəbulu işləri təmin edilir. Terminoqraflar altsistemi vasitəsilə alim və mütəxəssislər (peşəkar terminoqraflar), həmçinin termin yaradıcılığı işləri ilə məşğul olmaq istəyən vətəndaşlar (həvəskar terminoqraflar) tərəfindən hazırlanmış lüğətlərin ekspertizası həyata keçirilir.

Analitik emal altsistemi terminoloji reyestrə və veb-portalda toplanmış məlumatlar əsasında müxtəlif təyinatlı monitorinqlər, təhlillər və qiymətləndirmələrin aparılması funksiyasını həyata keçirir.



Şəkil 2. Milli terminoloji informasiya sisteminin arxitekturu

Belə ki, MTİS-in Terminoloji reyestrində toplanmış lüğətləri, terminləri xarakterizə edən parametrləri, istifadə reytingini və s. göstəriciləri təhlil etməklə dəyərli biliklər əldə etmək olar.

Beləliklə, MTİS-in sintezi terminoloji informatika və digər əlaqəli sahələr üzrə araşdırmaların aparılması, sistemin intellektuallaşdırılması və səmərəlilik göstəricilərinin yaxşılaşdırılması, funksional imkanlarının genişləndirilməsi, innovativ texnologiyaların tətbiqi üçün geniş imkanlar açır. MTİS dinamik inkişaf edən, İnternet mühitində istismar olunan, açıq milli informasiya resursudur. MTİS yaradılarkən beynəlxalq standartların tələbləri nəzərə alınmış və istifadəçilərə e-terminoloji xidmətlər göstəriləcəkdir. Terminoloji lüğətlərin vahid informasiya sistemində toplanması terminoloji tədqiqatların effektivliyini artıracaqdır. MTİS-in əsas əhəmiyyəti ondan ibarətdir ki, onun sayəsində Azərbaycan terminoloji bazası formalaşacaq.

İkinci fəslin ikinci bəndində MTİS-də OLAP və VS texnologiyalarına əsaslanan QQDS-in strukturu işlənmişdir [13, 14]. Əsas məqsəd MTİS-in Terminoloji reyestrində toplanmış terminlər və onları təyin edən metaverilənləri analiz etməklə dəyərli informasiyanın əldə edilməsidir. QQDS-in imkanları aşağıda ətraflı şərh olunur.

Qərarların qəbulunu dəstəkləyən sistem. Bu cür sistemlər informasiyanın operativ emalını, hazır hesabatların verilməsini təmin edir, çoxfunksionallıq kimi mühüm keyfiyyətə malikdir.

QQDS-in əsas məsələləri – verilənlərin yığılması, saxlanması, emalı və analizidir. Bu məsələlərin həlli üçün informasiya saxlanması xüsusi strukturu olan VS işlənmişdir.

Verilənlər saxlancı. VS baza, tarixi, inteqrasiya olunmuş verilənlər və metaverilənlərdən ibarətdir. VS-da müxtəlif mənbələrdən olan bircins verilənlər inteqrasiya olunaraq problem-yönümlü, strukturlaşmış olur. VS-nin struktur elementləri faktlar (Fact Table) və ölçülər cədvəlidir (Dimension Tables). VS-də yeni terminin meydana çıxması fakt olacaq, ölçü isə tarix, terminin mənbəyi, fəaliyyət sferası, terminin dili və s. olacaq. Təqdim edilən VS-də “Tarix” – “il – ay – gün” iyerarxik səviyyələrinə, “Dil” – “dil ailəsi – dil qrupu” iyerarxik səviyyələrinə malikdir. VS-nin arxitekturunun düzgün seçilməsi sistemin uğurlu işləməsində əsas amildir.

İkinci fəslin üçüncü bəndində MTİS-in Terminoloji reyestrinə toplanmış verilənlərin analitik emalı üçün işlənən polikubik OLAP-model verilmişdir [14].

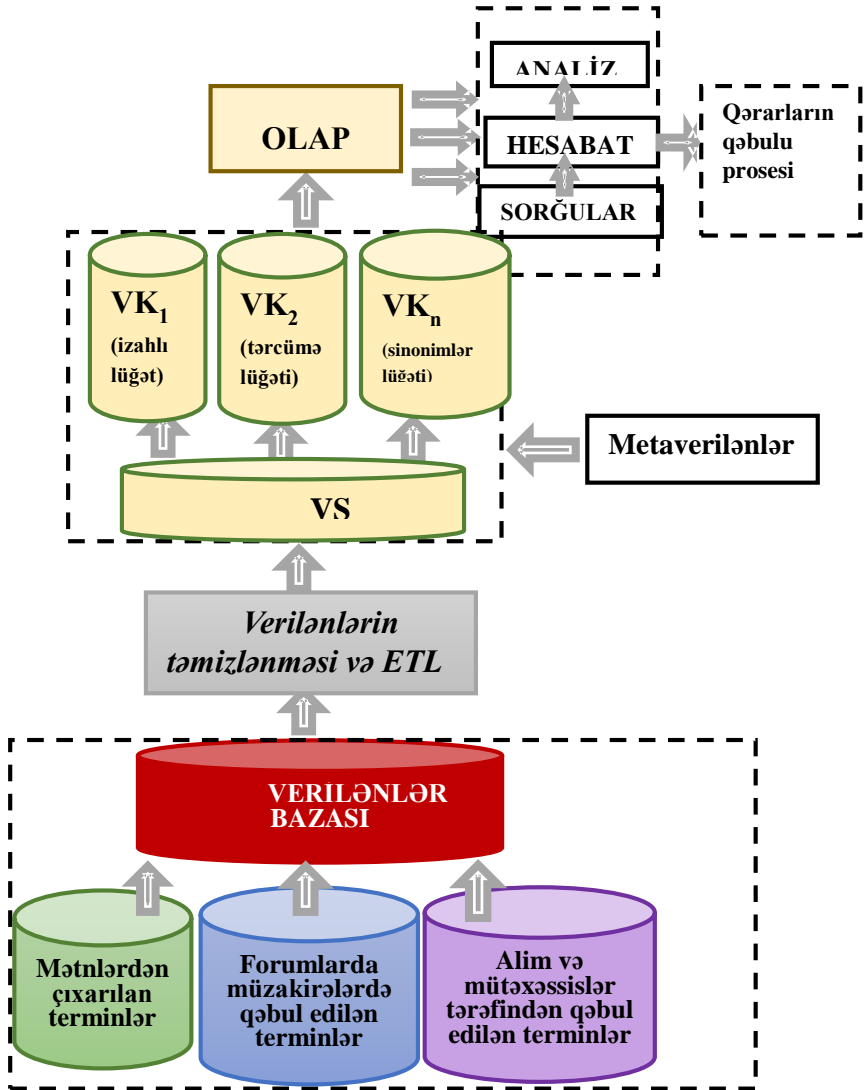
Polikubik OLAP-model. OLAP – müxtəlif mənbələrdən olan məlumatları səmərəli şəkildə təhlil etməyə imkan verən, əsaslandırılmış biznes qərarları qəbul etmək üçün informasiyaya sürətli çıxışı təmin edən texnologiyadır. Məqsəd verilənlərin toplanması, təşkili, ümumiləşdirilməsi və analizidir. OLAP son istifadəçiyə daha effektiv qərarların qəbulu üçün zəruri olan informasiyanı təqdim edərək bir neçə ölçüdə verilənlərin analizinə imkan verir. Tipik OLAP əməliyyatları kəsiklərdən ibarətdir – *dilim* və *kəsim* (slice&dice), *aqreqatlaşdırma* (roll-up) və *detallaşdırma* (drill-down). Aqreqatlaşdırma və detallaşdırma iyerarxik struktura malik ölçülər üzərində aparılır.

MTİS üçün QQDS-in arxitektur-texnoloji modeli təklif edilib (Şəkil3). Bu, VS-nin üçsəviyyəli modelinə əsaslanır.

