UOT 04:37 DOI: 10.25045/jpis.v10.i1.10

Ocaqverdiyeva S.S.
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
allahverdiyevasabira@gmail.com

VERİLƏNLƏRİN SANİTARİZASİYASININ BƏZİ AKTUAL PROBLEMLƏRİ HAQQINDA

Məqalədə verilənlərin sanitarizasiyasının (təmizlənməsinin), məqsəd və vəzifələri, əhatə dairəsi, tətbiq sahələri, perspektivləri araşdırılmışdır. Verilənlərin sanitarizasiyasının texnologiyaları, mövcud üsulları, bəzi elmi-nəzəri və aktual problemləri analiz olunmuşdur. Bu problemin həlli ilə bağlı görülən işlər təhlil olunmuş, müvafiq tövsiyə və təkliflər irəli sürülmüşdür.

Açar sözlər: *verilənlərin sanitarizasiyası, verilənlərin təmizlənməsi, konfidensial, fərdi məlumatlar.*

Giriş

İnternet sürətlə inkişaf edərək cəmiyyətin bütün sahələrinə hərtərəfli və dərin nüfuz edir, ondan həm fərdlər, həm də təşkilatlar müxtəlif məqsədlərlə istifadə edirlər. Müasir dövrdə dövlət idarəçiliyi, biznes, şəxsi məlumatlar, elm-təhsil, səhiyyə sistemi və digər sahələr ənənəvi mühitdən elektron mühitə inteqrasiya olunur. İnformasiya texnologiyaları və vasitələrinin inkişafı informasiya əldə etmək imkanını asanlaşdırsa da, informasiya təhlükəsizliyi baxımından ciddi təhdidlərə yol açır. İnternetin imkanlarının geniş olması və virtual mühitdə informasiyanın həddən artıq çox olması təhlükəsizliklə bağlı məsələlərin həllini bir qədər çətinləşdirir.

E-dövlətin qurulmasında təhlükəsizlik məsələləri həlli vacib olan problemlərdən biridir [1]. Bununla bağlı Azərbaycanda bir sıra qanunlar, dövlət proqramları qəbul edilmişdir və bu proses davam etməkdədir. "Azərbaycan Respublikasında informasiya cəmiyyətinin inkişafına dair 2014-2020-ci illər üçün Milli Strategiya"da elektron mühitdə təhlükəsizlik məsələlərinə xüsusi yer ayrılır. Həmin sənəddə uşaqlarin qanunazidd və təhlükəli kontentdən qorunması üçün "təhlükəsiz İnternet" mexanizminin işlənilməsi və tətbiqi" informasiya təhlükəsizliyinin təmin edilməsinin əsas hədəflərdən biri kimi qeyd olunur [2]. 2018-ci ildə "Uşaqların zərərli informasiyadan qorunması haqqında" Azərbaycan Respublikasının Qanunu qəbul edilmişdir [3]. Bu qanun uşaqların yaşına uyğun informasiya əldə etmək hüququnun həyata keçirilməsi ilə əlaqədar olaraq onların zərərli informasiyadan qorunması tədbirlərini müəyyən edir və bu sahədə yaranan münasibətləri tənzimləyir.

Məlumdur ki, İnternet resurslarına böyük sürətlə müxtəlif məzmunlu resurslar daxil edilir. İstifadəçilər, xüsusən, uşaqlar zorakılığı, qəddarlığı, cinayətkarlığı, uşaqların həyat və təhlükəsizliyinə potensial təhlükə yaradan əməlləri təbliğ edən məlumatlar, əxlaqdan kənar və qeyri-məişət leksika (söyüş, jarqon) və qorxu məzmunlu informasiya, erotik və pornoqrafik xarakterli informasiya və s. məlumatlarla qarşılaşırlar. İstər İnternet mühitində, istərsə də televiziya məkanında uşaqların ziyanlı kontentlə qarşılaşması isə bir çox neqativ hallara – onların təhlükəsizliyinin, psixologiyasının və eyni zamanda, informasiya ekologiyasının pozulmasına səbəb olur. Hal-hazırda mütəxəssisləri ən çox narahat edən məsələlərdən biri verilənlərin onlayn mühitdə təhlükəsizliyinin təmin edilməsidir. Belə yanaşmalardan biri də zərərli kontentin təmizlənməsi üçün verilənlərin sanitarizasiyası texnologiyasının tətbiqidir.

