

AZƏRBAYCAN RESPUBLİKASI

Əlyazması hüququnda

AZƏRBAYCAN DİLİNDƏKİ MƏTNLƏRİN MÜƏLLİFLİYİNİ İDENTİFİKASIYA EDƏN KOMPÜTER SİSTEMİ ÜÇÜN ALQORİTM VƏ PROQRAM TƏMİNATININ İŞLƏNİLMƏSİ

İxtisas: 1203.01- Kompüter elmləri

Elm sahəsi: Texnika

İddiaçı: **Rüstəm Bakir oğlu Əzimov**

Fəlsəfə doktoru elmi dərəcəsi almaq üçün təqdim edilmiş
dissertasiyanın

AVTOREFERATI

Bakı – 2024

Dissertasiya işi Azərbaycan Respublikası Elm və Təhsil Nazirliyinin İdarəetmə Sistemləri İnstitutunda yerinə yetirilmişdir.

Elmi rəhbər: AMEA-nın müxbir üzvü, r.e.d. professor
Kamil Rəcəb oğlu Ayda-zadə

Rəsmi opponentlər: texnika elmləri doktoru, professor
Bayram Qənimət oğlu İbrahimov

texnika elmləri doktoru, dosent
Mais Paşa oğlu Fərhadov

texnika üzrə fəlsəfə doktoru, dosent
Şəhla Surxay qızı Hüseynzadə

Azərbaycan Respublikasının Prezidenti yanında Ali Attestasiya Komissiyasının AR ETN İdarəetmə Sistemləri İnstitutunun nəzdində fəaliyyət göstərən ED 1.20 Dissertasiya şurası.



Dissertasiya şurasının
sədri:

AMEA-nın həqiqi üzvü, texnika elmləri
doktoru, professor
Əli Məhəmməd oğlu Abbasov

Dissertasiya şurasının
elmi katibi:

texnika üzrə fəlsəfə doktoru, dosent
Tahir Əli oğlu Əlizadə

Elmi seminarın
sədri:

texnika elmləri doktoru, dosent
Rövşən Ağakışi oğlu Quliyev

İŞİN ÜMUMİ XARAKTERİSTİKASI

Mövzunun aktuallığı və işlənmə dərəcəsi. Təbii dildə yazılmış mətnlərin emalına olan maraq müxtəlif məsələlərin aktuallaşmasına gətirib çıxarmışdır. Bu sahədə aktual istiqamətlərdən biri də mətnlərin kateqoriyalara ayrılmasıdır. Mətnlər mövzuya, sentimentallığa və yazı tərzinə görə kateqoriyalara bölünə bilər. Yazı tərzlərinə görə mətnlərin kateqoriyalara bölünməsi sənəd janrlarının identifikasiyası və mətnlərin müəlliflərinin tanınması üçün geniş istifadə olunur. Belə alınır ki, mətnlərin müəlliflərinin tanınması və ya mətnlərin müəlliflik identifikasiyası mətnlərin müəlliflərin yazı stilistikasına əsasən kateqoriyalara ayrılmasıdır.

Mətnlərin müəlliflərinin tanınması müəlliflərin mətnlərdəki yazı tərzlərinin – yazı stilistikalarının kəmiyyət analizi əsasında aparılır. Mətnin yazı tərzinin kəmiyyət analizinə, yəni stilometriyaya həsr olunmuş ilk işlərdən biri T.C. Mendenhol aiddir. T.C. Mendenhol 19-cu əsrin sonunda çap olunan bir işində müxtəlif uzunluqlu sözlərin saylarının (bir hərfli sözlərin sayı, iki hərfli sözlərin sayı və s.) əsərin ayrı-ayrı hissələrində invariant qaldığını göstərmişdir. T.C. Mendenhol 1901-ci ildə çap olunan işində bu xarakteristikalar üzrə Şekspirin və Bekonun əsərlərini göstəricilərini hansısa konkret riyazi prosedur istifadə etmədən öz manual analizi əsasında müqayisə etmişdir. A.A. Markov 1913-ci ildə çap olunmuş işində Puşkinin “Yevgeni Onegin” əsərinin mətni üzərində sait-samit sayları əsasında statistik analiz aparmışdır. Sonradan, 20-ci əsrin 60-cı illərində Mosteller və Ualles ilk Nyu York konstitusiyası barədə 18-ci əsrdə Nyu York qəzetlərinin nəşrlərindəki 77 essedən müəllifi şübhəli 12-sinin müəlliflik identifikasiyası üçün statistik analiz üsullarından istifadə etmişdir. Mətnlərin müəlliflərinin tanınması məsələlərinin tədqiqi əsasları T.C. Mendenhol, F. Mosteller, D.L. Wallace, E. Stamatatos, P. Juola, J. Deiderich, Y. Zhao, V. Keselj, M. Kestemont, A.V. Anisimov, K.R. Ayda-zadə, S.G. Talibov və s. alimlərin işlərində qoyulmuş və inkişaf etdirilmişdir.

Mətnlərin kateqoriyaya ayrılması istiqamətindəki digər məsələ siniflərində olduğu kimi mətnlərin müəlliflərinin tanınması məsələsi də obrazların tanınması (və ya surətlərin tanınması, ingilis dilində

“pattern recognition”, rus dilində “распознавание образов” adlandırılır) məsələsinə gətirilə bilər. Obrazların tanınması sahəsində respublikamızda xeyli iş görülmüşdür. 80-ci illərdə akademik T.A. Əliyevin rəhbərliyi ilə G.G. Abdullayeva və N.H. Qurbanova tərəfindən elektron qələm vasitəsilə ərəb əlifbasında yazılmış yazını tanıma sistemi yaradılmışdır 90-cı illərdə AMEA-nın müxbir üzvü R.A. Əliyevin rəhbərliyi ilə qeyri-səlis təsvirlərin tanınması sahəsində tədqiqatlar aparılmışdır. Prof. Ə.K. Kərimovun rəhbərliyi ilə R.İ. Əsədova (Davudova) tərəfindən dinamik sistemlərin optimal klassifikatorlarının yaradılması üçün işlər görülmüş, kompleks dinamik sistemlərin vəziyyətinin tanınması və klaster analizlər üzərində tədqiqatlar aparılmışdır. Prof. R.Q. Məmmədovun rəhbərliyi ilə A.S. Mütəllibova tərəfindən texniki görmə sistemlərinə baxılmışdır. Prof. O.Q. Nüsrətovun rəhbərliyi ilə əllə yazılmış çap hərflərini tanıma sistemi işlənmişdir. AMEA-nın müxbir üzvü K.R. Ayda-zadənin rəhbərliyi ilə E.E. Mustafayev tərəfindən çap əlyazma hərflərinin tanınması, C.Z. Həsənov tərəfindən əlyazmaların tanınması və S.S. Rüstəmov tərəfindən nitqi tanıma sistemləri işlənmişdir G.G. Abdullayevanın rəhbərliyi altında dəm qazından zəhərlənmələrin tanınması, ortopediya və onkologiya xəstələrinin avtomatlaşdırılmış diaqnostikası, xalçaların naxış və ornamentlərinin tanınması məsələləri tədqiq olunmuşdur. Prof. Ə.N. Nəsimovun rəhbərliyi altında sosial şəbəkələrdəki türk dilindəki mətnlərdə kiber zorakılıq hallarının təyini üzrə işlər aparılmışdır. M.P. Fərhadovun rəhbərliyi altında nitq tanıma əsasında mətndəki açar sözlərin aşkarlanması sistemi işlənmişdir. AMEA-nın müxbir üzvü R.M. Aliquliyevin rəhbərliyi altında mətnlərdəki plagiat hallarının təyini, mətnlərin avtomatik xülasələşdirilməsi üzrə elmi tədqiqatlar aparılmışdır.

Mətnlərə müəllifliyin tanınması istiqamətində görülmüş işlər istifadə olunan mətn əlaməti (müəlliflərinin mətnlərdəki stilistik göstəriciləri) növləri, mətn janrı və həcm barədə göstəriciləri, dil, istifadə olunan müəlliflik identifikasiyası yanaşması və sairəyə görə fərqlənir. Mətnlərə müəlliflərinin identifikasiyası barədə aparılmış elmi tədqiqatlar arasında qəzet guşələrindəki və digər qısa mətnlərin müəlliflərinin tanınması və bədii əsərlərin müəlliflərinin tanınması ilə əlaqəli bir sıra işlərə rast gəlinə bilər.

Azərbaycan dilindəki mətnlərin müəlliflərinin tanınması üzrə də işlər görülmüşdür. Məsələn, K.R. Ayda-zadənin və S.G. Talıbovun işlərindən biri Azərbaycan dilindəki publisistik üslubdakı kiçik həcmli qəzet məqalələri mətnlərinin müəllifliyinin tanınmasında simvol n-qramları (simvol n-qramı dedikdə hər hansı n simvolun kombinasiyası nəzərdə tutulur) statistik üsullar və dayaq vektorlar üsulu ilə istifadə olunmuşdur.

Mövcud müəllif tanıma kompüter sistemlərinin heç biri azərbaycan dilindəki mətnlərin müəlliflərini tanımaq üçün nəzərdə tutulmamışdır. Dissertasiya işində azərbaycan dilindəki mətnlərin xüsusiyyətləri dilin leksikonu əsasında əhatəli tədqiq olunmuşdur, mətnlərin stilistik əlamətləri araşdırılmış və azərbaycan dilindəki mətnlərin müəlliflərini tanıma kompüter sistemi işlənilmişdir

Dissertasiya işində azərbaycan dilindəki mətnlərin müəlliflərini tanıma sisteminin qurulması prinsipləri işlənilmiş, mətdəki müəllif stilistikasını ifadə edən müxtəlif növ əlamətlər, tanımada effektiv mətn əlamətlərinin seçilməsi prosedurları və müxtəlif məşin öyrənməsi üsullarının müəllif tanımada effektivliyi baxılan müəllif tanıma məsələsi təmsalında tədqiq edilmişdir.

Tədqiqatın obyektı və predmeti. Dissertasiya işinin tədqiqat obyektı azərbaycan dilindəki mətnlərin müəlliflərinin tanınması məsələləridir. Tədqiqatın predmeti isə müəllifi məlum mətnlərə əsaslanan yanaşmalar və onların məşin öyrənməsi üsulları ilə istifadəsidir.

