

Rəqəmsal Kriminalistikada Mətn Tipli Sübutların Analizi Üçün Təbii Dil Emalı Metodlarının Tətbiqi: Azərbaycan Dili Üzrə Yanaşma

Mehriban Əsədullasoy

Naxçıvan Dövlət Universiteti, Naxçıvan, Azərbaycan
quliyeva.mehriban@list.ru

Xülasə— Bu tədqiqatda rəqəmsal kriminalistika sahəsində mətn tipli rəqəmsal sübutların təhlilində təbii dil emalı (NLP) metodlarının tətbiqi məsələləri araşdırılır. Kiberməkanda baş verən hüquqpozuntular zamanı yazılı kommunikasiya vasitələrinin geniş yayılması bu məlumatların sistemli analizini zəruri edir. Məqalədə Azərbaycan dili üzrə mövcud resurs çatışmazlığı, terminoloji uyğunsuzluqlar və alqoritmik uyğunlaşdırma problemləri təhlil olunur. Bununla yanaşı, mətnlərin semantik emalı, müəlliflik xüsusiyyətlərinin müəyyənəndirilməsi və riskli məzmunun aşkarlanması kimi istiqamətlər üzrə NLP metodlarının tətbiq imkanları göstərilir və praktiki tövsiyələr təqdim olunur. Tədqiqat nəticələri göstərir ki, Azərbaycan dili üçün uyğunlaşdırılmış NLP modellərinin inkişafı rəqəmsal sübutların daha effektiv analizinə şərait yarada bilər. Bu yanaşma kibercinayətlərin aşkarlanması prosesində sübutların interpretasiyasının dəqiqliyini artırmağa xidmət edir.

Açar sözlər— rəqəmsal kriminalistika; təbii dil emalı; mətn analizi; rəqəmsal sübutlar; Azərbaycan dili; NLP metodları; kibercinayətkarlıq.

I. GİRİŞ

Müasir dövrdə rəqəmsal texnologiyaların geniş yayılması kibercinayətkarlıq sahəsində yeni çağırışların yaranmasına səbəb olmuşdur. İnformasiya mübadiləsinin əsas hissəsinin mətn əsaslı rəqəmsal mühitlər vasitəsilə həyata keçirilməsi bu məlumatların analitik emalını rəqəmsal kriminalistikanın mühüm istiqamətlərindən birinə çevirmişdir. Elektron yazışmalar, sosial şəbəkə mesajları və digər mətn resursları cinayət hadisələrinin aşkarlanması və rəqəmsal sübutların toplanmasında əhəmiyyətli rol oynayır [1].

Bu kontekstdə təbii dil emalı (Natural Language Processing – NLP) texnologiyaları mətnlərin avtomatik təhlili, semantik məzmunun müəyyən edilməsi və müəlliflik xüsusiyyətlərinin aşkarlanması kimi imkanlar təqdim edir [2]. Bununla belə, Azərbaycan dili üzrə bu texnologiyaların tətbiqi bir sıra metodoloji və texnoloji məhdudiyyətlərlə müşayiət olunur.

Bu məqalənin məqsədi rəqəmsal kriminalistikada mətn tipli sübutların təhlilində NLP metodlarının tətbiqi imkanlarını araşdırmaq, Azərbaycan dili kontekstində mövcud çətinlikləri müəyyənəndirmək və inkişaf perspektivlərini təhlil etməkdir. Rəqəmsal kriminalistika müasir informasiya cəmiyyətində kibercinayətlərin aşkarlanması, sübutların toplanması və təhlili prosesində mühüm rol oynayır. Bu sahədə aparılan tədqiqatların

əsas hissəsi rəqəmsal mühitdə formalaşan mətn tipli məlumatların emalına yönəlmişdir. Elektron yazışmalar, sosial media platformalarında paylaşılan mesajlar və digər mətn resursları cinayət hadisələrinin rekonstruksiyası üçün əsas sübut bazasını təşkil edir. Bu amillər rəqəmsal kriminalistika sahəsində mətnlərin avtomatik və dəqiq analizini çətinləşdirir. Bu səbəbdən, Azərbaycan dili üçün uyğunlaşdırılmış NLP metodlarının hazırlanması və tətbiqi həm elmi, həm də praktiki baxımdan aktuallıq kəsb edir [3].