Sanitarizasiya həssas məlumatların (mətn, audio, video və s.) təmizlənməsi prosesidir [4]. Burada əsas məqsəd informasiyanın gizli saxlanması və məlumatın sanitarizasiya olunduqdan sonra hədəf auditoriyasına təqdim olunmasıdır. Verilənlərin sanitarizasiyasının məqsədi həssas, fərdi, konfidensial, tövsiyə olunmayan və s. məlumatları təhlükələrdən qorumaqla yanaşı, onların geniş ictimaiyyətə çatdırılmasının qarşısının alınmasından və ya müəyyən məhdudiyyətlərin tətbiq

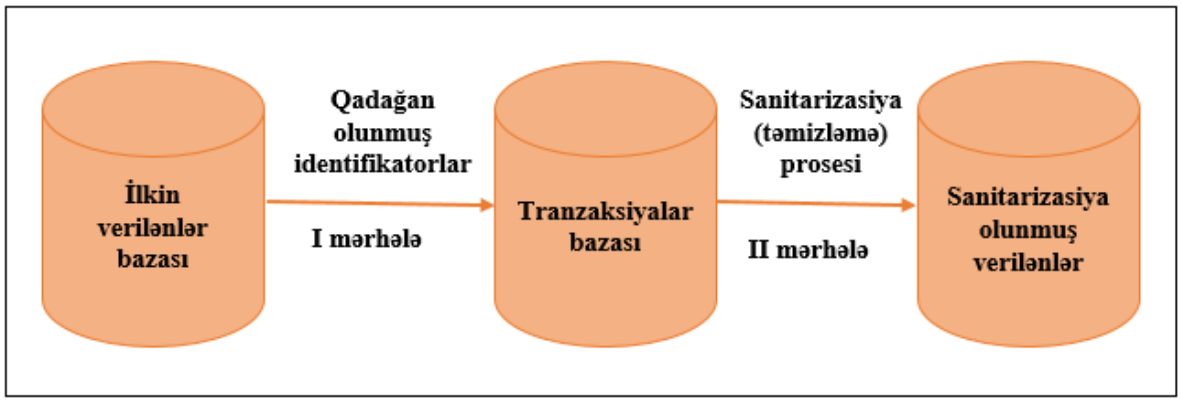
olunmasından ibarətdir [4, 5]. Məqalədə verilənlərin sanitarizasiyasının məqsədi, vəzifələri, təmizləmə üsulları şərh olunur və bu sahədə mövcud problemlərlə bağlı təkliflər irəli sürülür.

Verilənlərin sanitarizasiyasının məqsəd və vəzifələri

Qədim zamanlardan başlayaraq insanlar, adətən, məlumatın açıqlanmasının qarşısını almaq, yaxud məlumatı gizli saxlamaq məqsədilə onu ya maskalayır, ya da müxtəlif yollarla gizliliyini təmin edirdilər. Bununla da, məlumatı qorumağa nail olurdular. Lakin texnologiyaların yüksək inkişaf etdiyi müasir dövrdə elektron məlumatların konfidensiallığının tam qorunmasına zəmanət vermək çətindir. Məlumatların kənar şəxslər tərəfindən sui-istifadəsinin qarşısını almaq üçün sənədlər üzərində müxtəlif dəyişikliklərin edilməsi tələb olunur [6]. Bu məqsədlə məlumatlar üzərində sanitarizasiya işləri aparmaq daha əhəmiyyətlidir. Məlumatların təmizlənməsi zamanı ilkin olaraq metaverilənlərin sənəd daxilindən silinməsi prosesi həyata keçirilir [7]. Metaverilənlər - hər hansı bir sənəd daxilindəki informasiya mənbəyi haqqında məlumatlardır [8]. Bu məlumatlar sənədin xüsusiyyətlərini (başlıq, mövzu, açar sözlər, kateqoriya, status, şərhlər, düzəlişlər və ümumi redaktə kimi məlumatları) özündə birləşdirir. İstənilən metaverilənlərdə sənədin kim tərəfindən yazılması, onun məqsədi, fayl formatı, hansı dildə yazılması, sənədin İnternetdə yerləşdiyi ünvan və s. kimi bir çox həssas məlumatlar daxil edilir.