Tədqiqatın məqsəd və vəzifələri. Dissertasiya işinin əsas məqsədi azərbaycan dilindəki mətnlərin müəlliflərinin tanınması məsələlərinin tədqiqi, həll alqoritmlərinin və azərbaycan dilindəki mətnlərin müəlliflərinin tanınması kompüter sisteminin işlənilməsidir. Tədqiqatın məqsədinə uyğun olaraq dissertasiya işində aşağıdakı vəzifələr qarşıya qoyulmuşdur:

1. Azərbaycan dilindəki mətnlər üçün məşin öyrənməsindən istifadə edən müəllif tanıma üsullarının işlənilməsi.

2. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınmasında istifadə üçün müxtəlif növlərə aid mətn əlamətlərin qiymətlərinin hesablanması prosedurlarının işlənilməsi.

3. Azərbaycan dilindəki mətnlərin müəlliflərinin

tanınmasında istifadə üçün effektiv mətn əlamətlərinin seçilməsi üçün prosedurların işlənməsi.

4. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınmasına imkan verən müəllif tanıma kompüter sisteminin prinsipləri və proqram təminatının işlənməsi.

Tədqiqat metodları. Dissertasiya işində maşın öyrənməsi üsullarından, mətnlərin müəlliflərinin tanınması metodologiyaları, müasir informasiya texnologiyaları və proqramlaşdırma vasitələrindən istifadə olunub.

Müdafiyyə çıxarılan əsas müddəalar. Dissertasiyada müdafiyyə çıxarılan əsas müddəalar aşağıdakılardır:

1. Azərbaycan dilindəki böyük və kiçik həcmli bədii üslubdakı mətnlərin müəlliflərinin tanınması məsələsinin həlli.

2. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınmasında istifadə üçün mətn əlamətlərinin qiymətlərinin hesablanması alqoritmlərinin işlənməsi.

3. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınmasında istifadə üçün tanımada effektiv mətn əlamətlərinin seçilməsi prosedurlarının işlənməsi.

4. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınması kompüter sisteminin işlənməsi.

Tədqiqatın elmi yeniliyi. Dissertasiya işinin elmi yenilikləri aşağıdakılardır:

1. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınmasında istifadə üçün effektiv mətn əlamətlərinin seçilməsi prosedurları təklif olunmuşdur.

2. Müəllif tanıma üçün böyük həcmli sənədlərin (məsələn, roman, povest) və kiçik həcmli sənədlərin (məsələn, hekayələr) birlikdə emalı üçün yanaşma təklif olunmuş, təklif edilən iki empirik analizin nəticələrinin istifadəsi ilə əsaslandırılmışdır.

3. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınmasında istifadə üçün müxtəlif növlərə aid mətn əlamətlərinin qiymətlərinin hesablanması prosedurları təklif olunmuşdur.

4. Azərbaycan dilindəki mətnlər timsalında müxtəlif mətn əlaməti növlərinin və maşın öyrənməsinin müxtəlif üsullarının istifadəsi təhlil olunmuşdur.

Tədqiqatın nəzəri və praktiki əhəmiyyəti. Mətnlərin müəlliflik identifikasiyası; anonim və ya müəllifinin kim olması mübahisəli bədii əsərin müəllifinin identifikasiyasında, intihar məktublarının həqiqiliyinin yoxlanışında, kəşfiyyatda, zərərli kompüter proqramlarının müəllifinin identifikasiyasında, internetdəki bəzi mətnlərin müəlliflərinin təyində geniş istifadə olunur.

Aprobasiyası və tətbiqi. Dissertasiya işinin əsas nəticələri aşağıdakı yerli və beynəlxalq konfranslarda məruzə edilmişdir:

8th World Conference of Soft Computing (2021, Bakı, Azərbaycan); 5th International Conference on Problems of Cybernetics and Informatics (2023, Bakı, Azərbaycan); Международной научно-практической конференции с элементами научной школы - 2021, 2022, 2023 (Омск, Россия); Second International Bilateral Workshop on Science Between Dokuz Eylül University and Azerbaijan National Academy of Sciences (2022, İzmir, Türkiyə); “Riyaziyyatın tətbiqi məsələləri və yeni informasiya texnologiyaları” adlı IV Respublika elmi konfransı (2021, Sumqayıt, Azərbaycan); “Riyaziyyatın fundamental problemləri və intellektual texnologiyaların təhsildə tətbiqi” mövzusunda II Respublika elmi konfransı (2022, Sumqayıt, Azərbaycan); Azərbaycan dilinin İKT problemləri, İKT-nin Azərbaycan dili problemləri (2023, Bakı, Azərbaycan); Tələbə və Gənc Tədqiqatçıların Beynəlxalq Elmi Konfransları - 2022, 2023 (Bakı, Azərbaycan); 2nd International Conference on Information Technologies and Their Applications (2024, Bakı Azərbaycan); Azərbaycan Respublikası Elm və Təhsil Nazirliyinin İdarəetmə Sistemləri İnstitutunun elmi seminarları, (2024, Bakı Azərbaycan). Qısa müddət ərzində dissertasiya işinin bəzi nəticələrinin çap olunduğu bir işə [15] istinad olunmuşdur (baxın Alsanoosy, T., Shalbi, B., Noor, A. Authorship Attribution for English Short Texts // – Engineering, Technology & Applied Science Research – 2024. v. 14, no. 5, – pp. 16419-16426).

Dissertasiya işində süni neyron şəbəkələrin struktur və parametrik identifikasiyası üçün istifadə olunmuş proqram modulları və alınmış nəticələr Azərbaycan Respublikasının Dövlət İmtahan Mərkəzindəki “Formaların tanıma sistemi” tətbiqi proqram paketinin

təkmilləşdirilməsində istifadə olunmuşdur və bu barədə dissertasiya işinin əlavəsində tətbiq aktları əlavə edilib.

Çap olunmuş elmi əsərlər. Dissertasiya işi üzrə 21 elmi iş çap olunmuşdur, bunlardan 6-sı məqalə [2, 4, 12, 19–21], 7-si konfrans materialı [3, 8, 9, 15–18], 8-i tezisdir [1, 5–7, 10, 11, 13, 14]. 2 məqalə Clarivate Analytics agentliyinin Web of Science™ ESCI beynəlxalq bazasına, 1 məqalə PIIHQ bazasına, 1 məqalə Ukrayna AAK siyahısına daxildir. Konfrans materialları və tezislərin 4-ü Scopus, 5-i PIIHQ bazasına daxildir.

Dissertasiya işinin yerinə yetirildiyi təşkilatın adı. Dissertasiya işi Azərbaycan Respublikası Elm və Təhsil Nazirliyinin İdarəetmə Sistemləri İnstitutunda yerinə yetirilmişdir.

Dissertasiyanın struktur bölmələrinin ayrılıqda həcmi qeyd olunmaqla dissertasiyanın işarə ilə ümumi həcmi. Dissertasiya işi girişdən, 3 fəsildən, işin əsas nəticələrindən, 111 adda istifadə olunmuş ədəbiyyat siyahısından və Əlavədən ibarətdir. Dissertasiya işinin ümumi həcmi 156 səhifə, əsas həcmi isə 12 cədvəl, 12 şəkil olmaqla 126 (166435 işarə) səhifə təşkil edir. O cümlədən titul səhifəsi – 395 işarə, mündəricat – 1270 işarə, giriş – 18230 işarə, birinci fəsil – 55921 işarə, ikinci fəsil – 71491 işarə, üçüncü fəsil – 18190, nəticə – 938 işarədən ibarətdir.

İŞİN MƏZMUNU

Girişdə mövzunun aktuallığı əsaslandırılır, tədqiqatın predmeti, məqsədi və s. göstərilir, işin qısa məzmunu və əsas nəticələri verilir.

Birinci fəsildə mətnlərin müəlliflərinin tanınması məsələləri analiz olunmuşdur.

1.1 paraqrafında mətnlərin müəllif analizi istiqamətindəki məsələ sinifləri tədqiq olunmuş, müəllif tanıma məsələlərinin həm mətnlərin müəllif analizindəki, həm də mətnlərin kateqoriyalara ayrılmasındakı yeri təhlil olunmuşdur. Ardınca müəllif tanıma məsələlərinin həllində geniş istifadə olunan iki paradigmanın təhlili aparılmışdır.

Mətnlərin müəlliflik analizi anonim və ya müəllifinin kim olması mübahisəli bədii əsərin müəllifinin identifikasiyasında (xüsusilə, müəllif hüquqları ilə bağlı məhkəmə istintaqlarında), intihar məktublarının həqiqiliyinin yoxlanışında, kəşfiyyatda (məsələn, anonim ismarış və ya bəyanatların məlum bir terorist tərəfindən yazılıb-yazılmadığının təyində), zərərli kompüter proqramlarının (məsələn, kompüter virusları, ing. malware) müəllifinin identifikasiyasında, internetdəki bəzi mətnlərin (email məktubları, bloq yazıları, onlayn forum səhifələrindəki mətnlər) müəlliflərinin təyində geniş istifadə olunur.

Ümumilikdə bütün müəllif tanıma məsələlərinin həllində iki əsas həll paradigması istifadə olunur:

- müəllif profilinə əsaslanan paradigma – müəyyən müəllif namizədinin məlum mətnləri vahid bir mətn kimi başa düşülür, başqa sözlə, həll zamanı hər müəllif namizədinin ayrı-ayrı mətnləri birləşdirilərək vahid bir mətnə gətirilir, müəllif tanıma müəlliflərin vahid mətnlərindəki göstəricilərlə müəllifi tanınması nəzərdə tutulan verilmiş mətndəki göstəricilərin müqayisəsi ilə aparılır;

- mətn nümunələrinə əsaslanan paradigma – müəyyən müəllif namizədinin məlum mətnləri ayrı-ayrı mətn nümunələri kimi qəbul olunmaqla öyrətmədə istifadə olunan hər müəllifi məlum mətn bir müşahidə hesab olunur. Bu paradigma ilə maşın öyrənməsinin klassifikasiya üsullarından istifadə oluna bilər.

1.2 paragrafında bəzi mövcud müəllif tanıma kompüter sistemlərinin müqayisəli təhlili aparılmışdır.