II. NƏZƏRİ ƏSASLAR VƏ TƏDQIQAT YANAŞMASI

Rəqəmsal kriminalistika informasiya texnologiyalarının inkişafı ilə paralel olaraq formalaşan multidissiplinar sahədir və kompüter sistemlərində saxlanılan, eləcə də ötürülən rəqəmsal sübutların araşdırılmasını əhatə edir. Bu sahənin əsas məqsədi kibercinayətlərin aşkarlanması və sübut bazasının elmi əsaslarla təhlil edilməsidir. Mətn tipli rəqəmsal sübutlar rəqəmsal kriminalistikanın ən geniş yayılmış məlumat formalarından biridir. Bu sübutlar strukturlaşdırılmamış mətn formatında olduğuna görə onların avtomatik analizi kompüter dilçiliyi və təbii dil emalı (NLP) metodlarına əsaslanır. NLP yanaşmaları mətnlərin sintaktik və semantik təhlilini həyata keçirməyə imkan verir və nəticədə informasiya daxilində gizli əlaqələrin aşkarlanmasını asanlaşdırır. Azərbaycan dili kimi morfoloji baxımdan zəngin dillərdə NLP təbiiqləri əlavə çətinliklər yaradır. Söz formalarının çoxvariantlılığı, standartlaşdırılmış dil korpuslarının məhdudluğu və mövcud alqoritmlərin uyğunlaşdırılmaması bu sahədə əsas tədqiqat problemlərini təşkil edir [3].

NLP-nin nəzəri əsasları üç əsas yanaşmaya söykənir: qaydaya əsaslanan (rule-based), statistik və dərin öyrənmə (deep learning) modelləri. “Qaydaya əsaslanan metodlar dilin qrammatik və sintaktik qaydaları üzərində qurularaq mətnlərin emalını həyata keçirir, lakin mürəkkəb kontekstlərdə məhdudiyyətlər göstərir”. “Statistik yanaşmalar böyük mətn korpusları əsasında ehtimal modellərinə əsaslanaraq daha çevik nəticələr verir” [2]. Müasir mərhələdə isə “dərin öyrənmə modelləri mətnlərdə semantik əlaqələri avtomatik öyrənmək və daha dəqiq nəticələr əldə etmək imkanı yaradır” [4].

III. TƏBİİ DİLİN EMALI METODLARININ RƏQƏMSAL KRİMINALİSTİKADA TƏTBİQİ

Rəqəmsal kriminalistikada təbii dil emalı (Natural Language Processing – NLP) metodları mətn tipli rəqəmsal sübutların

avtomatik və yarımavtomatik təhlilində geniş tətbiq potensialına malikdir. Bu yanaşma xüsusilə strukturlaşdırılmamış mətnlərin – elektron yazışmaların, sosial media mesajlarının və onlayn forum yazılarının – emalı üçün effektiv hesab olunur. Belə məlumatlar çox zaman böyük həcmdə və qeyri-formal dildə olduğu üçün onların əl ilə təhlili həm vaxt aparan, həm də subyektiv nəticələrə səbəb ola bilən prosesdir [4].

NLP metodları bu problemlərin aradan qaldırılmasına imkan yaradır və mətnlərin əvvəlcədən emalı (tokenizasiya, lemmatizasiya), semantik analiz, sentiment analizi və açar sözlərin çıxarılması kimi mərhələləri əhatə edir. Bu mərhələlər vasitəsilə mətn daxilində gizli mənalara, niyyətlər və kommunikativ əlaqələr müəyyən edilə bilər [4].

Rəqəmsal kriminalistika kontekstində NLP-nin mühüm tətbiq sahələrindən biri müəllif identifikasiyasıdır. Bu yanaşma mətnin leksik və sintaktik xüsusiyyətlərinə əsaslanaraq yazının kim tərəfindən yazıldığını ehtimal etməyə imkan verir. Digər mühüm istiqamət isə riskli məzmunun aşkarlanmasıdır ki, burada sistemlər təhdid, şantaj və ya qanunsuz fəaliyyətə dair ifadələri avtomatik müəyyən edə bilər.