I səviyyə. QQDS-in birinci səviyyəsi MTİS-in Terminoloji reyestrində olan verilənlərin müxtəlif mənbələri təqdim edir. Bu, mətnlərin avtomatik emalı nəticəsində əldə olunan, forumlarda keçirilən müzakirələr nəticəsində alınan, müəyyən bir sahənin alim və mütəxəssisləri tərəfindən qəbul edilən və s. terminlər ola bilər. İkinci səviyyəyə keçid zamanı aralıq mərhələdə *ETL* (*Extraction, Transformation, Loading*) - verilənlərin çıxarılması, yenisinə çevrilməsi, yüklənməsi texnologiyasının vasitəsilə verilənlərin təmizlənməsi prosesi həyata keçirilir.

II səviyyə. VS-ni təqdim edir, o həm cari, həm də tarixi verilənləri özündə saxlayır. Terminlərə aid tarixi verilənlər de-terminləşmiş, həmçinin arxaizmlər kateqoriyasına keçmiş terminlərdir. VS-dən verilənlər köşkləri (VK) ayrılır. Hər bir VK əlaqəli məsələlərin həllinə istiqamətlənmiş verilənləri saxlayır. Hazırkı məsələdə VK lüğətlərin – izahlı, tərcümə (az-rus, az-ing. və s.), sinonimlər, tezauslar və s. tərtibatına yönəlib. Bunun üçün VS və onun əsasında qurulan VK bütün zəruri verilənlərə malik olmalıdır.

III səviyyə. QQDS-də VS-nin əsas komponenti olan OLAP texnologiyasıdır. OLAP-kub əsasında verilənlər analiz olunur.



Şəkil 3. Milli terminoloji informasiya sistemi üçün QQDS-in arxitektura-texnoloji modeli

OLAP-kub istənilən sayda ölçüyə malik ola bilər. Burada analiz aparılması üçün üç ölçü seçilib, onlar VS-nin atributlarıdır: tarix (Date),

fəaliyyət sahəsi (Field of activity), dil (Language). OLAP-kubun təqdim edilən nümunəsində onun oyuqları müvafiq tarixə daxil olan, uyğun fəaliyyət sahəsinə aid, müvafiq dildə yaradılan terminlərin sayını əks etdirən ədədləri özündə saxlayır. Bu ölçülərdən ikisi – dil və tarix iyerarxik olduğu üçün onlarda yerinə yetirilən SUM() aqreqat funksiyasının nəticələri yerləşdirilir. OLAP-kub bu qiymətləri detallaşdırmağa kömək edir. Bu əməliyyatların aparılması nəticəsində ikiölçülü cədvəl şəklində təqdim olunan müvafiq göstəricilərdən ibarət verilənlər alınmışdır.

Təklif edilən OLAP və VS texnologiyaları əsasında işlənən QQDS MTİS-i daha təkmil və dayanıqlı edəcək. O QQŞ-ə əhəmiyyətli dəstək göstərəcək, terminologiya sahəsində analitik fəaliyyəti təkmilləşdirməyə imkan verəcək, bu sahədə elmi tədqiqatların aparılmasında mühüm rol oynayacaqdır.

Üçüncü fəsilə (Milli terminoloji informasiya mühitinin intellektual analizi üçün metod və alqoritmlər) MTİS çərçivəsində termin yaradıcılığı proseslərinin avtomatlaşdırılması, dinamik proseslərin modelləşdirilməsi və intellektual analizi, nitqin populyarlıq səviyyəsinin qiymətləndirilməsi üçün metod və alqoritmlər işlənmişdir.

Üçüncü fəslin birinci bəndində termin yaradıcılığı proseslərinin avtomatlaşdırılması məsələsinə baxılmışdır [8, 9, 11, 15]. Termin yaradıcılığı müxtəlif bilik və təcrübə sahələrində terminlərin yaradılması, sistemləşdirilməsi, standartlaşdırılması və istifadəsi prosesidir.

Terminoloji proseslər dilin terminoloji sistemində baş verən dəyişikliklər məcmusudur. Əsas terminoloji proseslər - terminləşmə, de-terminləşmə, terminlərin semantik dəyişiklikləri, terminlərin homonimləşdirilməsidir. MTİS çərçivəsində terminoloji proseslərin analizi və qiymətləndirilməsi üçün indikator təklif edilmişdir. Təklif olunan yanaşma aşağıda ətraflı izah edilmişdir:

Terminoloji proseslərin qiymətləndirilməsi indikatoru. MTİS-in Terminoloji reyestrində (TR) olan hər bir termini (müzakirəyə qəbul olunmuş və təsdiq olunmuş) onu müşayiət edən metaverilənlər ilə təmin edək. Bu metaverilənlərə termini təqdim edən təşkilat və müəlliflərin (terminoqrafların) adı, ünvanı, terminin müzakirəyə təqdim edildiyi vaxt, həmin terminin təsdiq edildiyi vaxt, terminin hansı elmi sahəyə

aid olmasını göstərən əlamət, terminin hansı dildən alınması haqqında məlumat, həmin dildə terminin adı (yazılışı), terminin etimologiyası və s. aiddir. Formal olaraq s -ci bilik sahəsindən olan i -ci terminə aid metaverilənləri aşağıdakı vektorla işarə edək:

$$T_i^{(s)} = \left\langle t_{ij}^{(s)} \mid j \equiv \overline{1, m_s} \right\rangle, s = \overline{1, S}, i = \overline{1, n_s} \quad (1)$$

Burada $t_{ij}^{(s)}$ – s -ci bilik sahəsinə aid olan i -ci terminin j -ci atributudur (metaverilənləridir). TR terminlər (t_i) və onları müşayiət edən metaverilənlər bazalarından ibarətdir:

$$TR = \bigcup_{s=1}^S \bigcup_{i=1}^{n_s} (t_i, T_i^{(s)}) = \bigcup_{s=1}^S TR_s \quad (2)$$

$$TR_s = \bigcup_{i=1}^{n_s} (t_i^{(s)}, T_i^{(s)}), s = \overline{1, S} \quad (3)$$

Burada n_s dəyişəni s -ci bilik sahəsinə aid olan terminlərin sayı, m_s isə bu terminləri xarakterizə edən metaverilənlərin sayıdır.

Terminoloji prosesləri analiz etmək, qiymətləndirmək, monitorinqini aparmaq, müxtəlif bilik sahələri arasındakı müqayisələr aparmaq və s. məqsədlər üçün $n_s(t)$ indikatorunu daxil edək. Burada $n_s(t)$ t -ci zaman, s -ci bilik sahəsi üzrə Terminoloji reyestrədə olan terminlərin sayıdır. Təbii olaraq, Terminoloji reyestrədə t -ci zamana görə terminlərin ümumi sayı $N_s(t)$ aşağıdakı kimi olacaqdır:

$$N_s(t) = \sum_{s=1}^S n_s(t) \quad (4)$$

s -ci sahə üzrə termin yaradıcılığın prosesinin inkişaf dinamikasını xarakterizə etmək üçün $\Delta n_s(t)$ indikatorunu daxil edək:

$$\Delta n_s(t) = n_s(t) - n_s(t-1), t=1,2,\dots,n \quad (5)$$

Bu indikatorun köməyiylə bilik sahələri üzrə termin yaradıcılığın prosesinin tərəqqisini (tənəzzülünü) müəyyənləşdirmək olar. İki sahənin ($U_s(t)$ və $U_{s'}(t), s \neq s'$) kəsişməsinin dinamikasına baxaq:

$$|U_s(t) \cap U_s(t)| = \Delta U_{ss}(t) \quad (6)$$

Əgər $\Delta U_{ss}(t) > \Delta U_{ss}(t-1)$ şərti ödənərsə 2 sahənin inteqrasiyası baş verib; əgər $\Delta U_{ss}(t) < \Delta U_{ss}(t-1)$ şərti ödənərsə 2 sahənin diferensiyası baş verib. Bunu 3 sahə, 4 sahə və s. arasında da aparmaq olar. Bu MTİS-in həyata keçirdiyi əsas funksiyalardan biridir.

Üçüncü fəslin birinci bəndində predmet sahəsinin terminlərinin semantik şəbəkəsinin avtomatik qurulması və terminlərin mətnlərdən avtomatik çıxarılması texnologiyaları da təhlil edilmiş, tətbiqi yanaşma verilmişdir [8, 9, 11, 15].