Müəllif tanıma üçün, xaricdə xeyli tanıma sistemləri hazırlanmışdır. Bunların arasında ən geniş yayılmışlarından Dükeyn Universitetinin Java Graphical Authorship Attribution Program, NeoNeuro şirkətinin NeoNeuro Authorship Attribution və Flint AI şirkətinin Flint AI Authorship Determination kompüter sistemləridir. Adları çəkilən sistemlərin əksəriyyəti yalnız ingilis dilindəki və bəziləri bəzi digər dillərdəki mətnlərin müəlliflərini tanıyırlar.

Adları çəkilən və digər bir çox müəllif tanıma kompüter sistemlərinin heç biri azərbaycan dilindəki mətnlərin müəlliflərini tanımaq üçün nəzərdə tutulmamışdır.

1.3 paraqrafında baxılan azərbaycan dilindəki mətnlərin müəlliflərinin tanınması məsələsinin xüsusiyyətləri barədə qeyd olunmuş, işdə timsalında kompüter eksperimentləri aparılmış müəllifi məlum mətnlər çoxluğundakı (datasetdəki) böyük həcmli və kiçik həcmli bədii əsərlərin birlikdə emalında istifadə olunan yanaşmanın effektivliyinin qiymətləndirilməsi üçün aparılmış kompüter eksperimentlərinin nəticələri təhlil olunmuşdur. Həmin paraqrafın üçüncü bəndində müəlliflərin əsərlərindəki stilistik xarakteristikaları onların statistik göstəriciləri əsasında təhlil olunmuşdur.

İşdə baxılan müəllif tanıma məsələsində istifadə olunan dataset azərbaycanlı 11 yazıçının böyük həcmli və kiçik həcmli əsərlərinin yığılması nəticəsində hazırlanmışdır. Datasetdə müəllif başına düşən böyük həcmli əsər sayı 1-5 arasında, kiçik həcmli əsər sayı 0-46 arasında dəyişir.

Baxılan müəllif tanıma məsələsində roman-povest və hekayələrin müəlliflərinin tanınması bir yerdə aparılır. Tanımada effektiv əlamətlərin seçilməsi və riyazi modellərin parametrlərinin təyini üçün istifadə olunacaq öyrətmə bazasındakı müəllifi məlum mətnlərin ümumilikdə az olması müəllif tanıma məsələsinin etibarlı və dəqiq şəkildə həll olunmasını daha da çətinləşdirir. Baxılan məsələdəki müəllif namizədlərinin – yazıçıların böyük bir qismi artıq ölmüşdür, ona görə də, müəllif tanıma məsələsinin həllində istifadə olunacaq müəllifi məlum mətnlərin sayını təbii yolla artırmaq əlçatan deyil. Baxılan məsələnin etibarlı və dəqiq həlli üçün həm müəllifi naməlum əsərlərin sayının ümumilikdə azlığı, həm də böyük və kiçik həcmli əsərlərin birlikdə emalındakı çətinliklərin üstəsindən gəlmək məqsədilə böyük həcmli əsərlərin müəyyən sayda hissələrə bölünməsi yanaşmasından istifadə olunub. Lakin hər hansı böyük həcmli əsəri hissələrə böləndə alınan ayrı-ayrı mətnlərin müəlliflik stilistikası baxımından ayrı-ayrı nümunələr – əsərlər hesab oluna biləcəyi müəyyən edilməlidir. Həmin mətnlərin ayrı-ayrı əsər nümunələri hesab olunması üçün iki şərt ödənməlidir:

- böyük həcmli bir əsəri böləndə alınan mətnlər müəlliflik stilistikası baxımından həmin böyük həcmli mətnə kifayət qədər yaxın olmalıdır,

- müəllifin kiçik həcmli əsərlərindəki müəlliflik stilistikası

nəzərə alınmaqla böyük həcmli bir əsərin bölünməsinə alınan hissələr stilistik baxımdan bir-birindən fərqlənə bilər.

Baxılan məsələdə 1-ci şərtin ödənməsinin yoxlanışı üçün böyük həcmli bir əsərin ayrı-ayrı mətnlərə bölünməsi zamanı mətnlərdə müəllifin stilistik göstəricilərinin saxlanılıb-saxlanılmadığı mətnlərdəki simvol n-qramlarının tezlikləri əlamətləri timsalında analiz olunmuşdur. Simvol n-qramı dedikdə müəyyən n sayda simvolun kombinasiyası nəzərdə tutulur. Simvol n-qramı tezliyi isə müəyyən simvol n-qramının müəyyən mətnə istifadə olunma tezliyidir. Bunun üçün bir böyük həcmli əsərin mətni (demək olar bərabər simvol sayıyla) 10 mətnə bölünmüşdür. Simvol 2-qramlarının bu mətnlərin hər birindəki tezlikləri və bu tezlik qiymətlərinin standart yayınması hesablanmışdır. Standart yayınma qiymətləri kiçik olan simvol 2-qramlarının da mövcud olması böyük həcmli həmin mətnlərdəki müəlliflik xarakteristikalarının bu mətn bölünəndə alınan mətnlərdə də saxlanıldığını göstərir. Başqa sözlə, böyük həcmli əsər bölünəndə alınan mətnlərdə müəllifin həmin böyük həcmli əsərdəki xarakteristikaları saxlanılır (Cədvəl 1).

Cədvəl 1-də soldan sağa 1-ci sütunda sətir nömrəsi, 2-ci sütunda tezliklərindən istifadə olunan simvol n-qramı, 3-cü sütunda birinci mətn hissəsindəki (burada mətn hissəsi nömrəsi şərti olaraq verilir) tezliyi, 4-cü sütunda ikinci mətn hissəsindəki tezliyi, 6-cı sütunda sonuncu, yəni onuncu mətn hissəsindəki tezliyi, 7-ci sütunda böyük həcmli mətnin özündəki tezliyi, 8-ci, yəni soldan axırıncı sütunda isə nəhayət mətnin ayrı-ayrı hissələrindəki tezliklərin standart yayınması verilmişdir. Cədvəlin sütun adlarındakı işarələr və onların təsvirləri aşağıda verilmişdir:

$\alpha_{i_m}^n$ – ixtiyari i_m^n multiindeksli müəyyən n simvollu simvol n-qramı, burada $\alpha_{i_m}^n = (\alpha_{i_1}, \dots, \alpha_{i_n})$, $i_m^n = (i_1, \dots, i_n)$, m simvol əlifbasındakı simvolların sayıdır. Qeyd olunan multiindeksdəki i_v , $v = 1, 2, \dots, n$, indekslərinin hər biri verilmiş $\{1, 2, \dots, m\}$ çoxluğundakı qiymətləri ala bilər, yəni $1 \leq i_v \leq m$, $v = 1, 2, \dots, n$;

$\varphi_{j1i_m}^i$ – i nömrəli müəllifin j nömrəli mətni T_j^i -nin (onu böldükdə alınan) 1 nömrəli hissəsində $\alpha_{i_m}^n$ simvol n-qramının tezliyi;

$\varphi_{j2i_m}^i$ və $\varphi_{j10i_m}^i - T_j^i$ mətninin 2 və 10 nömrəli hissələrindəki

$\alpha_{i_m}^n$ simvol n-qramının tezlikəri;

$f_{3j i_m}^i - T_j^i$ mətnində $\alpha_{i_m}^n$ simvol n-qramının tezliyi;

$\sigma_{j i_m}^i - \alpha_{i_m}^n$ simvol n-qramının T_j^i mətninin hissələrindəki $\varphi_{j 1 i_m}^i$,

$\varphi_{j 2 i_m}^i, \dots, \varphi_{j 10 i_m}^i$ qiymətlərinin standart yayınması;

burada $j = 1, 2, \dots, l_i$, $i = 1, 2, \dots, L$, L baxılan məsələdəki müəllif namizədlərinin sayıdır. Qeyd edək ki, müəyyən $\alpha_{i_m}^n$ simvol n-qramlarının T_j^i mətninin digər (3, 4, ..., 9-cu) hissələrindəki tezlikləri bu analizi aparmaq üçün kifayət etdiyinə görə verilməyib.

Yuxarıda da qeyd olunduğu kimi verilmiş T_j^i mətnində $\alpha_{i_m}^n$ simvol n-qramının tezliyi $f_{3j i_m}^i$ ilə işarə olunub. Burada indeksdəki “3” bu əlamətin simvol n-qramı tezliyi növünə aid olduğunu bildirir. Dissertasiya işinin ikinci fəslində dissertasiya işində baxılan müəllif tanıma məsələsində istifadə olunan beş növə aid əlamətlərin qiymətlərinin hesablanması qaydası təsvir edilmişdir. Burada – birinci fəslin üçüncü paraqrafında T_j^i mətnində $\alpha_{i_m}^n$ simvol n-qramının tezliyi $f_{3j i_m}^i$ işarəsindəki “3” ikinci fəsilə mətn əlamətlərinin qiymətlərinin hesablanması qaydaları və tanımadada effektiv əlamətlərin seçilməsi prosedurlarının təsvirində istifadə olunan işarələrlə eyni olması üçün qeyd olunub.

İkinci şərtin baxılan məsələdə ödənməsinin yoxlanıldığı analizdə böyük həcmli əsər bölündükdə alınan mətnlərdəki stilistik göstəricilərin kiçik həcmli əsərlərdəki göstəricilərlə yenə simvol n-qramları tezliklərinin timsalında aparılan müqayisəsinə əsaslanır. Təklif edilən bu analizdə bir böyük həcmli və birdən çox kiçik həcmli əsərin mətnlərindən istifadə olunur.

Verilmiş müəllif A^i -nin mətnlərinin nömrələri çoxluğunu $J^i = \bar{J}^i \cup \underline{J}^i$ ilə işarə edək, burada $\bar{J}^i = \{\bar{j}_k^i: k = 1, 2, \dots, \bar{\xi}_i\}$ və $\underline{J}^i = \{\underline{j}_k^i: k = 1, 2, \dots, \underline{\xi}_i\}$ uyğun olaraq böyük həcmli və kiçik həcmli mətnlərin nömrələri çoxluğuudur.