Verilənlərin sanitarizasiyası texnologiyası əvvəllər yalnız çap olunmuş materiallar üzərində redaktə kimi həyata keçirilirdi, indi isə bu xüsusi emal prossesindən onlayn resursların təmizlənməsində də istifadə edilir [4]. Araşdırmalar göstərir ki, fərdlərə və ya təşkilatlara məxsus olan məlumatların konfidensiallığının pozulması riski getdikcə artır [9].

Məlumatların gizliliyinin təmin olunması üçün onları anonimləşdirməklə yanaşı, sənədlər müəyyən auditoriyaya təqdim olunmazdan və ya paylaşılmazdan əvvəl həssas məlumatlar üzərində "təmizləmə" işləri aparılır. Əvvəlcə identifikatorlar müəyyən edilir, məlumatlar verilənlər bazasından silinir, yaxud da bəzi hissələr maskalanır [10]. Bu proses müəyyən şərtlər daxilndə olmaqla, iki mərhələdə yerinə yetirilir: I mərhələdə qadağan olunmuş identifikatorlar seçilir, II mərhələdə isə sanitarizasiya prosesi reallaşdırılır (Şəkil 1).



Şəkil 1. Mərhələli sanitarizasiya prosesinin təsviri [10]

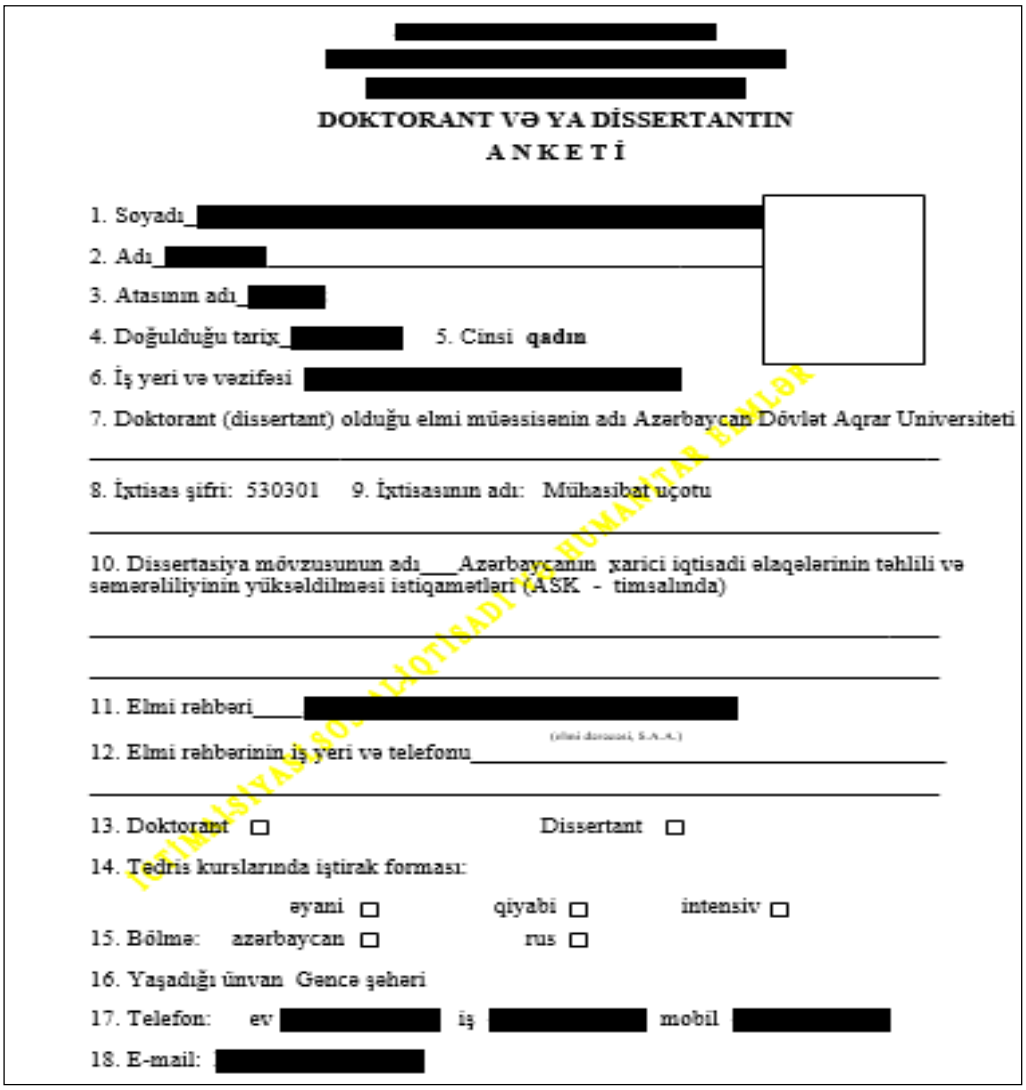
İnformasiyanı icazəsiz girişlərdən qorumaq üçün məlumatların konfidensiallığının təmin edilməsi vacibdir. Fərdlərə məxsus olan şəxsi, həssas informasiyanın gizlədilməsi (buraya kommersiya sirrini təşkil edən məlumatlar, işçilərin şəxsi məlumatları, informasiya sahibinin nüfuzuna zərər verə bilən və digər konfidensial məlumatlar daxildir) bu məlumatların qorunmasında daha əhəmiyyətlidir.