Beləliklə, Azərbaycanda müxtəlif sahələr üzrə hazırlanmış terminoloji lüğətlərin vahid informasiya sistemində toplanması nəticəsində termin yaradıcılığının səmərəliliyi yüksələcəkdir. MTİS-in konsepsiyası çərçivəsində Milli terminoloji veb-portal yaradılmışdır ki, bu da Azərbaycanda termin yaradıcılığı işlərinin avtomatlaşdırılması istiqamətində mühüm addımdır.

Üçüncü fəslin ikinci bəndində terminləşmə və de-terminləşmə prosesləri təhlil olunmuş, MTİS çərçivəsində dinamik proseslər riyazi qanunauyğunluqlar əsasında formalaşdırılmış və sistemin ümumi analizi üçün populyasiya modelinə əsaslanan metod təklif edilmişdir [17]. Təklif olunan yanaşmanın hər bir mərhələsi aşağıda ətraflı izah edilmişdir:

De-terminləşmə həddi. Köhnə terminlərin yeni məna alması – de-terminləşməsi və yeni terminlərin yaranması terminologiyanın dinamikliyinin nəticəsidir. Terminlərin de-terminləşməsi həddi həmin termini istifadə edən xalqın sayından asılıdır, yəni, bu termin həmin xalqın ümumi leksikonunda istifadə olunma tezliyi ilə müəyyənləşir. Başqa sözlə, əgər T terminlər bazasından olan T_i termininin İnternet və digər resurslarda istifadə olunma sayı f_i , baxılan xalqın sayı P olarsa, onda

$$\frac{f_i}{P} \leq d, i \in T, d \in [0,1] \quad (7)$$

şərtini ödəyənlər termin, ödəməyənlər isə de-terminləşmiş sayılır. Bu göstərici həmin xalqın savadlılığından, təhsilindən asılıdır. Bu həddi

analiz etməklə cəmiyyətdə müxtəlif dövrlər üçün biliklərin populyarlaşması, eyni zamanda xalqın savadlanması haqqında müəyyən biliklər əldə etmək olar.

Populyasiya modeli. Terminlər müəyyən ömür yaşamış insanlar tərəfindən yaradılır, onda insan populyasiyasının sayı ilə yeni terminlərin formalaşması prosesi arasında analogiya aparmaq olar. Populyasiya dinamikasının modelinə analogi olaraq terminlərin say dinamikasına görə

$$T'(t) = kT(t) \quad (8)$$

olacaq, burada k - artım sürətidir. $T'(t)$ t - anında TB-yə daxil olan terminlərin sayı, $T(t)$ t anına qədər TB-də olan terminlərin sayıdır. Göründüyü kimi, t zamanında terminlərin artım sürəti həmin dövrdəki terminlərin sayı ilə mütənasibdir. t_0 başlanğıc anında terminlərin sayını T_0 kimi təyin edək:

$$T(t_0) = T_0 \quad (9)$$

(8) və (9)-un ümumi həlli eksponensial funksiyadır:

$$T(t) = T_0 e^{kt} \quad (10)$$

e - natural logarifmin əsasını təşkil edir. Aydındır ki, $k > 0$ olarsa, onda terminlərin sayı zamanın artması ilə sonsuz artacaq, $k < 0$ olarsa, terminlərin sayı zamanın artması ilə sıfıra doğru azalacaq, $k = 0$ olduqda, terminlərin sayı zamanla dəyişilməz qalacaq.

Terminoloji bazada olan terminlərin orta sayı ilə ümumi leksikondakı sözlərin orta sayını müqayisə etməklə dildə gedən prosesləri analiz edib qərar qəbul etmək olar. Bunun üçün xüsusi indikator daxil edilib və təhlil olunub.

TB-də terminlərin orta sayı. TB-nin doldurulma sürəti $T'(t) = \eta(t)$ olsun. t anına qədər orta hesabla TB-də olan terminlərin sayı aşağıdakı kimi təyin olunur:

$$N_T(t) = \frac{1}{t - t_0} \int_{t_0}^t \eta(t) dt \quad (11)$$

Leksikondakı sözlərin orta sayı. Leksikondakı sözlərin sayının dəyişmə sürəti $l(t) = L'(t)$ olsun, t anına qədər dilin ümumi leksikonunda olan sözlərin orta sayı aşağıdakı kimi təyin olunur:

$$N_L(t) = \frac{1}{t - t_0} \int_0^t l(t) dt \quad (12)$$

Aydındır ki, $N_T(t) \ll N_L(t)$. (11) və (12) formullarının nisbəti vasitəsi ilə dildə gedən prosesləri analiz edib qiymətləndirmək olar. Bu məqsədlə $n(t)$ – indikatoru daxil edilib:

$$n(t) = \frac{N_T(t)}{N_L(t)} \quad (13)$$

$N_L(t)$ - monoton artan funksiyadır. Beləliklə, $n(t)$ funksiyasının artması və ya azalması yeni terminlərin dilə daxil olması və bəzi terminlərin müntəzəm olaraq istifadə tezliyinin limit qiymətini aşmaqla ümumi leksikona daxil olması hesabına baş verə bilər. Bunu analiz etməklə yeni terminlərin təhsildə, maariflənmədə, populyarlaşmada geniş istifadə olunub olunmaması, sosial media vasitəsilə de-terminləşməsi və ya əksinə, de-terminləşməməsi haqqında bilik əldə etmək olar və s.

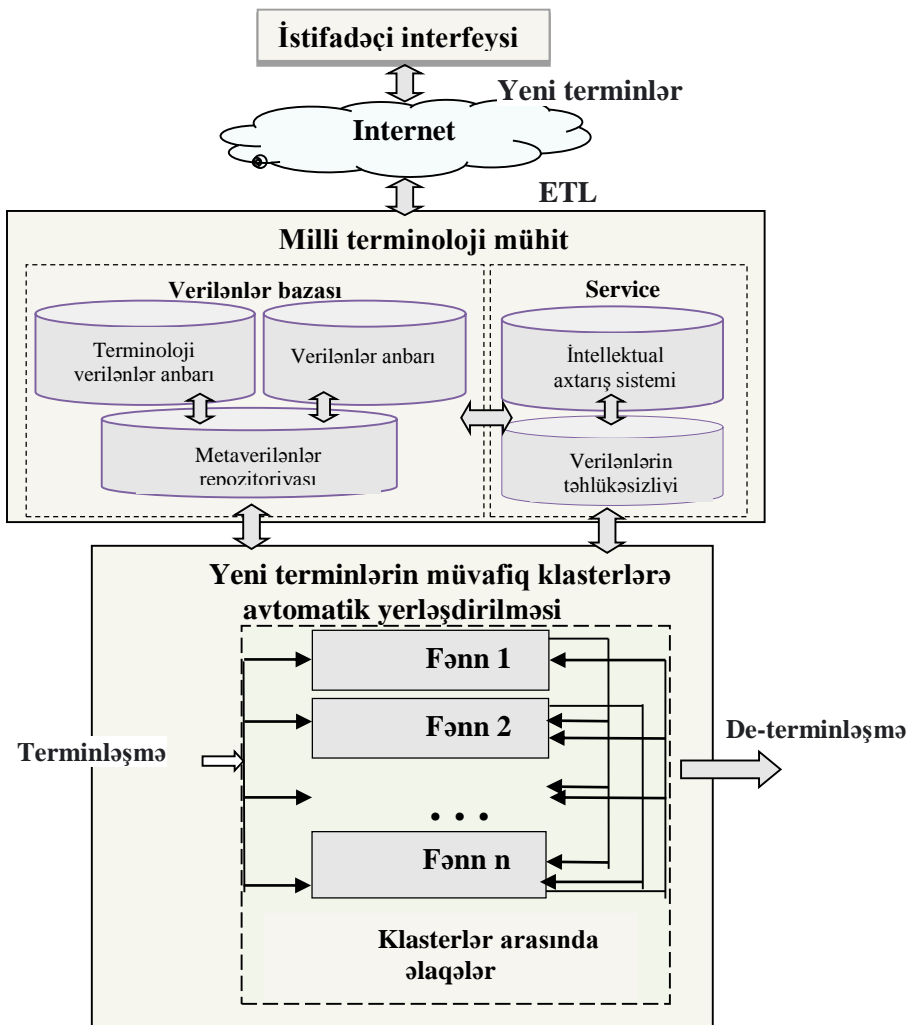
Üçüncü fəslin ikinci bəndində MTİS çərçivəsində dinamik proseslərin ümumi modeli (Şəkil 5), bu proseslərin analizi və qiymətləndirilməsi üçün maşın təliminə əsaslanan metodlar təklif olunmuşdur. Yeni terminlərin uyğun fənlər üzrə paylanmasının avtomatlaşdırılması üçün k-means klasterləşmə alqoritmi və Naive Bayes təsnifatlandırma metodu təklif edilmişdir [20].