Verilmiş $\alpha_{i_m}^n$ simvol n-qramının müəllif A^i -nin \bar{j}_1^i nömrəli böyük həcmli əsəri bölünəndə alınan $T_{\bar{j}_1^i}^i, T_{\bar{j}_2^i}^i, \dots, T_{\bar{j}_1^i \mathcal{N}_{\bar{j}_1^i}^i}^i$

mətnlərindəki $\varphi_{j_1^i 1 i_m^n}^i, \varphi_{j_1^i 2 i_m^n}^i, \dots, \varphi_{j_1^i \mathcal{N}_{j_1^i}^i i_m^n}^i$ tezliklərinin bir-birilərinə nəzərən orta fərqlənmə dərəcəsi aşağıdakı kimidir:

$$\mathcal{F}_{i_m^n}^{1,1} = \frac{2}{\left(\mathcal{N}_{j_1^i}^i\right)^2} \sum_{x=1}^{\mathcal{N}_{j_1^i}^i} \sum_{y=1}^{\mathcal{N}_{j_1^i}^i/2} \text{abs}\left(\varphi_{j_1^i x i_m^n}^i - \varphi_{j_1^i y i_m^n}^i\right),$$

$\text{abs}(\cdot)$ ədədin mütləq qiymətidir, \mathcal{N}_j^i müəllif A^i -nin j nömrəli T_j^i mətninin bölündüyü hissələrin sayıdır.

Verilmiş $\alpha_{i_m^n}^n$ simvol n-qramının müəllif A^i -nin $\underline{j}_1^i, \underline{j}_2^i, \dots, \underline{j}_{\xi_i}^i$ nömrəli kiçik həcmli $T_{\underline{j}_1^i}^i, T_{\underline{j}_2^i}^i, \dots, T_{\underline{j}_{\xi_i}^i}^i$ mətnlərindəki $f_{3\underline{j}_1^i i_m^n}^i, f_{3\underline{j}_2^i i_m^n}^i, \dots, f_{3\underline{j}_{\xi_i}^i i_m^n}^i$ tezliklərinin bir-birilərinə nəzərən orta fərqlənmə dərəcəsi aşağıdakı kimidir:

$$\mathcal{F}_{i_m^n}^{2,2} = \frac{2}{\left(\xi_i\right)^2} \sum_{v=1}^{\xi_i} \sum_{z=1}^{\xi_i/2} \text{abs}\left(f_{3\underline{j}_v^i i_m^n}^i - f_{3\underline{j}_z^i i_m^n}^i\right).$$

Verilmiş $\alpha_{i_m^n}^n$ simvol n-qramı üzrə müəllif A^i -nin böyük həcmli mətn əsərinin hissələri $T_{j_1^i 1}^i, T_{j_1^i 2}^i, \dots, T_{j_1^i \mathcal{N}_{j_1^i}^i}^i$ ilə kiçik həcmli $T_{\underline{j}_1^i}^i, T_{\underline{j}_2^i}^i, \dots, T_{\underline{j}_{\xi_i}^i}^i$ mətnlərinin bir-birilərinə nəzərən orta fərqlənmə dərəcəsi aşağıdakı kimidir:

$$\mathcal{F}_{i_m^n}^{1,2} = \frac{2}{\mathcal{N}_{j_1^i}^i \xi_i} \sum_{x=1}^{\mathcal{N}_{j_1^i}^i} \sum_{v=1}^{\xi_i} \text{abs}\left(\varphi_{j_1^i x i_m^n}^i - f_{3\underline{j}_v^i i_m^n}^i\right).$$

Müəyyən $\alpha_{i_m^n}^n$ simvol n-qramı üzrə böyük həcmli əsərin hissələrinin bir-birinə stilistik baxımdan fərqlənmə dərəcəsinin nə kiçik həcmli əsərlərin bir-birinə stilistik baxımdan fərqlənmə dərəcəsinə, nə də böyük həcmli əsərin hissələrinin bir-birinə stilistik baxımdan fərqlənmə dərəcəsinə yaxın olmaması aşağıdakı $\mathcal{F}_{i_m^n}^n$ -in kiçik olması kimi hesab olunur:

$$\dot{\mathcal{F}}_{i_m^n} = \text{abs} \left(\mathcal{F}_{i_m^n}^{1,2} - \frac{\mathcal{F}_{i_m^n}^{1,1} + \mathcal{F}_{i_m^n}^{2,2}}{2} \right).$$

Müəyyən $\alpha_{i_m^n}$ simvol n-qramı üzrə $\dot{\mathcal{F}}_{i_m^n}$ qiymətinin sadə interpretasiyası məqsədilə aşağıdakı kəmiyyət daxil edilir:

$$\ddot{\mathcal{F}}_{i_m^n} = \frac{\dot{\mathcal{F}}_{i_m^n}}{\text{abs} \left(\mathcal{F}_{i_m^n}^{1,1} - \mathcal{F}_{i_m^n}^{2,2} \right)} \times 100.$$

Cədvəl 1

Böyük həcmli əsərin ayrı-ayrı hissələrinin müqayisəsinin nəticələri

N	$\alpha_{i_m^n}$	$\varphi_{j_1 i_m^n}^i$	$\varphi_{j_2 i_m^n}^i$...	$\varphi_{j_{10} i_m^n}^i$	$f_{j_3 i_m^n}^i$	$\sigma_{j i_m^n}^i$
1	an	0.0140	0.0127	...	0.0128	0.0125	0.0020
2	da	0.0129	0.0128	...	0.0126	0.0122	0.0018
3	in	0.0092	0.0118	...	0.0113	0.0121	0.0017
4	ar	0.0131	0.0126	...	0.0121	0.0119	0.0015
5	la	0.0125	0.0122	...	0.0105	0.0110	0.0015

Cədvəl 2-də sağdan birinci sütundakı faiz qiymətlərin kiçik olmasından görünür ki, elə mətn əlamətləri tapmaq olar ki, böyük həcmli bir əsəri böldükdə alınan mətnlər müəllif stilistikası baxımından nə həmin böyük həcmli əsərə, nə də kiçik həcmli əsərlərə yaxın olsun. Cədvəl 2-5-də həcm məhdudiyətini nəzərə yalnız 5 sətir verilmişdir.

Cədvəl 2

Böyük həcmli əsərin ayrı-ayrı hissələri ilə kiçik həcmli əsərlərin müqayisəsinin nəticələri

N	$\alpha_{i_m^n}$	$\mathcal{F}_{i_m^n}^{1,1}$	$\mathcal{F}_{i_m^n}^{1,2}$	$\mathcal{F}_{i_m^n}^{2,2}$	$\ddot{\mathcal{F}}_{i_m^n}$
1	ae	0	2.86197E-06	5.34234E-06	3.57
2	vn	9.63262E-06	7.64103E-06	5.34234E-06	3.58
3	vg	4.81631E-06	9.79213E-05	0.000178458	3.62
4	lz	4.81631E-06	6.60234E-05	0.000118915	3.64
5	lh	4.81631E-06	2.16771E-05	3.61856E-05	3.75

İşdə istifadə olunan mətn əlaməti növləri arasından simvol n-qramı tezliyi və söz tezliyi növü mümkün əlamətlərin sayının çox olması (məsələn, mətndəki tezlikləri mətn əlaməti kimi istifadə oluna biləcək sözlərin çox olması) ilə fərqlənir. Ona görə də, bu növlərə aid əlamətlər arasından tanımada effektiv olacaq əlamətləri seçmək lazım gəlir. İşdə bu iki növə aid əlamətlərin seçilməsi üçün dörd müxtəlif əlamət seçimi proseduru istifadə olunmuşdur. Bu əlamət seçimi prosedurlarının riyazi təsviri verilmiş və baxılan müəllif tanıma məsələsi üzərində aparılmış kompüter eksperimentlərinin nəticələri əsasında tanımada effektivlikləri analiz edilmişdir. Birinci fəslin üçüncü paragrafında isə iki yazıçının əsərləri üzərində $n = 2$ simvollu simvol n-qramlarının tezlikləri və ümumilikdə baxılan məsələdəki müəlliflər tərəfindən ümumilikdə ən çox istifadə olunan sözlərin tezlikləri analiz olunmuşdur. Cədvəl 3-də iki müəllifin – Süleyman Rəhimov və İsmayıl Şıxlının mətnlərindəki bəzi simvol n-qramlarının tezliklərinin orta qiyməti və dispersiyası verilmişdir, burada

$$\mu_{i_m}^i = \frac{1}{l_i} \sum_{j=1}^{l_i} f_{3ji_m}^i,$$

$$\sigma_{i_m}^i = \frac{1}{l_i} \sum_{j=1}^{l_i} (f_{3ji_m}^i - \mu_{i_m}^i)^2.$$

Yuxarıdakı $\sigma_{i_m}^i$ işarəsində siqmanın (“ σ ”) üzərindəki nöqtə onu $\sigma_{ji_m}^i$ işarəsindən rahat ayıra bilmək üçün nəzərdə tutulub, burada $\sigma_{ji_m}^i$ $\alpha_{i_m}^n$ simvol n-qramının T_j^i mətninin T_{jk}^i , $k = 1, 2, \dots, \mathcal{N}_j^i$, hissələrindəki $\varphi_{j1i_m}^i$, $\varphi_{j2i_m}^i$, $\varphi_{j10i_m}^i$ tezliklərinin standart yayınması idi, lakin $\sigma_{i_m}^i$ i nömrəli müəllif A^i -nin T_j^i , $j = 1, 2, \dots, l_i$, mətnlərində $\alpha_{i_m}^n$ simvol n-qramının tezliklərinin standart yayınmasıdır.

Cədvəl 3-dən görünür ki, müxtəlif simvol n-qramları üzrə müxtəlif müəlliflərin statistik göstəriciləri bir-birindən fərqlənə bilər. Aydın ki, bu cədvəldəki dispersiya qiymətləri orta qiymətə görə miqyasca dəyişir. Ona görə də, bütün müəlliflər üzrə nəticələrin daha

asan interpretasiya oluna bilməsi üçün aşağıdakı variasiya əmsalından istifadə edirik:

$$d_{i_m}^i = \begin{cases} \frac{\hat{\sigma}_{i_m}^i}{\hat{\mu}_{i_m}^i}, & \text{əgər } \mu_k^j \geq \rho, \\ \rho, & \text{əks halda,} \end{cases}$$

burada ρ hansısa kiçik ədəd, ρ isə hansısa çox böyük ədəddir, məsələn, $\rho = 10^{-4}$, $\rho = 10^{12}$ və eksperimentlərin nəticələri əsasında seçilir. Bu $d_{j_m}^i$ kəmiyyətlərinin bəzi simvol n-qramları üzrə qiymətləri isə cədvəl 4-də verilmişdir. Bu cədvəldəki qiymətlərdən görünür ki, müxtəlif əlamətlər üzrə müxtəlif müəlliflərin statistik göstəriciləri dəyişir.