NLP əsaslı sistemlərdə istifadə olunan maşın öyrənməsi alqoritmləri zamanla öz dəqiqliyini artırmaq qabiliyyətinə malikdir. Bu isə sistemlərin yeni tipli cinayət davranışlarına uyğunlaşmasını təmin edir. Kontekstual analiz metodları mətnin yalnız söz səviyyəsində deyil, bütöv məna səviyyəsində qiymətləndirilməsinə imkan yaradır. Eyni zamanda, çoxdilli mətnlərin emalı beynəlxalq kibercinayətlərin araşdırılmasında xüsusi əhəmiyyət daşıyır. Avtomatik filtrasıya sistemləri böyük həcmli məlumat içərisindən aidiyyəti olan sübutların seçilməsini asanlaşdırır. NLP texnologiyalarının tətbiqi sübutların emalında insan faktorundan qaynaqlanan səhvləri minimuma endirir. Bu yanaşma hüquqi prosedurların daha obyektiv və elmi əsaslarla aparılmasına şərait yaradır. Nəticə etibarilə, NLP rəqəmsal kriminalistika sistemlərinin funksional imkanlarını əhəmiyyətli dərəcədə genişləndirir [5].

IV. AZƏRBAYCAN DİLİNDƏ NLP TƏTBİQİNİN ÇƏTİNLİKLƏRİ

Azərbaycan dili morfoloji baxımdan zəngin və aqqlütinativ bir dil olduğu üçün NLP sistemlərinin bu dilə uyğunlaşdırılması əlavə çətinliklər yaradır. Söz köklərinə çoxsaylı şəkilçilərin əlavə olunması nəticəsində eyni məna müxtəlif formalarda ifadə oluna bilər ki, bu da avtomatik analiz prosesini mürəkkəbləşdirir. Bundan əlavə, Azərbaycan dili üçün geniş və standartlaşdırılmış korpusların məhdudluğu NLP modellərinin öyrədilməsini çətinləşdirir. Mövcud resursların azlığı nəticəsində modellərin dəqiqliyi aşağı düşür və nəticələrin etibarlılığı azalır. Terminoloji baza çatışmazlığı rəqəmsal kriminalistika sahəsində vahid terminologiyanın formalaşmasını məhdudlaşdırır. Bu problemlər Azərbaycan dili üçün xüsusi optimallaşdırılmış NLP modellərinin yaradılmasını və lokal dil resurslarının inkişaf etdirilməsini zəruri edir [6].

Bu çətinliklər təkcə texniki deyil, həm də metodoloji xarakter daşıyır və kompleks yanaşma tələb edir. Xüsusilə morfoloji zənginlik səbəbindən lemmatizasiya və söz formalarının normallaşdırılması daha mürəkkəb mərhələyə çevrilir. Korpus çatışmazlığı isə maşın öyrənməsi modellərinin ümumiləşdirmə qabiliyyətini zəiflədir. Nəticədə, sistemlər real mətnlər üzərində tətbiq olunduqda səhv interpretasiya riski artır.

Bu vəziyyət kriminalistik analizlərin etibarlılığına birbaşa təsir göstərir. Bu səbəbdən adaptiv və Azərbaycan dilinə uyğunlaşdırılmış modellərin hazırlanması elmi-praktiki zərurət kimi çıxış edir. Belə yanaşma həm dil resurslarının inkişafına, həm də rəqəmsal kriminalistika sistemlərinin dəqiqliyinin artırılmasına xidmət edir [6].

V. PRAKTİKİ TƏTBİQLƏR

Rəqəmsal kriminalistikada təbii dil emalı metodlarının praktik tətbiqi əsasən real rəqəmsal sübutların analizi üzərində qurulur. Bu sübutlar çox vaxt strukturlaşdırılmamış mətn formasında olduğu üçün onların avtomatik emalı cinayət hadisələrinin araşdırılmasında operativlik və obyektivliyi təmin edir. Xüsusilə mesajlaşma platformalarından, e-poçt sistemlərindən və sosial şəbəkələrdən əldə edilən mətn məlumatları kriminal araşdırmaların əsas informasiya mənbələrindən biridir [5].