Təklif olunan metod: Tutaq ki, tədqiq olunan mühit üçün baxılan mövzu ilə bağlı verilənlər bazası yaradılmışdır. Təklif olunan yanaşmanın hər bir mərhələsi aşağıda ətraflı izah edilmişdir:

Addım 1. Əvvəlcə MTİS-in veb brauzerindən müxtəlif sahələr üzrə mətnlər toplanılır. Bu mətnlərə sənəd kimi baxılır: $D = \{d_1, d_2, \dots, d_n\}$ burada n – sənədlərin sayıdır.

Addım 2. Toplanan sənədlər ilkin emal olunur. Sənədlərdən terminlər çıxarılır və semantik təhlil edilir.

Addım 3. Mətnlərdən terminlər çıxarılır, $T = \{t_1, t_2, \dots, t_m\}$. Daha sonra sənədlər çoxluğu “Term Frequency-Inverse Document (TF-IDF)” sxeminin köməyiylə vektor kimi təsvir olunur.



Şəkil 4. Milli terminoloji sistemdə dinamik proseslərin ümumi modeli

Sənədlər arasındakı məsafəni hesablamaq üçün Evklid məsafəsindən istifadə olunur, $d_i = [w_{i1}, w_{i2}, \dots, w_{im}]$, $i = 1, \dots, n$, burada w_{ij} – j -ci terminin i -ci sənəddə çəkisidir və aşağıdakı kimi hesablanır:

$$w_{ij} = tf_{ij} \log \frac{n}{n_j}; i = 1, \dots, n; \quad (14)$$

burada tf_{ij} i -ci sənəddə j -ci sözün tezliyi, $n_j - t_j$ sözünün rast gəlinədiyi sənədlərin sayı, n – sənədlərin ümumi sayı, m – terminlərin sayıdır.

Addım 4. Kosinus metrikasından istifadə etməklə, iki vektor arasındakı yaxınlıq aşağıdakı kimi hesablanır:

$$sim(d_i, d_l) = \frac{\sum_{j=1}^m w_{ij} w_{lj}}{\sqrt{\sum_{j=1}^m w_{ij}^2 \cdot \sum_{j=1}^m w_{lj}^2}}, \quad i, l = 1, 2, \dots, n \quad (15)$$

burada $d_i = [w_{i1}, w_{i2}, \dots, w_{im}]$ və $d_l = [w_{l1}, w_{l2}, \dots, w_{lm}]$ - d_i və d_l sənədlərinə uyğun vektorlar, w_{ij} və w_{lj} - müqayisə olunan vektorların rastgəlmə tezliklərinin j -ci elementləridir. Burada vektorlar istifadə olunan sözlər üzrə düzülüb, onların uzunluqları üst-üstə düşür və iki vektordan birinin maksimal uzunluğuna bərabərdir.

Addım 5. Sənədlər vektor şəklində təsvir olunduqdan sonra klasterləşdirilir. 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.

Bazaya daxil olan sənədlər $C = \{C_1, C_2, \dots, C_p, \dots, C_k\}$ klasterləri arasında qruplaşdırılmalıdır. Klasterləşmə problemini həll etmək üçün bəzi qeydləri aydınlaşdırmaq lazımdır:

- $x_{ip} = \begin{cases} 1, & \text{if } d_i \in C_p \\ 0, & \text{əks halda} \end{cases}$

$R = [R^1, R^2, \dots, R^g]$ – sənədlərin ümumi mərkəzidir.

burada R -in l -ci koordinatı aşağıdakı kimi müəyyən edilir:

$$R^l = \frac{1}{n} \sum_{i=1}^n w_{ij}; l = 1, 2, \dots, g \quad (16)$$

- $R_p = [R_p^1, R_p^2, \dots, R_p^g]$ – C_p klasterinin mərkəzidir.

R_p -nin l -ci koordinatı aşağıdakı kimi hesablanır:

$$R_p^l = \frac{1}{|C_p|} \sum_{i=1}^n w_{il} x_{ip}; l = 1, \dots, m \quad (17)$$

burada $|C_p| = \sum_{i=1}^n x_{ip}$, $p = 1, \dots, k$, C_p – klasterdəki sənədlərin sayıdır.

Klasterləşmə zamanı klasterlərin mərkəzləri ilə bütün sənəd dəstlərinin mərkəzi arasında yaxınlıq dərəcəsini minimuma endirmək lazımdır.

$$\max f(x) = \sum_{p=1}^k \frac{1}{|C_p|} \sum_{i=1}^n \text{sim}(d_i, R_p) x_{ip} \quad (18)$$

Klasterləşmənin nəticələrini qiymətləndirmək üçün “təmizlik” (purity) əmsalından istifadə olunmuşdur:

$$\text{Purity} = \frac{1}{N} \sum_{k=1}^K \max_j |C_p \cap T_j| \quad (19)$$

burada, N – terminlərin ümumi sayı, K – klasterlərin ümumi sayı, C_p – p -ci klasterə aid terminlər dəsti, T_j – j -ci həqiqi sinfə aid terminlər dəstidir.

Eksperiment üçün k-means klasterləşmə metodundan istifadə edərək müxtəlif təmizlik əmsalı (Purity Coefficient) qiymətlərinə (0,1, 0,2, 0,3, 0,4, 0,5, 0,6 və 0,7) malik olan terminləri qruplaşdırdıq. Nəticə üzrə verilənlərin ümumi həcmi qiymətləndirmək üçün resursların (150, 300, 450, 600, 750 termin) sayını da artırırıq. Cədvəldən görüldüyü kimi, klasterləşdirmənin nəticələri resursların sayından asılıdır. Resursların sayı artdıqca nəticə yaxşılaşır. Təmizliyin artması klasterlər və həqiqi sinflər arasında uyğunluğu artırır.

Addım 6. Növbəti mərhələ yeni terminin artıq müəyyən olunmuş klasterə (fənnə) avtomatik daxil edilməsi üçün Naive Bayes klassifikatorundan istifadə edilib. Naive Bayes böyük həcmli verilənlərin analizi üçün çox uyğun olan sadə və sürətli təsnifat üsuludur. Əsas ideya yeni terminlərin hansı sinfə aid olma ehtimalını qiymətləndirmək üçün onların əlamətlərindən istifadə etməkdir.

Cədvəl. a -nın müxtəlif qiymətlərində klasterləşmənin təmizlik (purity) əmsalı

a	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Resursların sayı	Təmizlik (purity) əmsalı						
150	0.54	0.62	0.68	0.73	0.78	0.82	0.85
300	0.57	0.65	0.71	0.76	0.81	0.84	0.87
450	0.58	0.66	0.72	0.77	0.83	0.86	0.88
600	0.60	0.68	0.73	0.79	0.84	0.87	0.89
750	0.61	0.69	0.75	0.80	0.85	0.88	0.90

Tutaq ki, $C = \{c_1, c_2, \dots, c_m\}$ siniflərdir. Terminlər sinfini w_i kimi qəbul edirik: $\{w_i, i = 1, \dots, m\}$, m -terminlərin sayıdır. w_i terminlərinin $T(x)$ - əlamətlər çoxluğuna aid olması ehtimalını $p(w_i)$ kimi müəyyən edək. Əlamətlər dedikdə sözlərin semantik, qramatik, morfoloji xarakteristikaları nəzərdə tutulur.