Cədvəl 3

İki müəllifin mətnləri üzərində simvol n-qramları tezlikləri üzrə statistik göstəricilər

N	$\alpha_{i_m}^n$	$\hat{\mu}_{i_m}^i$		$[\hat{\sigma}_{i_m}^i]^2$	
		Süleyman Rəhimov ($i = 9$)	İsmayıl Şıxlı ($i = 10$)	Süleyman Rəhimov ($i = 9$)	İsmayıl Şıxlı ($i = 10$)
1	aa	0.00025	0.00042	3.20254E-08	2.74829E-08
2	ab	0.00235	0.00146	9.76353E-07	5.05273E-08
3	ac	0.00125	0.00135	3.12249E-08	2.30432E-08
4	aç	0.00103	0.00092	7.07818E-08	3.59291E-09
5	ad	0.00374	0.00380	4.79537E-07	1.15944E-07

Cədvəl 5-də baxılan müəllif tanıma məsələsindəki müəllif namizədlərinin öyrətmə bazasındakı mətnlərində ən çox istifadə olunan beş sözün bu müəlliflər tərəfindən öyrətmə bazasındakı mətnlərində istifadə tezliyi verilmişdir.

Cədvəl 4

Müxtəlif müəlliflərin mətnləri üzərində bəzi simvol n-qramları tezlikləri üzrə statistik göstəricilər

N	$\alpha_{i_m}^n$	$d_{i_m}^1$	$d_{i_m}^2$	$d_{i_m}^3$...	$d_{i_m}^9$	$d_{i_m}^{10}$	$d_{i_m}^{11}$
1	aa	0.55	0.78	0.55	...	1.02	0.60	0.96
2	ab	0.97	0.45	0.36	...	0.59	0.29	0.70
3	ac	0.55	0.77	0.23	...	0.28	0.17	0.42
4	aç	0.66	0.86	0.34	...	0.43	0.19	0.66
5	ad	0.23	0.39	0.16	...	0.20	0.13	0.49

Cədvəl 5

Öyrətmə bazasında ən çox işlənən 5 sözün müəlliflərin mətnlərindəki işlənmə tezlikləri (faizlə)

v	ω_v	$\phi_v^{5,1,1}$	$\phi_v^{5,1,2}$	$\phi_v^{5,1,3}$...	$\phi_v^{5,1,10}$	$\phi_v^{5,1,11}$	$\phi_v^{5,2}$
1	bir	2.0	2.7	1.9	...	1.4	2.7	0.0
2	bu	2.1	1.6	1.6	...	0.9	1.5	0.0
3	ki	1.9	2.7	1.4	...	0.8	1.2	0.0
4	də	1.5	1.1	1.3	...	1.1	0.4	0.0
5	və	2.6	3.1	1.0	...	0.8	2.4	0.0

Simvol n-qramları və sözlərin tezlikləri və digər statistik göstəricilərdən (cədvəl 3-5-də verilmişdir) aydın olur ki, müxtəlif əlamətlər müxtəlif müəllif namizədləri tərəfindən istifadə göstəriciləri bir-birindən fərqlənir. Ona görə də, işdə tanımada effektiv əlamətlərin seçimi prosedurlarından istifadə olunmuşdur.

İkinci fəsilə azərbaycan dilindəki mətnlərin müəlliflərinin tanınması üçün istifadə olunan metodologiya təhlil olunmuşdur.

2.1 paraqrafında işdə istifadə olunan tanıma üsulları barədə qeyd olunmuşdur.

İşdə baxılan müəllif tanıma məsələsində maşın öyrənməsinin dayaq vektorlar üsulu (ing. support vector machine, SVM), random forest (RF) üsulları və birbaşa əlaqəli süni neyron şəbəkə (ing.

multilayer feedforward artificial neural network, ANN) modellərindən istifadə olunmuşdur.

2.2 paraqrafında məlum mətn əlamətlərinin qısa ədəbiyyat icmalı aparılmış, baxılan müəllif tanıma məsələsində istifadə olunan növlərə aid mətn əlamətlərinin qiymətlərinin hesablanması qaydaları və tanımada effektiv əlamətlərin seçilməsi üçün təklif edilən prosedurlar təsvir olunmuşdur.

Müəllif tanımada hər hansı verilmiş mətn müəyyən əlamət qrupundakı müəyyən mətn əlamətləri ilə xarakterizə olunur. Bu əlamət qrupundakı əlamətlər müxtəlif əlamət növlərinə aid ola bilər. İşdə cümlə uzunluğu tezliyi, söz uzunluğu tezliyi, simvol n-qramı tezliyi, (işdə təklif olunan) simvol n-qramı tezliklərinin dispersiyası, söz tezliyi növlərinə aid əlamətlərdən ibarət əlamət qruplarından istifadə olunmuşdur. Bu əlamət qruplarındakı əlamətlərin müəyyən bir mətndəki qiymətləri həmin mətnin əlamət vektorunu təşkil edir. Aşağıdakı işarələrin istifadəsi ilə mətn əlamətlərinin qiymətlərinin hesablanması və tanımada effektiv mətn əlamətlərinin seçilməsi prosedurlarının riyazi təsvirləri verilmişdir.

Verilmiş müəllif namizədləri çoxluğunu $A = \{A^i: i = 1, 2, \dots, L\}$ ilə işarə edək, burada $L = |A|$ müəllif namizədlərinin sayıdır (burada və sonra $|D|$ işarəsi ixtiyari D çoxluğunun elementlərinin sayını göstərir). Müəllif tanımada istifadə olunacaq müəllif A^i -nin yazdığı mətnlər çoxluğunu $T^i = \{T_j^i: j = 1, 2, \dots, l_i\}$ ilə işarə edək, burada T_j^i müəllif A^i -nin j nömrəli mətnini, l_i müəllif A^i -nin mətnlərinin sayını göstərir, $i = 1, 2, \dots, L$.

Mətnlərin müəllifliyinin tanınması məsələsində müəllif namizədlərinin yazdıqları mətnlər çoxluğunu (dataset) $T = \{T^i: i = 1, 2, \dots, L\}$ ilə işarə edək. Hər hansı T_j^i mətni s_{jk}^i , $k = 1, 2, \dots, N_j^{si}$, cümlələrindən ibarət, hər hansı s_{jk}^i cümləsi w_{jkv}^i , $v = 1, 2, \dots, N_{jk}^{wi}$, sözlərindən ibarət, hər hansı w_{jkv}^i sözü $c_{jkv\eta}^i$, $\eta = 1, 2, \dots, N_{jkv}^{ci}$, hərflərindən çoxluqlardır və onlar aşağıdakı kimi işarə olunurlar:

$$T_j^i = \{s_{jk}^i: k = 1, 2, \dots, N_j^{si}\}, \quad j = 1, 2, \dots, l_i, \quad i = 1, 2, \dots, L,$$

$$s_{jk}^i = \{w_{jkv}^i: v = 1, 2, \dots, N_{jk}^{wi}\}, \quad k = 1, 2, \dots, N_j^{si}, \quad j = 1, 2, \dots, l_i, \quad i = 1, 2, \dots, L,$$

$$w_{j_{kv}}^i = \{c_{j_{kv}\eta}^i: \eta = 1, 2, \dots, N_{j_{kv}}^{ci}\}, \quad v = 1, 2, \dots, N_{jk}^{wi}, k = 1, 2, \dots, N_j^{si}, j = 1, 2, \dots, l_i, i = 1, 2, \dots, L.$$

Burada T_j^i A^i müəllifinin j nömrəli mətnidir; s_{jk}^i , $k = 1, 2, \dots, N_j^{si}$, T_j^i mətnindəki cümlələrdir; $w_{j_{kv}}^i$, $v = 1, 2, \dots, N_{jk}^{wi}$, T_j^i mətnindəki k nömrəli cümlədəki sözlərdir, burada $k = 1, 2, \dots, N_j^{si}$; $c_{j_{kv}\eta}^i \in \alpha^1$, $\eta = 1, 2, \dots, N_{j_{kv}}^{ci}$, T_j^i mətnindəki k nömrəli cümlədəki v nömrəli sözdəki hərfləri bildirir, burada $v = 1, 2, \dots, N_{jk}^{wi}$, $k = 1, 2, \dots, N_j^{si}$, və N_j^{si} , N_{jk}^{wi} , $N_{j_{kv}}^{ci}$ uyğun olaraq T_j^i mətnindəki cümlələrin sayı, T_j^i mətnindəki k nömrəli cümlədəki sözlərin sayı, T_j^i mətnindəki k nömrəli cümlədəki v nömrəli sözdəki hərflərin sayıdır, α^1 simvol əlifbası çoxluğudur, $|\alpha^1|$ həmin simvol əlifbası çoxluğundakı simvolların sayıdır.

Müəllif tanıma məsələsinin həllində istifadə olunan müəllifi məlum mətnlər çoxluğu - dataset T iki bir-birilə kəsişməyən altçoxluğa bölünür $T = \bar{T} \cup \tilde{T}$, $\bar{T} \cap \tilde{T} = \emptyset$. Bunlardan biri \bar{T} öyrətmə bazası (ing. training set), digəri \tilde{T} test bazasıdır (ing. test set).

Cümlə uzunluğu tezliyi. Cümlə uzunluğu dedikdə cümlədəki sözlərin sayı, söz uzunluğu dedikdə sözdəki hərflərin sayı nəzərdə tutulacaq. Mətnlər çoxluğu T -dəki i nömrəli müəllifin j nömrəli T_j^i mətnində cümlə uzunluğu (cümlədəki sözlərin sayı) v olan cümlələrdən istifadə sayı aşağıdakı kimidir:

$$b_{1jv}^i = |\{s_{jk}^i: N_{jk}^{wi} = v, k = 1, 2, \dots, N_j^{si}\}|,$$

burada $v \in I = \{1, 2, \dots\}$, $j = 1, 2, \dots, l_i$, $i = 1, 2, \dots, L$.