Praktik analiz prosesində NLP əsaslı sistemlər mətnləri ilkin emal edərək sözlərin normallaşdırılması, cümlə strukturunun parçalanması və semantik əlaqələrin müəyyən edilməsi mərhələlərini həyata keçirir. Bu mərhələlər nəticəsində mətn daxilində gizli niyyətlər, təhdid xarakterli ifadələr və kommunikativ davranış modelləri aşkarlana bilər. Məsələn, şübhəli şəxslərin yazışmalarında istifadə olunan leksik seçimlər onların emosional vəziyyəti və niyyətləri haqqında dolaylı məlumat əldə etməyə imkan yarada bilər. Bundan əlavə, klasterləşdirmə və təsnifat alqoritmləri vasitəsilə oxşar məzmunlu mesajların qruplaşdırılması mümkündür ki, bu da böyük həcmli məlumatların təhlilini xeyli asanlaşdırır. Bu yanaşma xüsusilə kibercinayət şəbəkələrinin aşkarlanmasında effektiv hesab olunur [5].

VI. TƏKMİLLƏŞDİRMƏ YOLLARI VƏ PERSPEKTİVLƏR

Azərbaycan dili üçün NLP əsaslı rəqəmsal kriminalistika sistemlərinin inkişafı bir sıra strateji istiqamətlərin həyata keçirilməsini tələb edir. İlk növbədə, geniş həcmli və keyfiyyətli dil korpuslarının yaradılması vacibdir. Bu korpuslar müxtəlif üslubları, terminologiyaları və real kommunikasiya nümunələrini əhatə etməlidir. İkinci mühüm istiqamət terminoloji standartlaşdırmaqdır. Rəqəmsal kriminalistika sahəsində vahid anlayış sisteminin formalaşdırılması həm elmi tədqiqatların, həm də praktik tətbiqlərin effektivliyini artırır. Eyni zamanda, Azərbaycan dilinin morfoloji xüsusiyyətlərini nəzərə alan xüsusi NLP modellərinin hazırlanması zəruridir [7].

Gələcək perspektivlər arasında dərin öyrənmə (deep learning) əsaslı modellərin tətbiqi, avtomatik sübut analizi sistemlərinin yaradılması və hüquq-mühafizə orqanları üçün intellektual qərar dəstək sistemlərinin inkişafı xüsusi əhəmiyyət kəsb edir. Bu istiqamətlər rəqəmsal kriminalistikanın daha dəqiq, sürətli və etibarlı işləməsinə şərait yarada bilər.

Bununla yanaşı, NLP sistemlərinin inkişafında açıq mənbəli alətlərin və platformaların istifadəsi tədqiqatların sürətini artıran mühüm amillərdən biridir. Azərbaycan dili üçün xüsusi optimallaşdırılmış modellərin hazırlanması beynəlxalq təcrübə ilə inteqrasiyanı da asanlaşdırır. Eyni zamanda, annotasiya olunmuş məlumat bazalarının yaradılması maşın öyrənməsi modellərinin dəqiqliyini yüksəldir. Multidisiplinar yanaşma dilçilik, informatika və kriminalistika sahələrinin birgə

əməkdaşlığını zəruri edir. Bu əməkdaşlıq nəticəsində daha kompleks və effektiv analiz sistemləri formalaşdırıla bilər. Praktik tətbiqlərdə real-vaxt rejimində işləyən süni intellekt sistemləri xüsusi əhəmiyyət daşıyır. Bu sistemlər hüquq-mühafizə orqanlarının operativ qərarvermə imkanlarını genişləndirir. Nəticə etibarilə, Azərbaycan dili üzrə NLP əsaslı kriminalistika sistemlərinin inkişafı milli təhlükəsizlik baxımından strateji əhəmiyyət kəsb edir [7].

Bu istiqamətdə ilk növbədə milli dil korpuslarının yaradılması üçün real mətn mənbələrinin sistemli şəkildə toplanması zəruridir. Eyni zamanda, sosial şəbəkələr, elektron yazışmalar və rəqəmsal platformalardan əldə olunan məlumatların strukturlaşdırılması vacib addımlardan biridir. Terminoloji standartlaşdırma üçün hüquq, kriminalistika və informatika sahələri üzrə vahid termin bazasının formalaşdırılması həyata keçirilməlidir. Azərbaycan dili üçün xüsusi leksik və morfoloji analiz modullarının hazırlanması NLP sistemlərinin effektivliyini artıracaqdır. Açıq mənbəli NLP kitabxanalarının yerli dil xüsusiyyətlərinə uyğunlaşdırılması praktik tətbiqləri sürətləndirə bilər. Maşın öyrənməsi modellərinin təlimi üçün annotasiya olunmuş korpusların hazırlanması prioritet istiqamət olmalıdır. Real-vaxtda işləyən analitik sistemlərin prototiplərinin hazırlanması hüquq-mühafizə orqanlarının operativliyini yüksəldəcəkdir. Bu tədbirlərin kompleks şəkildə həyata keçirilməsi Azərbaycan dili üzrə rəqəmsal kriminalistika ekosisteminin formalaşmasına imkan yaradacaqdır [7].