Şərt aşağıdakı kimidir:

$$L(w_j, c_i) = \begin{cases} 0, w_j \notin c_i \\ 1, w_j \in c_i \end{cases} \quad (20)$$

Nəzərdə tutulur ki, giriş nümunəsi w_j sinfinə aiddirsə ϕ_i həlli itkiyə yol verir. Şərti risk $x \in w_j$ şərti daxilində itki funksiyası ilə müəyyən edilir:

$$R(\phi_i / x) = \sum_{j=1}^m L(\phi_i / w_j) p(w_j / x) \quad (21)$$

$p(w_j / x)$ ehtimalı ($x \in w_j$) Bayes teoremi vasitəsilə aşağıdakı kimi hesablanır:

$$p(w_j / x) = \frac{p(w_j) p(x / w_j)}{\sum_{k=1}^m p(w_k) p_k(x / w_k)} \quad (22)$$

Beləliklə, Naive Bayes klassifikatorundan istifadə etməklə, biz öz fərdi təsnifatımızı yarada və terminlərin əlamətlərinə görə uyğun fənlər üzrə avtomatik paylanması məsələsini həll edə bilirik. Məsələnin bu cür həlli Big data problemini səmərəli həll etməyə, həm də sistemdə axtarışın keyfiyyətinin yüksəldilməsinə kömək edir.

Üçüncü fəslin üçüncü bəndində milli terminoloji informasiya sistemi çərçivəsində nitqin populyarlıq səviyyəsinin qiymətləndirilməsinin avtomatlaşdırılması məsələsinə baxılmışdır. Bu məqsədlə fərdlərin nitqlərinin populyarlıq səviyyəsinin qiymətləndirilməsi metodu təklif edilmişdir [19].

Nitqin populyarlıq səviyyəsinin qiymətləndirilməsi metodu.

Məqsəd çıxışların, mətnlərin ictimaiyyət və ya hədəf auditoriyası tərəfindən anlama səviyyəsini qiymətləndirməkdir. Bunun üçün statistik tezlik modelinə əsaslanan metod təklif edilmişdir. Uniqramlar və biqramlar üçün nitqin populyarlıq indeksi işlənmişdir.

Aşağıdakı işarələmələri daxil edək:

- $D = \{d_1, \dots, d_n\}$ – fərdin generasiya etdiyi mətnlər (sənədlər) çoxluğu; burada n – sənədlərin ümumi sayıdır.
- $T = \{t_1, \dots, t_m\}$ – D sənədlər çoxluğunda rast gəlinən sözlər çoxluğu; burada m – sözlərin sayıdır;
- $F^{1-gram} = \{f^{1-gram}_1, \dots, f^{1-gram}_m\}$ – dilin tezlik lüğətində sözlərin $(1-gram)$ işlənmə tezlikləri çoxluğu; burada $f_j^{1-gram} t_j$ sözünün işlənmə tezliyidir;
- $s_j = \{t_{(j,1)}, \dots, t_{(j,m_j)}\}$ – t_j sözünün sinonimlər çoxluğu; burada m_j – sinonimlərin sayıdır; $m_j \geq 1$; $j = 1, \dots, m$. Hesab olunur ki, bu çoxluğun 1-ci elementi (sinonim) $t_{(j,1)}$ elə t_j sözünün özüdür.
- $F_j^{1-gram} = \{f_{(j,1)}^{1-gram}, \dots, f_{(j,m_j)}^{1-gram}\}$ – dilin tezlik lüğətində t_j sözünün sinonimlərinin $t_j = \{t_{(j,1)}, \dots, t_{(j,m_j)}\}$ işlənmə tezlikləri çoxluğu; burada $f_{(j,k)}^{1-gram}$ – k -cı sinonimin $t_{(j,k)}$ işlənmə tezliyidir, $k = 1, \dots, m_j$; $j = 1, \dots, m$.

- $F^{2-gram} = \{f_{ij}^{2-gram} | i, j = 1, \dots, m; i \neq j\}$ – dilin tezlik lüğətində bigramların ($2-gram$) işlənmə tezlikləri çoxluğu; burada $f_{ij}^{2-gram}(t_i t_j)$ biqramının işlənmə tezliyidir.

- $F_{ij}^{2-gram} = \left\{ f_{(i,k)j}^{2-gram}, f_{i(j,p)}^{2-gram} \left| \begin{array}{l} k = 1, \dots, m_i; i = 1, \dots, m \\ p = 1, \dots, m_j; j = 1, \dots, m; i \neq j \end{array} \right. \right\};$

- $f_{(i,k)j}^{2-gram}$ - dilin tezlik lüğətində $(t_{(i,k)} t_j)$ biqramının işlənmə tezliyidir.

- $f_{i(j,p)}^{2-gram}$ - dilin tezlik lüğətində $(t_i t_{(j,p)})$ biqramının işlənmə tezliyidir.

- $(t_{(i,k)} t_j)$ - t_i sözünün k -cı sinonimi ilə t_j sözünün əmələ gətirdiyi biqramdır.

- $(t_i t_{(j,p)})$ - t_i sözü ilə t_j sözünün p -ci sinoniminin əmələ gətirdiyi biqramdır.

Nəzərdə tutulur ki, milli terminoloji informasiya sisteminin verilənlər bazasında terminlər, onların sinonimləri və işlənmə tezlikləri haqqında dolğun məlumat vardır.

Uniqramlar üçün nitqin populyarlıq indeksi. Dilin tezlik lüğətində hər bir sözün (sinonimləri nəzərə almaqla) işlənmə tezliyinin ən kiçik və ən böyük qiyməti belə təyin olunur:

$$f_{j_min}^{1-gram} = \min_{k=1, \dots, m_j} \{f_{(j,k)}^{1-gram}\}, j = 1, \dots, m \quad (23)$$

$$f_{j_max}^{1-gram} = \max_{k=1, \dots, m_j} \{f_{(j,k)}^{1-gram}\}, j = 1, \dots, m \quad (24)$$

Əgər sinonimlər çoxluğu s_j yalnız bir elementdən ibarətdirsə, yəni sözün sinonimi yoxdursa, onda $f_{j_min}^{1-gram} = f_{j_max}^{1-gram}$.

(23) və (24)-dən alınır ki:

$$f_j^{1-gram} \in [f_{j_min}^{1-gram}, f_{j_max}^{1-gram}], j = 1, \dots, m \quad (25)$$

(23)-(25)-də bütün sözlər üzrə cəmləmə aparıb və orta qiyməti hesablasaq, D çoxluğu üçün sözlərin işlənmə tezliyinin orta qiymətini alarıq:

$$f_{\min}^{1-gram} = \frac{1}{m} \sum_{j=1}^m f_{j_{\min}}^{1-gram} \quad (26)$$

$$f_{\max}^{1-gram} = \frac{1}{m} \sum_{j=1}^m f_{j_{\max}}^{1-gram} \quad (27)$$

$$f^{1-gram} = \frac{1}{m} \sum_{j=1}^m f_j^{1-gram} \quad (28)$$

(25)-(28) düsturlarından istifadə etməklə, nitqin populyarlıq indeksi belə təyin edilir:

$$p^{1-gram} = \frac{f^{1-gram} - f_{\min}^{1-gram}}{f_{\max}^{1-gram} - f_{\min}^{1-gram}} \quad (29)$$

(29)-dən alınır ki, $p^{1-gram} \in [0,1]$. Tərifdən görünür ki, p -nin qiyməti 0-a yaxınlaşdıqca nitqin populyarlıq səviyyəsi aşağı və əksinə, p -nin qiyməti 1-ə yaxınlaşdıqca nitqin populyarlıq səviyyəsi yüksək olacaq. Beləliklə, sinonimləri seçməklə, nitqin populyarlıq səviyyəsi idarə edilə bilər. Qeyd etmək lazımdır ki, söz birləşmələrinin (iki-iki, üç-üç, dörd-dörd və s.) tezliyinə də baxmaq olar. Bunu ən sadə model olan N -qramlardan (bir söz, iki söz, üç söz və s.) istifadə etməklə, eyni zamanda tezlik lüğəti əsasında etmək olar.