Mətnlər çoxluğu T -dəki i nömrəli müəllifin j nömrəli T_j^i mətnindəki cümlə uzunluğu (cümlədəki sözlərin sayı) v olan cümlələrin tezliyi aşağıdakı kimi hesablanır:

$$f_{1jv}^i = \frac{b_{1jv}^i}{N_j^{si}},$$

burada $v \in I$.

Söz uzunluğu tezliyi. Hər hansı T_j^i mətnində söz uzunluğu (sözdəki hərflərin sayı) v olan sözlərdən istifadə sayı aşağıdakı kimidir:

$b_{2jv}^i = |\{w_{jkv}^i: N_{jkv}^{ci} = v, v = 1, 2, \dots, N_{jk}^{wi}, k = 1, 2, \dots, N_j^{si}\}|$,
burada $v \in I$.

Mətnlər çoxluğu T -dəki i nömrəli müəllifin j nömrəli T_j^i mətnindəki söz uzunluğu (sözdəki hərflərin sayı) v olan sözlərin tezliyi aşağıdakı kimi hesablanır:

$$f_{2jv}^i = \frac{b_{2jv}^i}{\sum_{k=1}^{N_j^{si}} N_{jk}^{wi}}$$

burada $v \in I$.

Simvol n-qramı tezliyi. Ədəbiyyatda simvol n-qramı dedikdə verilmiş n ($n = 1, 2, \dots$) simvolun verilmiş hər hansı kombinasiyasının nəzərdə tutulmağı qəbul olunub. Verilmiş n üçün hər hansı simvol n-qramındakı simvolların seçildiyi simvol əlifbası çoxluğunu, yəni bütün simvol 1-qramları çoxluğunu aşağıdakı kimi işarə edək:

$$\alpha^1 = \{\alpha_k: k = 1, 2, \dots, |\alpha^1|\}.$$

Bundan sonra \hat{i}_m^n n ölçülü multiindeksi üçün $\hat{i}_m^n = (i_1, \dots, i_n)$ işarəsindən istifadə olunacaq. Burada $i_v, v = 1, 2, \dots, n$, indekslərinin hər biri verilmiş $\{1, 2, \dots, m\}$ çoxluğundakı qiymətləri ala bilər, yəni $1 \leq i_v \leq m, v = 1, 2, \dots, n$.

Simvolları simvol əlifbası çoxluğu α^1 -dən seçilən ixtiyari \hat{i}_m^n multiindeksli simvol n-qramı $\alpha_{\hat{i}_m^n} = (\alpha_{i_1}, \dots, \alpha_{i_n})$ kimi istifadə olunacaq, burada $m = |\alpha^1|$.

Bütün simvol n-qramları çoxluğunu isə aşağıdakı kimi işarə edək:

$$\alpha^n = \{\alpha_{\hat{i}_m^n}: 1 \leq i_v \leq m, v = 1, 2, \dots, n, m = |\alpha^1|\}.$$

Mətnlər çoxluğu T -dəki i nömrəli müəllifin j nömrəli T_j^i mətnindəki simvol n-qramı tezliklərinin hesablanması qaydasının təsviri üçün T_j^i mətnini aşağıdakı kimi $\tilde{c}_{jk}^i, k = 1, 2, \dots, \tilde{N}_j^i$, simvollarının çoxluğu şəklində işarə edək:

$$T_j^i = \{\tilde{c}_{j1}^i, \dots, \tilde{c}_{j\tilde{N}_j^i}^i\},$$

burada $\tilde{N}_j^i T_j^i$ mətnindəki simvolların sayıdır.

Hər hansı T_j^i mətnində ixtiyari simvol n-qramı $\alpha_{\hat{i}_m^n} \in \alpha^n$ -dən

istifadə olunma sayı aşağıdakı kimidir:

$$b_{3j\hat{i}_m}^i = \left| \left\{ \{ \tilde{c}_{jk}^i, \dots, \tilde{c}_{jk+n-1}^i \} : \{ \tilde{c}_{jk}^i, \dots, \tilde{c}_{jk+n-1}^i \} = \alpha_{i_m}^n, k = 1, 2, \dots, N_j^i \right\} \right|, m = |\alpha^1|,$$

$N_j^i T_j^i$ mətnindən çıxarıla bilən simvol n-qramlarının sayıdır.

Hər hansı T_j^i mətnindəki ixtiyari simvol n-qramı $\alpha_{i_m}^n \in \alpha^n$ -nin tezliyi aşağıdakı kimi hesablanır:

$$f_{3j\hat{i}_m}^i = \frac{b_{3j\hat{i}_m}^i}{N_j^i}, m = |\alpha^1|.$$

Simvol n-qramı tezliklərinin dispersiyası. İxtiyari verilmiş mətnin ayrı-ayrı hissələrindəki simvol n-qramı tezliklərinin statistik göstəriciləri (məsələn, dispersiyaları) həmin mətnin əlamətləri kimi istifadə oluna bilər. İşdə bu növə aid əlamətlərin tanımada effektivliyinin analizi aparılmışdır.

İxtiyari n üçün hər hansı simvol n-qramının verilmiş mətndəki statistik göstəricilərinin qiymətlərinin hesablanması üçün əvvəl həmin mətn müəyyən hissələrə bölünməli, həmin simvol n-qramının bu mətn hissələrinin hər birindəki tezliyi tapılmalı, bu tezliklərin dispersiya qiyməti hesablanmalıdır.

Hər hansı T_j^i mətninin \mathcal{N}_j^i sayda ayrı-ayrı, bir-birilə kəsişməyən hissələri aşağıdakı kimidir:

$$T_j^i = \left\{ T_{jk}^i : k = 1, 2, \dots, \mathcal{N}_j^i, T_{jk_1}^i \cap T_{jk_2}^i = \emptyset, 1 \leq k_1, k_2 \leq \mathcal{N}_j^i, k_1 \neq k_2 \right\},$$

burada, T_{jk}^i T_j^i mətninin k nömrəli hissəsidir (mətn), \mathcal{N}_j^i T_j^i mətninin hissələrinin sayıdır, $j = 1, 2, \dots, l_i$, $i = 1, 2, \dots, L$.

İxtiyari $\alpha_{i_m}^n \in \alpha^n$ simvol n-qramının T_j^i mətninin k nömrəli T_{jk}^i hissəsindəki tezliyini $\varphi_{jk\hat{i}_m}^i$ ilə işarə edək, burada $m = |\alpha^1|$, $j = 1, 2, \dots, l_i$, $i = 1, 2, \dots, L$.

İxtiyari $\alpha_{i_m}^n \in \alpha^n$ simvol n-qramının T_j^i mətninin T_{jk}^i hissəsindəki $\varphi_{jk\hat{i}_m}^i$, $k = 1, 2, \dots, \mathcal{N}_j^i$, tezliklərinin dispersiyası aşağıdakı kimi hesablanır:

$$f_{A_j l_m^i}^i = \sigma_{j l_m^i}^i = \frac{1}{N_j^i} \sum_{k=1}^{N_j^i} (\varphi_{j k l_m^i}^i - \mu_{j l_m^i}^i)^2,$$

burada $\mu_{j l_m^i}^i = \frac{1}{N_j^i} \sum_{k=1}^{N_j^i} \varphi_{j k l_m^i}^i$ ixtiyari $\alpha_{l_m^i}^i \in \alpha^n$ simvol n-qramının T_j^i mətninin T_{jk}^i hissəsindəki $\varphi_{j k l_m^i}^i$, $k = 1, 2, \dots, N_j^i$, tezliklərinin orta qiymətidir, $m = |\alpha^1|$, $j = 1, 2, \dots, l_i$, $i = 1, 2, \dots, L$.

Söz tezliyi. Öyrətmə bazasındakı mətnlər çoxluğu \bar{T} -dəki T_j^i , burada $T_j^i \in \bar{T}$, $j = 1, 2, \dots, l_i$, $i = 1, 2, \dots, L$, mətnlərindəki $w_{j k v}^i$, $v = 1, 2, \dots, N_{j k}^{w_i}$, $k = 1, 2, \dots, N_j^{s_i}$, burada $T_j^i \in \bar{T}$, $j = 1, 2, \dots, l_i$, $i = 1, 2, \dots, L$, sözlərinin təkrarlanmayan çoxluğunu aşağıdakı kimi işarə edək:

$$W = \{ \omega_p : T_j^i \in \bar{T}, \omega_p \in s_{j k}^i, k = 1, 2, \dots, N_j^{s_i}, j = 1, 2, \dots, l_i, i = 1, 2, \dots, L, \omega_{p_1} \neq \omega_{p_2} \text{ əgər } p_1 \neq p_2 \},$$

burada $\omega_p = \{ \alpha_r^p : \alpha_r^p \in \alpha^1, r = 1, 2, \dots, |\omega_p| \}$, α^1 simvol əlifbası çoxluğu.

Hər hansı T_j^i mətnində $\omega_v \in W$ sözündən istifadə olunma sayı aşağıdakı kimidir:

$$b_{5 j v}^i = \left| \{ w_{j k v}^i : w_{j k v}^i = \omega_v, N_{j k v}^{c_i} = |\omega_v|, v = 1, 2, \dots, N_{j k}^{w_i}, k = 1, 2, \dots, N_j^{s_i} \} \right|,$$

burada $|\omega_v| = \omega_v$ sözündəki hərflərin sayıdır, $v \in \{1, \dots, |W|\}$.

Hər hansı T_j^i mətnində $\omega_v \in W$ sözünün tezliyi aşağıdakı kimi hesablanır:

$$f_{5 j v}^i = \frac{b_{5 j v}^i}{\tilde{N}_j^i},$$

burada $\tilde{N}_j^i = \sum_{k=1}^{N_j^{s_i}} N_{j k}^{w_i} T_j^i$ mətnindəki sözlərin sayıdır, $N_{j k}^{w_i} T_j^i$ mətnindəki k nömrəli cümlədəki sözlərin sayıdır və $v \in \{1, \dots, |W|\}$, $j = 1, 2, \dots, l_i$, $i = 1, 2, \dots, L$.