NƏTİCƏ

Rəqəmsal texnologiyaların cəmiyyətin bütün sahələrinə inteqrasiyası informasiya mühitinin yeni strukturunu formalaşdırmışdır. İnternet və elektron informasiya daşıyıcıları artıq gündəlik həyatın ayrılmaz hissəsinə çevrilmiş, sosial və hüquqi münasibətlərə birbaşa təsir göstərmişdir. Bu kontekstdə qeyd olunur ki: "İnternet və elektron informasiya daşıyıcıları cəmiyyətin və hər bir insanın fərdi həyatının bir hissəsinə çevrilmişdir. İndi həyatı internet və rəqəmsal texnologiyalardan istifadə etmədən təsəvvür etmək mümkün deyil. Beynəlxalq cinayətlərin, o cümlədən qloballaşma prosesləri ilə əlaqəli transmilli mütəşəkkil cinayətlərin artması tendensiyasından çıxış edərək demək olar ki, milli və beynəlxalq səviyyədə cinayətlərin hamısı İnternetdən və müasir elektron avadanlıqlardan istifadə etməklə törədilir.

Tədqiqat göstərir ki, NLP metodları rəqəmsal kriminalistikada mətn tipli sübutların analizində effektiv vasitədir. Bu metodlar semantik analiz, müəllif identifikasiyası və riskli məzmunun aşkarlanmasında yüksək potensiala malikdir. Azərbaycan dili üçün əsas məhdudiyyətlər korpus çatışmazlığı və morfoloji mürəkkəblikdir. Bu problemlərin

aradan qaldırılması üçün xüsusi adaptiv modellərin hazırlanması vacibdir. Nəticə olaraq, NLP texnologiyalarının inkişafı rəqəmsal kriminalistika analizlərinin dəqiqliyini artıraraq hüquq-mühafizə sistemlərinin effektivliyini yüksəldir.

ƏDƏBİYYAT

- [1] "Azərbaycanda kibertəhlükəsizlik hesabatı," Elektron Təhlükəsizlik Xidməti, Bakı, Azərbaycan, Texniki Hesabat, 2022.
- [2] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. Sebastopol, CA, USA: O'Reilly Media, 2009.
- [3] M. Əsədullayev, Elmi innovasiyalar. Bakı, Azərbaycan: Elm nəşriyyatı, 2025.
- [4] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. draft, 2023.
- [5] M. Əsgərova, Cinayət hüququ problemləri. Bakı, Azərbaycan: Qanun nəşriyyatı, 2025.
- [6] S. Məmmədova, Rəqəmsal sosiologiya. Bakı, Azərbaycan: Elm nəşriyyatı, 2025.
- [7] M. Yusifov, Dövlətçilik və dil. Bakı, Azərbaycan: Nurlan nəşriyyatı, 2004.

Application of Natural Language Processing Methods for Analysis of Text-Type Evidence in Digital Forensics: an Approach to the Azerbaijan Language

Mehriban Asadullayev

Nakhchivan State University, Nakhchivan, Azerbaijan
quliyeva.mehriban@list.ru

Abstract - This study examines the application of natural language processing (NLP) methods in the analysis of text-type digital evidence in the field of digital forensics. The widespread use of written communication tools in cyberspace makes systematic analysis of this information necessary. The article analyzes the current resource shortage, terminological inconsistencies, and algorithmic adaptation problems in the Azerbaijani language. In addition, the application possibilities of NLP methods in such areas as semantic processing of texts, identification of authorship features, and detection of risky content are shown and practical recommendations are provided. The research results show that the development of NLP models adapted for the Azerbaijani language can create conditions for more effective analysis of digital evidence. This approach serves to increase the accuracy of interpretation of evidence in the process of detecting cybercrimes.

Keywords - digital forensics; natural language processing; text analysis; cybercrime; Azerbaijani language; NLP methods; digital evidence.