Biqramlar üçün nitqin populyarlıq indeksi. Dilin tezlik lüğətində hər bir (t_i, t_j) biqramının (sinonimləri nəzərə almaqla) işlənmə tezliyinin ən kiçik və ən böyük qiyməti belə təyin olunur:

$$f_{i_{\min}, j}^{2-gram} = \min_{k=1, \dots, m_i} \{f_{(i,k)j}^{2-gram}\}, i, j = 1, \dots, m, i \neq j \quad (30)$$

$$f_{i_{\max}, j}^{2-gram} = \max_{k=1, \dots, m_i} \{f_{(i,k)j}^{2-gram}\} \quad (31)$$

$$f_{i, j_{\min}}^{2-gram} = \min_{p=1, \dots, m_j} \{f_{i(j,p)}^{2-gram}\} \quad (32)$$

$$f_{i, j_{\max}}^{2-gram} = \max_{p=1, \dots, m_j} \{f_{i(j,p)}^{2-gram}\} \quad (33)$$

(30)-(33)-dən alınır ki,

$$f_{ij_min}^{2-gram} = \min \{ f_{i_min, j}^{2-gram}, f_{i, j_min}^{2-gram} \} \quad (34)$$

$$f_{ij_max}^{2-gram} = \max \{ f_{i_max, j}^{2-gram}, f_{i, j_max}^{2-gram} \} \quad (35)$$

(34) və (35)-dən alınır ki,

$$f_{ij}^{2-gram} \in [f_{ij_min}^{2-gram}, f_{ij_max}^{2-gram}] \quad (36)$$

(30)-(36)-də bütün sözlər üzrə cəmləmə aparıb və orta qiyməti hesablasaq, D çoxluğu üçün biqramların işlənmə tezliyinin orta qiymətini alırıq:

$$f_{min}^{2-gram} = \frac{1}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m f_{ij_min}^{2-gram} \quad (37)$$

$$f_{max}^{2-gram} = \frac{1}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m f_{ij_max}^{2-gram} \quad (38)$$

$$f^{2-gram} = \frac{1}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m f_{ij}^{2-gram} \quad (39)$$

(36)-(39) düsturlarından istifadə etməklə, bigramlar üçün nitqin populyarlıq indeksi belə təyin edilir:

$$p^{2-gram} = \frac{f^{2-gram} - f_{min}^{2-gram}}{f_{max}^{2-gram} - f_{min}^{2-gram}} \quad (40)$$

Fərdə aid yazılı və şifahi nitq materiallarını zamana görə dövrləşdirib ayrılıqda qiymətləndirmək və sonra dinamikasına baxmaq olar və s.

Populyarlıq indeksinin zamandan asılılıq dinamikası. Nitq materiallarının populyarlıq səviyyəsinin zamandan asılılığı aşağıdakı kimi təyin olunur:

$$\begin{cases} \frac{dp(t)}{dt} = ap(t) \\ p(0) = p_0 \end{cases} \quad (41)$$

Buradan alınır ki,

$$p(t) = p_0 e^{at} \quad (42)$$

burada a – nitq materiallarının populyarlıq səviyyəsinin artma sürəti, e - natural loqarifmin əsasını təşkil edir. Beləliklə, fərdin nitqinin populyarlıq səviyyəsinin zamandan asılılığı eksponensial funksiyadır. Göründüyü kimi, əgər $a > 0$ olarsa, yəni fərdə aid nitq materiallarının populyarlıq səviyyəsi zamanla yüksəlsə, deməli fərd dili yaxşı öyrənir, əgər $a < 0$ olarsa, yəni fərdə aid nitq materiallarının populyarlıq səviyyəsi zamanla 0-a yaxınlaşarsa, deməli fərd dil biliyini itirir, əgər $a = 0$ olarsa, yəni fərdə aid nitq materiallarının populyarlıq səviyyəsi zamanla dəyişməz qalır, deməli fərdin dil biliyi sabit qalır.

Nitqin populyarlıq səviyyəsi qiymətləndirilərkən, onun həm ümumi, həm də elmi populyarlıq səviyyəsinin qiymətləndirilməsi və onların müqayisə edilməsi zəruridir. Belə ki, bu qiymətləndirmə nəticəsində fərdin dil və elmi bilik səviyyəsini müəyyənləşdirmək və qərar qəbul etmək olar. Qiymətləndirmə $p \in [0,1]$ intervalında aparılır: ümumi populyarlıq səviyyəsi yaxşı olduqda 1, zəif olduqda 0 ilə qiymətləndirilir. Nitqin populyarlıq səviyyəsinin dəyəri müxtəlif ola bilər. Belə ki, nitqin ümumi və elmi populyarlıq səviyyəsi eyni zamanda yaxşı və ya zəif ola bilər. Elə hallar ola bilər ki, nitqin ümumi populyarlıq səviyyəsi yaxşı, lakin elmi populyarlıq səviyyəsi zəif olsun və əksinə.

Nitqin ümumi populyarlıq indeksinin qiyməti (40) formuluna analoji olaraq aşağıdakı kimi təyin olunur:

$$p_L = \frac{f_L - f_L^{\min}}{f_L^{\max} - f_L^{\min}}, p_L \in [0,1] \quad (43)$$

Nitqin elmi populyarlıq indeksinin qiyməti analoji olaraq aşağıdakı kimi təyin olunur:

$$p_T = \frac{f_T - f_T^{\min}}{f_T^{\max} - f_T^{\min}}, p_T \in [0,1] \quad (44)$$

Nitqin ümumi və elmi populyarlıq indekslərini vaciblik əmsalına görə cəmləyək:

$$p = \alpha_L p_L + \alpha_T p_T \quad (45)$$

α – vaciblik əmsalıdır. Beləliklə, fərdlərin nitqinin (yazısının) populyarlıq indeksini müntəzəm olaraq analiz etməklə fərdin dil və elmi bilik səviyyəsini müəyyənləşdirmək və qərar qəbul etmək olar.

Dördüncü fəsilə (“MTİS-in yaradılması imkanları və perspektivləri”) MTİS-in analizində istifadə olunan verilənlərin səmərəli toplanması və təşkili üçün təkliflər verilmiş, dissertasiyada işlənən metod və alqoritmlərin eksperimental yoxlanması həyata keçirilmişdir. MTİS-in reallaşdırılmasından gözlənilən nəticələr verilmiş və sistemin perspektivləri istiqamətində tövsiyələr işlənmişdir.

Dördüncü fəslin birinci bəndində MTİS-də QQDS-in verilənlər saxlanmasına müvafiq proqram təminatı modulları əsasında sorğulara uyğun hesabatlar formalaşdırılmışdır (21,22). OLAP Microsoft Exel-də PivotTable Servise qurulması vasitəsilə vizuallaşdırılmışdır. Analiz üçün VS-yə müxtəlif mənbələrdən daxil edilən 50 termin və həmin terminlər haqqında məlumatlar (terminin seçilən müddəti, seçilən dili, seçilən sahəsi və müəllifləri) üzərində PivotTable qurulub. Analiz üçün seçilən verilənlər relyasiyadır, yəni ayrı-ayrı cədvəllərdə saxlanırsalar da ümumi dəyərlərə görə birləşdirilə bilən verilənlərdir. Müxtəlif kəsimlərdə aparılan sorğular nəticəsində hesabatlar alınmışdır (Şəkil 5-7).

Beləliklə, tədqiqat göstərdi ki, terminlər və onların əlamətlər vektorunu analiz etməklə milli terminoloji informasiya mühitində gedən prosesləri izləmək və idarə etmək mümkündür. Belə ki, bir il ərzində dilimizə nə qədər terminin keçdiyini, əsasən hansı sahədən, hansı dildən daha çox terminlərin əlavə olunduğunu və s. təyin etməklə, dilin ekologiyası, təhlükəsizliyi haqqında, ölkədə hansı fəaliyyət sahəsinin inkişaf etdiyi haqqında əldə etdiyimiz məlumatlar əsasında qərar qəbul etmək imkanı əldə edirik. Analiz üç əlamət üzrə aparılıb (terminin dili, fəaliyyət sahəsi və yaranma tarixi). Təklif edilən polikubik OLAP-model əsasında əlamətlər vektorunu genişləndirməklə daha çox parametrlər üzrə analiz aparmaq və qərar qəbul etmək imkanı vardır.