Tədqiqatlarımızın nəticələri göstərir ki, ixtiyari sözün verilmiş mətndəki tezliyi ilə birlikdə müəllif namizədlərinin vahid mətnlərindəki (burada müəllif namizədinin vahid mətni dedikdə

həmin müəllif namizədinin öyrətmə bazasındakı mətnlərinin birləşməsindən alınan mətni nəzərdə tuturuq) tezlikləri də mətnin əlamətləri arasında istifadə olunduqda bu, tanıma nəticələrinə tanımada effektivlik baxımından yaxşı təsir edir. Verilmiş sözün müəllif namizədlərinin vahid mətnlərindəki tezlikləri ilə yanaşı həmin sözün öyrətmə bazasındakı mətnlərin birləşdirilməsi nəticəsində alınan mətnlərdəki tezlikləri də mətn əlamətləri arasında istifadə oluna bilər. İxtiyari sözün müəllif namizədlərinin vahid mətnlərindəki və öyrətmə bazasındakı mətnlərin birləşdirilməsindən alınan mətndəki tezliklərinin hesablanması qaydaları aşağıda verilmişdir.

Müəllif A^i -nin öyrətmə bazasındakı mətnləri çoxluğu aşağıdakı kimidir:

$$\bar{T}^i = T^i \cap \bar{T},$$

burada T^i – müəllif A^i -nin mətnləri çoxluğu, $\bar{T} \subset T$ – öyrətmə bazasındakı mətnlər çoxluğudur, $i = 1, 2, \dots, L$.

Hər hansı müəllif namizədi A^i -nin öyrətmə bazasındakı mətnlərinin birləşdirilməsindən alınan mətndə ixtiyari $\omega_v \in W$ sözünün tezliyi aşağıdakı kimidir:

$$\phi_v^{51i} = \frac{\sum_{j=1}^i b_{5jv}^i}{\sum_{j=1}^i \bar{N}_j^i}, \quad T_j^i \in \bar{T}^i,$$

burada $v \in \{1, \dots, |W|\}$.

Öyrətmə bazasındakı mətnlərin birləşdirilməsindən alınan mətndə ixtiyari $\omega_v \in W$ sözünün tezliyi aşağıdakı kimidir:

$$\phi_v^{52} = \frac{\sum_{i=1}^L \sum_{j=1}^i b_{5jv}^i}{\sum_{i=1}^L \sum_{j=1}^i \bar{N}_j^i}, \quad T_j^i \in \bar{T},$$

burada, $v \in \{1, \dots, |W|\}$.

Mətnlərə müəllifliyin tanınması məsələlərində mətnləri elə əlamətlərlə təsvir etmək lazım gəlir ki, müxtəlif müəlliflərin mətnləri bir-birindən fərqlənsin, yəni bu əlamətlər müəlliflər arasında diskriminativ xarakter daşsın. Bəzi əlamətlər digərlərinə nəzərən müəlliflər arasında daha diskriminativ xarakter daşıya bilər deyərək mümkün əlamətlər arasından daha diskriminativ olanlarını seçmək lazım gəlir. Müəyyən əlamət növü üzrə daha diskriminativ olanlarının seçilməsi üçün müəyyən əlamət seçimi prosedurlarından istifadə oluna bilər. Bu cür əlamət seçimi prosedurları, adətən, elə əlamət

növlərinə aid əlamətlərin seçimi üçün istifadə olunur ki, o növə aid əlamətlər istifadəçinin öz evristik bilikləri və ya kiçik miqyaslı manual (ağlabatan zamanda kompütersiz də aparıla bilən) analizləri əsasında seçim edə bilməyəcəyi qədər çoxdur. İşdə baxılan müəllif tanıma məsələsində sözlərin və simvol n-qramlarının seçilməsi üçün əlamət seçimi prosedurlarından istifadə olunmuşdur.

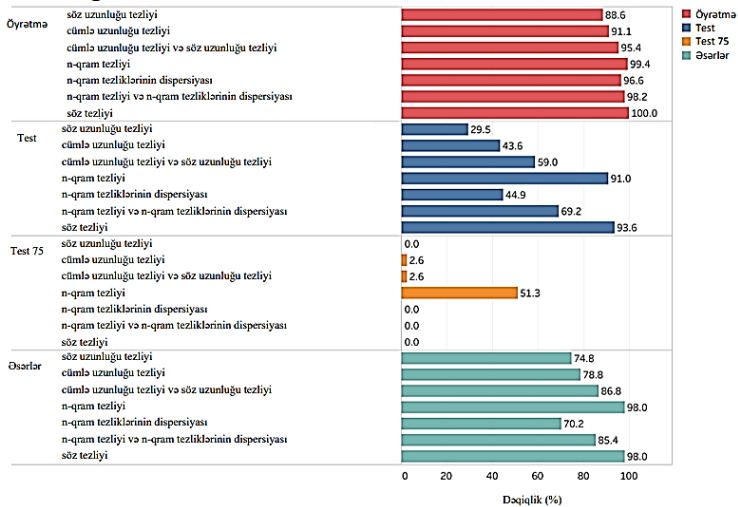
İşdə baxılan müəllif tanıma məsələsində müəllif tanımada mətn əlamətləri kimi tezlikləri istifadə olunacaq sözlərin seçilməsi üçün aşağıda təsvir olunmuş iki müxtəlif əlamət seçimi proseduru, simvol n-qramlarının seçilməsi üçün iki əlamət seçimi proseduru təklif edilmişdir (xatırladaq ki, bu əlamət seçimi prosedurları öyrətmə bazasındakı müəllifi məlum mətnlər əsasında həyata keçirilir). Sözlərin seçilməsi üçün istifadə olunan əlamət seçimi prosedurlarından birində (müəlliflərin tez-tez işlətdiyi sözlərin seçilməsi proseduru (birinci prosedur)) müəlliflərin hər birinin mətnlərində ən çox (ya da orta və ya ən az) istifadə etdiyi sözlər müəlliflər üzrə ayrı-ayrılıqda təyin olunur, ardınca hər müəllifin leksikonunda ən çox (ya da orta və ya ən az) işlətdiyi sözlərdən müəyyən qədər seçməklə müəlliflərin leksikonlarında əksəriyyət təşkil edən sözlərdən ibarət yığma sözlər çoxluğu yaradılır. Söz seçimi üçün istifadə olunan digər prosedurda (öyrətmə bazasında tez-tez istifadə olunan sözlərin seçilməsi proseduru (ikinci prosedur)) ayrı-ayrı müəlliflərin mətnlərinə ayrı-ayrılıqda yox, birlikdə baxılır, (müəllif fərqlərini nəzərə almadan) öyrətmə bazasındakı mətnlərdə ən çox (ya da orta və ya ən az) istifadə olunan sözlər seçilir. Simvol n-qramlarının seçilməsi üçün istifadə olunan əlamət seçimi prosedurlarından biri (öyrətmə bazasında tez-tez istifadə olunan simvol n-qramlarının seçilməsi proseduru (üçüncü prosedur)) sözlərin seçimi üçün istifadə olunan ikinci (sonuncu) prosedurun simvol n-qramları üzrə analoqudur: (müəllif fərqlərini nəzərə almadan) öyrətmə bazasındakı mətnlərdə ən çox (ya da orta və ya ən az) istifadə olunan simvol n-qramları seçilir. Simvol n-qramlarının seçilməsi üçün istifadə olunan digər əlamət seçimi prosedurunda (müəllifləri xarakterizə etmə dərəcələri fərqlənən simvol n-qramlarının seçilməsi proseduru (dördüncü prosedur)) isə verilmiş n üçün bütün simvol n-qramlarının baxılan məsələdəki müəllifləri xarakterizə etmə

dərəcələri (öyrətmə bazasındakı mətnlər əsasında) təyin olunur, ardınca müəllifləri xarakterizə etmə dərəcələri fərqlənən simvol n-qramları seçilir.

Bu əlamət seçimi prosedurlarının da istifadəsi ilə əlamət qruplarının tanımada effektivlikləri (tanımada effektivlik dedikdə müəyyən əlamət qrupu ilə istifadə olunan modelin və ya üsulun adekvatlığı nəzərdə tutulur) təhlil olunmuşdur.

2.3 paraqrafında baxılan müəllif tanıma məsələsində maşın öyrənməsi üsullarının mətn əlamətləri növlərinə aid əlamətlərdən ibarət müxtəlif əlamət qrupları ilə istifadələrinin tanımada effektivlikləri qiymətləndirilmişdir.

İşdə baxılan müəllif tanıma məsələsindəki müəlliflərin böyük və kiçik həcmli əsərlərindən ibarət dataset-0 üzərində ən yüksək tanıma dəqiqliyi alınan iki əlamət qrupunun (bu iki əlamət qrupundan birində simvol n-qramı tezliyi növünə aid əlamətlər mövcuddur, digərində söz tezliyi növünə aid əlamətlər vardır) şəkil 1-də verilmiş nəticələrindən aydın olur ki, bu iki əlamət qrupu ilə dataset-0 üzərində alınan tanıma dəqiqlikləri eyni olsa da, simvol n-qramları tezliklərindən ibarət əlamət qrupu ilə müəllif tanıma daha böyük inam dərəcəsi ilə aparıla bilər.



Şəkil 1. Müxtəlif əlamət növlərinə aid əlamətlərin istifadəsi ilə alınmış maksimal tanıma dəqiqlikləri

Üçüncü fəsildə azərbaycan dilindəki mətnlərin müəlliflərini tanıma proqram təminatının təsviri verilmişdir.

3.1 paraqrafında işlənmiş sistemin strukturu və sistemlə iş mərhələləri qısaca icmal olunmuşdur.

3.2 paraqrafında müəllifi məlum mətnlərin, eləcə də müəllifi tanınmalı (müəllifi naməlum hesab olunan) mətnlərin sistemə daxil edilməsi, sistemin öyrədilməsi, yəni burada sistemin müəllif tanıma üçün hazırlanması, tanıma və nəticələrinin nümayişi prosesləri təsvir olunmuşdur.

3.3 paraqrafında ayrı-ayrı müəllif tanıma üsullarının kollektiv istifadəsi üçün sistemdə realizə edilməsi mümkün kollektiv qərar qəbul etmə yanaşmaları təhlil olunmuşdur.