	A	B	C	D
1	Field of act	computer science		
2				
3	Количество по полю Author	Названия столбцов		
4		Deutsch	English	Общий итог
5	Названия строк			
6	Algorithm		1	1
7	Anisotropic filtering			
8	AntiCam			
9	Authentication			
10	BASIC		1	1
11	Computer			
12	Differential evolution		1	1
13	Digital electronic computer	1		1
14	Java		1	1
15	Kabel			
16	Modem		1	1
17	Semantic web		1	1
18	World Wide Web		1	1
19	Общий итог	2	6	8

Şəkil 5. Bütün dillər üçün terminlərin sayının kompyuter elmləri üzrə paylanması

1	Language	(All)								
2										
3	Quantity across the field Date	Column names								
4	Row names	biochemistry	chemistry	computer science	ethics	history	economy	mathematics	pedagogics	physics
5	Abscissa							1		
6	Algorithm			1						
7	Anisotropic filtering			1						
8	AntiCam			1						
9	Aqua-Lung									1
10	Authentication			1						
11	Aviation									1
12	Barometer	1								
13	BASIC			1						
14	Beurs						1			
15	Bunsen-Rosco Law		1							
16	Capitalism					1				
17	Catalysis	1								
18	Catharsis									
19	Computer			1						
20	Deontology				1					
21	Differential evolution			1						
22	Diffusion	1								
23	Digital electronic computer			1						
24	Ellipse							1		
25	Exponent							1		
26	Golden ratio							1		
27	Imperium					1				
28	Integral							1		
29	Isomerism	1								
30	lava			1						

Şəkil 6. Müxtəlif fəaliyyət sahələrində terminlərin sayının bütün dillər üzrə paylanması

Quantity across the field	Language	Column names											chemistry total	computer science	ethics
Row names			1830	1833	1835	1858	1860	1862	1928	1930	1951				
Abscissa															
Algorithm														1	
Anisotropic filtering														1	
AntiCam														1	
Aqua-Lung															
Authentication													1		
Aviation															
Barometer					1										
BASIC														1	
Beurs															
Bunsen-Rosco Law									1				1		
Capitalism															
Catalysis						1							1		
Catharsis															
Computer														1	
Deontology															
Differential evolution													1	1	
Diffusion					1								1		
Digital electronic computer														1	
Ellipse															
Exponent															
Golden ratio															
Imperium															
Integral															
Isomerism													1		

Şəkil 7. Bütün fəaliyyət sahələrinə və tarixə görə terminlərin sayının bütün dillər üzrə paylanması

Dördüncü fəslin ikinci bəndində MTİS çərçivəsində yaradılan milli terminoloji veb-portalın funksiyaları və imkanları təhlil olunmuş (4), MTİS-in reallaşdırılması imkanları və perspektivləri müəyyənləşdirilmişdir:

Milli terminoloji veb-portal. Milli terminoloji veb-portalın (www.terminologiya.az) proqram təminatı: Adobe Photoshop CS6, Adobe Illustrator CC17, HTML5, CSS3, JavaScript, jQuery, AJAX, PHP5.6, MySQL. Veb-portalın keyfiyyət göstəriciləri *Webometrics sistemi*, *LiveInternet veb-portalı*, *Google-analytics sistemi* vasitəsilə qiymətləndirilir. Veb-portalın informasiya-axtarış sistemi vasitəsilə termin, onun izahı, digər dillərə tərcüməsi, onun mənbəsi və s. kimi məlumatların əldə edilməsi termin yaradıcılığı işinin keyfiyyətini yüksəldir. “İctimai tribuna” bölməsi vasitəsilə vətəndaşlarla onlayn əlaqələr qurulur. MTİS və onun veb-portalının işlənməsi terminlərin təhlilinə yönəlmiş xüsusi biliklər bazasının yaradılmasına və

Azərbaycanda termin yaradıcılığının səmərəliliyinin yüksəldilməsinə geniş imkanlar vardır.

MTİS-in reallaşdırılması imkanları və perspektivləri. Milli terminoloji informasiya sisteminin sintezi və intellektuallaşdırılması istiqamətində aparılan tədqidatlardan əldə olunan nəticələr sistemin reallaşdırılmasının imkan və perspektivlərini müəyyənləşdirdi:

- İstifadəçilərə e-terminoloji xidmətlər göstərəcəkdir;
- Termin yaradıcılığı prosesinə sərbəst olaraq insanların qoşulması, irad və təkliflərini bildirməsi, hazırlanmış terminoloji lüğətlərin müzakirəsinə həsr olunmuş onlayn forumlarda iştirak etməsi vətəndaş terminologiyasının formalaşmasına təkan verəcəkdir;
- Müxtəlif sahələr üzrə hazırlanmış terminoloji lüğətlərin vahid informasiya sistemində toplanması nəticəsində termin yaradıcılığının səmərəliliyi yüksələcəkdir (terminlərin harmonikləşdirilməsi və s.);
- Terminoloji lüğətləri hazırlamış alim və mütəxəssislərin müəlliflik hüquqlarının qorunması üçün şəffaflıq və asan əlyətərlik nəticəsində geniş imkanlar yaranacaqdır;
- Terminoloji lüğətlərə, ayrı-ayrı terminlərə dair əks-əlaqə məlumatlarını emal etməklə müxtəlif məzmunlu təhlillər və qiymətləndirmələr aparmaq, reytinglər təşkil etmək mümkün olacaqdır;
- Milli terminoloji informasiya sisteminin beynəlxalq və xarici ölkələrin analoji sistemləri ilə inteqrasiyası sürətlənəcəkdir;
- Azərbaycan dilinin terminoloji bazasının formalaşması və inkişafına böyük imkan yaradacaqdır;
- Terminologiya sahəsində aparılan tədqiqatlar üçün informasiya bazası rolunu oynayacaqdır;
- Ölkəmizdə terminoloji fəaliyyətin beynəlxalq və milli standartlar əsasında həyata keçirilməsinə dəstək olacaqdır və s.

DİSSERTASIYA İŞİNİN ƏSAS NƏTİCƏLƏRİ

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

1. Terminoloji informatikanın elmi-texnoloji problemləri analiz olunmuş, əsas inkişaf istiqamətləri müəyyən olunmuş və bu istiqamətlərdə elmi araşdırmaların aparılmasının vacibliyi əsaslandırılmışdır [1, 2, 4, 5, 6, 12, 16,18];

2. MTİS-in konseptual modeli və arxitektur-texnoloji prinsipləri işlənmiş, funksional alt sistemləri təhlil edilmiş, imkan və perspektivləri müəyyənləşdirilmişdir [3,7,10];

3. Milli terminoloji informasiya sisteminin qiymətləndirilməsi üçün OLAP texnologiyalarına əsaslanan qərarların qəbulunu dəstəkləyən sistem və polikubik OLAP-model işlənmişdir [13,14];

4. MTİS çərçivəsində termin yaradıcılığı proseslərinin avtomatlaşdırılması üçün statistik analizə əsaslanan metod işlənmişdir [8, 9, 11, 15].

5. MTİS çərçivəsində dinamik proseslərin modelləşdirilməsi üçün populyasiya artım modelinə əsaslanan metod işlənmişdir [17];

6. Terminoloji dinamik proseslərin intellektual analizi üçün səlisləşmə alqoritmi və Naive Bayes təsnifatlandırma modelinə əsaslanan metod işlənmişdir [20];

7. MTİS əsasında fərdlərin nitqlərinin populyarlıq səviyyəsinin qiymətləndirilməsi üçün statistik tezlik modelinə əsaslanan metod işlənmişdir [19].

Dissertasiya işinin əsas nəticələri aşağıdakı elmi işlərdə dərc edilmişdir:

1. Əliquliyev R.M., Qurbanova Ə.M. Virtual məkanda Azərbaycan dilinin problemləri və həlli yolları // Terminologiya məsələləri, - Bakı: - 2009, - s.11-18
2. Əliquliyev R.M., Qurbanova Ə.M. Terminoloji informasiya sistemi və inkişaf perspektivləri // Terminologiya məsələləri, - Bakı, - 2010, - s. 27-33.
3. Gurbanova A. Conceptual essences of creation of terminological information system in Azerbaijan // The Third International

- Conference Problems of Cybernetics and Informatics (PCI), vol.1, - 2010, - Baku, - pp. 157-159.
4. Qurbanova Ə.M. Terminoloji informatika sahəsində aparılan elmi-texnoloji araşdırmalar və onların müqayisəli təhlili // İnformasiya cəmiyyəti problemləri, - Bakı, - 2013, №2(8), - s. 3-12.
 5. Qurbanova Ə.M. Qloballaşma şəraitində Azərbaycan dilinə terminoloji təhlükələr haqqında // İnformasiya cəmiyyəti problemləri, - Bakı, - 2015, №2, - s. 87-95.
 6. Qurbanova Ə.M. Terminoloji informatika: formalaşma mərhələləri və inkişaf tendensiyaları // İnformasiya cəmiyyəti problemləri, - Bakı, - 2016, №1, - s. 71-77
 7. Qurbanova Ə.M. Verilmiş predmet sahəsində terminoloji informasiya sisteminin işlənməsi // İnformasiya cəmiyyəti problemləri, Bakı, - 2016, №2, - s.74–79.
 8. Qurbanova Ə.M. Predmet sahəsinin terminlərinin semantik şəbəkəsinin avtomatik qurulması texnologiyaları // İnformasiya texnologiyaları problemləri, - Bakı, - 2018, №1, - s. 84–91.
 9. Qurbanova Ə.M. Termin yaradıcılığı işlərinin avtomatlaşdırılması metodlarının təhlili // İnformasiya texnologiyaları problemləri, - Bakı, - 2018, №2, - s. 92-100.
 10. Alguliyev R., Gurbanova A. The Conceptual Foundations of National Terminological Information System // International Journal of Education and Management Engineering, MECS Press, - 2018, 4, - p. 19-30.
 11. Гурбанова А.М. Преимущества автоматизации терминографической работы // «TurkLang» VI International Conference on Computer Processing of Turkic Languages, - Tashkent (Uzbekistan), - 2018, - с. 265-272.
 12. Alguliyev R., Yusifov F., Gurbanova A. Protection of Azerbaijani language in e-government and development prospects // «TurkLang» VI International Conference on Computer Processing of Turkic Languages, - Tashkent (Uzbekistan), - 18-20 October 2018, - p. 282-288.
 13. Набибекова Г.Ч., Гурбанова А.М. Система поддержки принятия решений в среде национальной терминологической информации // Международная конференция «Развитие

- информатизации и государственной системы научно-технической информации», - Минск, - 2019, - с. 41-45
14. Alguliyev R., Nabibayova G., Gurbanova A. Development of a Decision Support System with the use of OLAP-Technologies in the National Terminological Information Environment // International Journal of Modern Education and Computer Science, Vol.11 (6), - 2019, - p. 43-52. (Scopus)
 15. Qurbanova Ə.M. Mətnlərdən terminlərin avtomatik çıxarılması metodlarının araşdırılması və müqayisəli analizi // İnformasiya texnologiyaları problemləri, - Bakı, - 2021, №1, - s. 55-69.
 16. Alguliyev R., Yusifov F., Gurbanova A. Protection of Azerbaijani Language in e-government platform // e-Journal of Linguistics, Vol. 15, No.2, - 2021, - p. 155-161.
 17. Алигулиев Р.М., Гурбанова А.М. Математическое моделирование динамических процессов в рамках терминологической информационной системы // Информационные технологии, 2022, № 5, с. 226-233.
 18. Gurbanova A.M. Problems and Prospects for Minority Languages in the Age of Industry 4.0. // The 4th International Symposium on Computer Science, Digital Economy and Intelligent Systems (CSDEIS2022), The Springer serie Lecture Notes on Data Engineering and Communications Technologies, - 2023, - Wuhan (China), volume 158, pp. 722-734. (Scopus)
 19. Alguliyev R., Aliguliyev R., Gurbanova A. A method of evaluating the popularity of individuals' speeches based on the National Terminological Information System // IEEE 17th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan, - 2023, - pp. 29-33, (Scopus)
 20. Gurbanova A., Alakbarova I., Mammadzade S. Development of a method for intellectual analysis of terminological dynamic processes // Proceedings of the 2nd Information Technologies and Its Applications Conference, Part 3, – Baku, Azerbaijan: – 23-25 april 2024, – pp. 1–15.

Həmmüəlliflərlə dərc olunmuş işlərdə iddiaçının şəxsi rolu:

[1] Virtual məkanda Azərbaycan dilinin vəziyyəti, istifadə səviyyəsi və problemləri üzrə aparılan araşdırmaların təhlili və problemlərin həlli istiqamətində yeni yanaşmanın təklif edilməsi.

[2] Azərbaycanda terminologiya sahəsində mövcud vəziyyət və problemlərin təhlil olunması, bu istiqamətdə İKT-nin tətbiqi imkanlarının müəyyənləşdirilməsi, terminologiya fəaliyyətinə yeni yanaşmanın təklif edilməsi.

[10] Milli terminoloji informasiya sisteminin konseptual modelinin işlənməsi, sistemin yaratdığı imkan və perspektivlərin müəyyənləşdirilməsi.

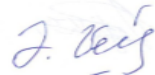
[12, 16] Qlobal informasiya məkanının e-Azərbaycan seqmentində Azərbaycan dilinin qorunması və inkişafı istiqamətində linqvistik texnologiyaların tətbiqi imkanlarının müəyyənləşdirilməsi, e-dövlət mühitində Azərbaycan dili ekosisteminin konseptual modelinin işlənməsi, ekosistemin funksional alt sistemlərinin müəyyənləşməsi.

[13, 14] Tədqiqat məsələsinin qoyuluşunda, MTİS-də OLAP əsasında Qərarların qəbulunu dəstəkləyən sistemin işlənməsi və sistemin verilənlərinin analiz olunması.

[17] MTİS çərçivəsində dinamik proseslərin riyazi modelləşdirilməsi üzrə aparılan işlərin təhlili və yeni yanaşmanın təklif edilməsi.

[19] MTİS əsasında nitqin populyarlıq səviyyəsinin qiymətləndirilməsinin avtomatlaşdırılması üzrə aparılan işlərin araşdırılması və yeni yanaşmanın təklif edilməsi.

[20] MTİS çərçivəsində dinamik proseslərin analizi və qiymətləndirilməsinin avtomatlaşdırılması üzrə aparılan işlərin təhlili və yeni yanaşmanın təklif edilməsi.



Dissertasiyanın müdafiəsi 7 fevral 2025-ci il tarixində saat 16⁰⁰-da Azərbaycan Respublikası Elm və Təhsil Nazirliyinin İnformasiya Texnologiyaları İnstitutunun nəzdində fəaliyyət göstərən ED 1.35 Dissertasiya Şurasının iclasında keçiriləcək.

Ünvan: Az 1141, Bakı şəhəri, B.Vahabzadə küçəsi, 9A.

Dissertasiya ilə Azərbaycan Respublikası Elm və Təhsil Nazirliyinin İnformasiya Texnologiyaları İnstitutunun kitabxanasında tanış olmaq mümkündür.

Dissertasiya və avtoreferatın elektron versiyaları Azərbaycan Respublikası Elm və Təhsil Nazirliyinin İnformasiya Texnologiyaları İnstitutunun rəsmi internet saytında (www.ict.az) yerləşdirilmişdir.

Avtoreferat 28 dekabr 2024-cü il tarixində zəruri ünvanlara göndərilmişdir.

Çapa imzalanıb: 26.12.2024

Kağızın formatı: 60 × 84 ^{1/16}

Həcm: 36400 işarə

Tiraj: 100 nüsxə