Əlavədə Tətbiq aktları və Azərbaycan dilindəki mətnlərin müəlliflik identifikasiyası üçün proqram təminatının əsas proqram modullarının kodlarının mətni verilmişdir.

Sonda ailəmə və İdarəetmə Sistemləri İnstitutunun kollektivinə sonsuz təşəkkürlər!

NƏTİCƏ

1. Azərbaycan dilindəki mətnlər üçün maşın öyrənməsindən istifadə edən müəllif tanıma üsulları işlənmişdir.

2. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınmasında istifadə üçün müxtəlif növlərə aid mətn əlamətlərin qiymətlərinin hesablanması prosedurları işlənmişdir.

3. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınmasında istifadə üçün effektiv mətn əlamətlərinin seçilməsi üçün prosedurlar təklif olunmuş, işlənmişdir.

4. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınmasında istifadə üçün müxtəlif növə aid mətn əlamətlərinin maşın öyrənməsinin müxtəlif üsulları ilə istifadəsi tədqiq olunmuşdur.

5. Azərbaycan dilindəki mətnlərin müəlliflərinin tanınmasına imkan verən müəllif tanıma kompüter sisteminin proqram təminatı işlənmişdir.

6. Süni neyron şəbəkələrin struktur və parametrik

identifikasiyası üçün istifadə olunmuş proqram modulları və alınmış nəticələr Azərbaycan Respublikasının Dövlət İmtahan Mərkəzindəki “Formaların tanıma sistemi” tətbiqi proqram paketinin təkmilləşdirilməsində istifadə olunmuşdur.

Dissertasiyanın mövzusunə dair dərc olunmuş elmi işlərin siyahısı

1. Мустафаев, Э.Э., Ахмедлы, Н.А., Азимов, Р.Б. Сравнительный анализ применения нейронных и сверточных нейронных сетей распознавания рукопечатных букв Азербайджанского алфавита // Информационный бюллетень Омского научно-образовательного центра ОмГТУ и ИМ СО РАН в области математики и информатики, – Омск, Россия:– 22-29 апреля 2021, т.5, № 1, – с.94-95. **(РИИЦ)**
2. Mustafayev, E.E., Azimov, R.B. Comparative analysis of the application of multilayer and convolutional neural networks for recognition of handwritten letters of the Azerbaijani alphabet // – Kiev, Ukraine: Cybernetics and Computer Technologies, – 2021. №3, – pp. 65–73. **(jurnal Ukrayna AAK siyahısına daxildir)** <https://doi.org/10.34229/2707-451X.21.3.6>.
3. Əzimov, R.B. Azərbaycan dilinin çap əlyazma hərflərinin tanınmasında bəzi əlamətlərin müqayisəsi // “Riyaziyyatın tətbiqi məsələləri və yeni informasiya texnologiyaları” adlı IV Respublika elmi konfransı, – Sumqayıt, Azərbaycan: 9-10 dekabr 2021, № 9, s.79-84. **(РИИЦ)**
4. Мустафаев, Э.Э., Азимов, Р.Б. Использование многослойных и сверточных нейронных сетей для распознавания рукопечатных букв на примере азербайджанского алфавита // – Омск, Россия: Прикладная математика и фундаментальная информатика, – 2022. т.8, №2, – с. 38-45. **(РИИЦ)**
5. Azimov, R.B., Mustafayev, E.M. Comparison of SVM and ANN methods for recognition of authorship of texts // Applied

Mathematics and Fundamental Informatics: Proceedings of the XII Intern. youth scientific-practical. conf. with elements of science. schools, – Омск: 16-21 may 2022, – pp. 60-61.

(PИИЦ)

6. Əzimov, R.B. Azərbaycan çapəlyazma əlifbasının tanınmasına yanaşmaların müqayisəli təhlili // Tələbə və Gənc Tədqiqatçıların III Beynəlxalq Elmi Konfranslarının “Proseslərin avtomatlaşdırılması və informasiya təhlükəsizliyi-2022” konfransı, – Bakı, Azərbaycan: – 26-27 aprel 2022, – s. 157-159.
7. Aida-zade, K.R., Mustafayev, E.M., Azimov, R.B. Features analysis for application in a computer recognition systems of Azerbaijani texts authorship // Second International Bilateral Workshop on Science Between Dokuz Eylül University and Azerbaijan National Academy of Sciences, – İzmir, Türkiyə: – 18 november 2022, – p. 11.
8. Əzimov, R.B., Mustafayev, E.M. Azərbaycan dilindəki mətnlərin müəlliflərini tanıyan kompüter sistemində istifadə üçün müxtəlif əlamət qruplarının müqayisəli təhlili // “Riyaziyyatın fundamental problemləri və intellektual texnologiyaların təhsildə tətbiqi” mövzusunda II Respublika elmi konfransı, – Sumqayıt, Azərbaycan: – 15-16 Dekabr 2022, s. 34-39.
9. Azimov, R.B., Aida-zade, K.R. Analysis of features of texts for use in an author recognition system // ICT problems of the Azerbaijani language, Azerbaijani language problems of the ICT, – Bakı, Azerbaijan: – 21-22 Fevral 2023, s. 22-25.
10. Əzimov, R.B. Müəllif Tanıma Sistemində Mətn Əlamətləri Siniflərinin İstifadəsinin Təhlili // Tələbə və Gənc Tədqiqatçıların IV Beynəlxalq Elmi Konfranslarının “Rəqəmsal Transformasiya-2023” konfransı, – Bakı, Azerbaijan: – 18-19 aprel 2023, s. 286-288.
11. Azimov, R.B. Analysis of use of text feature classes in an author recognition system // Applied Mathematics and Fundamental Informatics: Proceedings of the XII Intern. youth scientific-practical. conf. with elements of science, – Omsk, Russia: – 15-

- 20 may 2023, s. 88-89. **(РИИЦ)**
12. Azimov, R.B. Approaches to the recognition of handwritten letters of the Azerbaijani language and their analysis // Informatics and Control Problems, – 2023, v. 43, no. 3, – pp. 32-40. <https://doi.org/10.54381/icp.2023.1.05>.
 13. Azimov, R.B. Analysis of the Use of Methods and Feature Groups for Author Recognition on the Example of Texts in the Azerbaijani Language // Abstracts of V International Conference on “Problems of Cybernetics and Informatics” (PCI 2023), – Baku, Azerbaijan: – p. 90.
 14. Khalilov, C.J., Mustafayev, E.E., Mahmudov, I.M., Azimov, R.B. Computer System of Analysis of the Mass Exam Results // Abstracts of V International Conference on “Problems of Cybernetics and Informatics” (PCI 2023), – Baku, Azerbaijan: – pp. 29-30.
 15. Azimov, R.B. Analysis of the Use of Methods and Feature Groups for Author Recognition on the Example of Texts in the Azerbaijani Language // 2023 5th International Conference on Problems of Cybernetics and Informatics (PCI), – Baku, Azerbaijan: – 2023, – pp. 1-4. **(Scopus)** <https://doi.org/10.1109/PCI60110.2023.10325956>.
 16. Khalilov, C.J., Mustafayev, E.E., Mahmudov, I.M., Azimov, R.B. Computer System of Analysis of the Mass Exam Results // 2023 5th International Conference on Problems of Cybernetics and Informatics (PCI), – Baku, Azerbaijan: – 2023, – pp. 1-4. **(Scopus)** <https://doi.org/10.1109/PCI60110.2023.10326014>.
 17. Azimov, R.B. Comparison of artificial and convolutional neural networks in recognition of handwritten letters of Azerbaijani alphabet // Recent Developments and the New Directions of Research, Foundations, and Applications. Studies in Fuzziness and Soft Computing (8th World Conference of Soft Computing), – Baku, Azerbaijan: Springer, – 3-5 February, – 2022, – v. 422, pp. 377-385. **(Scopus)** https://doi.org/10.1007/978-3-031-20153-0_31.
 18. Mustafayev, E.E., Azimov, R.B. Computer System of

- Evaluation of the Mass Exam Results Based on Recognition of Handprinted Azerbaijani Characters // Proceedings of the Information Technologies and Its Applications Conference, – Baku, Azerbaijan: Springer, – 23-25 April, – 2024, – v. 2, pp. 171-183. **(Scopus)**
19. Azimov, R.B., Efthimios, P. A Comparative Study of Machine Learning Methods and Text Features for Text Authorship Recognition in the Example of Azerbaijani Language Texts // Algorithms, – 2024. V. 17, no. 6, – 242 (25 pages). <https://doi.org/10.3390/a17060242>. **(Web of Science, ESCI)**
 20. Aida-zade, K.R., Azimov, R.B.. Analysis of the use of text features in the authorship identification of literary works in the Azerbaijani language // Informatics and Control Problems, – Baku, Azerbaijan: – 2024, v. 44, no. 1, – pp. 51-58.
 21. Azimov, R.B. Comparative Analysis of using Different Text Features, Models, and Methods in Text Author Recognition // Cybernetics and Systems Analysis, – Kiev, Ukraine: – 2024. V. 60, no. 5, – pp. 711-725. <https://doi.org/10.1007/s10559-024-00709-z>. **(Web of Science, ESCI)**

Dissertasiyanın müdafiəsi 6 dekabr 2024-cü il tarixində saat 14:00-da Azərbaycan Respublikası Elm və Təhsil Nazirliyinin İdarəetmə Sistemləri İnstitutunun nəzdində fəaliyyət göstərən ED 1.20 Dissertasiya şurasının iclasında keçiriləcək.

Ünvan: AZ1141, Bakı şəhəri, B.Vahabzadə küç., 68

Dissertasiya işi ilə Azərbaycan Respublikası Elm və Təhsil Nazirliyinin İdarəetmə Sistemləri İnstitutunun kitabxanasında tanış olmaq mümkündür.

Dissertasiya və avtoreferatın elektron versiyaları Azərbaycan Respublikası Elm və Təhsil Nazirliyinin İdarəetmə Sistemləri İnstitutunun rəsmi internet saytında yerləşdirilmişdir.

Avtoreferat 4 noyabr 2024-cü il tarixində zəruri ünvanlara göndərilmişdir.

Çapa imzalanıb: 01.11.2024

Kağızın formatı: A5

Həcm: 37669

Tiraj: 100