Verilənlərin sanitarizasiyası mətn sənədlərinə, *pdf* formatında olan materiallara, multimedia fayllarına və müəyyən yaddaş qurğularına tətbiq edilir. Bununla əlaqədar aşağıdakıları qeyd etmək olar [11, 12]:

- qalıq məlumatın diskdən silinməsi;
- disk qurğusundan məlumatın silinməsi;

– yalnız faylın silinməsi;
– diskin fiziki məhv edilməsi;
– məlumatın şifrələnməsi;
– ötürülən məlumatların şifrə açarlarının qorunması;
– istifadəçilər üçün şəffaflığın təmin olunması (avtomatik şifrələmə);
– diskin daxili şifrələnməsi (yaddaş qurğusu tərəfindən məlumatın şifrələnməsi);
– məlumatın deşifrə olunması üçün açar;
– diskdə olan məlumatların sanitarizasiyası;
– diskin yenidən istifadəsi üçün istifadəçi məlumatlarının təhlükəsiz silinməsi.

Mətnlərin sanitarizasiyası zamanı həssas məlumatlar müəyyənləşdirilərək müəyyən şərtlər daxilində sanitarizasiya edilir. Bu zaman bəzi hissələr qaralanmış, silinmiş şəkildə göstərilir, beləliklə, həssas məlumatlar sənəddən təmizlənmiş olur. Aşağıda göstərilən nümunədə (Şəkil 2) olduğu kimi, sanitarizasiya olunmuş məlumatların tam olaraq görünməsi mümkünsüz olur.



DOKTORANT VƏ YA DİSSERTANTIN
A N K E T İ

1. Soyadı_
2. Adı_
3. Atasının adı_
4. Doğulduğu tarix_ 5. Cinsi **qadın**
6. İş yeri və vəzifəsi
7. Doktorant (dissertant) olduğu elmi müəssisənin adı Azərbaycan Dövlət Aqrar Universiteti
8. İxtisas şifri: 530301 9. İxtisasının adı: Mühasibat uçotu
10. Dissertasiya mövzusunun adı___Azərbaycanın xarici iqtisadi əlaqələrinin təhlili və səmərəliliyinin yüksəldilməsi istiqamətləri (ASK - timsalında)
11. Elmi rəhbəri____
12. Elmi rəhbərinin iş yeri və telefonu_____
13. Doktorant ☐ Dissertant ☐
14. Tədris kurslarında iştirak forması:
əyani ☐ qiyabi ☐ intensiv ☐
15. Bölmə: azərbaycan ☐ rus ☐
16. Yaşadığı ünvan Gəncə şəhəri
17. Telefon: ev iş mobil
18. E-mail:

Şəkil 2. Həssas məlumatlardan təmizlənmiş sənəd nümunəsi

Sanitarizasiya prosesi daha çox dövlət idarələri, tibb müəssisələri tərəfindən həyata keçirilir. Bu zaman rəhbərlik tərəfindən hansı informasiyanın geniş kütləyə açıq olub-olmaması haqda təlimat verilir və onlar klassifikasiya olunaraq oxşar görünüşlü, lakin yanlış məlumatlar ilə əvəz edilir. Böyük şirkətlər də öz maliyyə sirlərinin yayılmasının qarşısını almaq məqsədilə məlumatlar üzərində sanitarizasiya tədbirləri həyata keçirirlər. Belə ki, müəssisəyə məxsus olan həssas

məlumatlar yalnız korporativ mühitdə istifadə olunduğundan, həmin sənəd daxilində olan həssas hissələr maskalanır və ya dəyişdirilir.

Verilənlərin sanitarizasiyası texnologiyaları

Sanitarizasiya verilənlər bazasının hər hansı bir proqramla silinməsi, sanitarizasiya olunacaq qurğunun başqa qurğu ilə əlaqələndirilməsi, yaxud da fiziki olaraq qurğunun məhvi prosesini özündə birləşdirir. Sanitarizasiya olunan yaddaş qurğuları arasında maqnit disklər, yaddaş qurğuları, *CD* və *DVD*-lər əsas yer tutur [13].

Sanitarizasiya metodlarından istifadə etməklə qurğulardan informasiya bərpaolunmaz şəkildə silinir, yaxud məhv edilir [6, 13]. Sanitarizasiya olunmuş qurğuda olan məlumatlar faylbərpa proqramları vasitəsilə də bərpa oluna bilmir, bərpa olunduqda isə istifadəyə yararsız vəziyyətdə olur.

Verilənlərin təmizlənməsi zamanı aşağıda qeyd olunan üç hala baxılır: verilənlərin redaktəsi, verilənlərin təhlükəsizliyi və effektiv sanitarizasiya metodları [14].

Verilənlərin redaktəsi dövlət orqanları tərəfindən hüquqi sənədlərin konfidensiallığının qorunması məqsədilə istifadə olunur. Sənəd daxilində olan həssas informasiyanın ilkin variantını dəyişdirmək və ya maskalamaq əvəzinə verilənlərin redaktəsi kimi daha etibarlı metoddan istifadə edilir. Redaktə zamanı həssas sahələr və ya məzmun tamamilə aradan qaldırılır, yaxud informasiya qarışdırılır [14]. Məlumatları gizlətmək məqsədilə istifadə olunan əvvəlki üsulların tətbiqi zamanı mətn tipli sənədlərdə gizli məlumatların rəngi dəyişdirilirdi. Yazının fonu ilə onun yazıldığı şriftin rəngi təxminən eyni olurdu. Lakin bu bir qədər sadə üsul olduğundan oxucular tərəfindən axtarış zamanı gizli mətnin aşkarlanması mümkün idi.

Verilənlərin təhlükəsizliyinin təmin olunması informasiyanın tamlığının, əlyetərliliyinin pozulması və onların məlumat bazalarından silinməsinin qarşısını almaqdan ibarətdir. Onlayn və oflayn rejimdə yayılmış rəqəmsal verilənlər üzərində sanitarizasiya işlərinin yerinə yetirilməsi bir qədər çətindir. Çünki mövcud prosedur verilənləri sanitarizasiya etməyin və ya silməyin əleyhinədir [15]. Məlumatlar diskdən silinərkən yalnız onun indeksi təmizlənir. Keş-yaddaşda, informasiya buludunda yerləşdirməklə və verilənlərin sərt diskdə ehtiyat nüsxəsini çıxarmaqla və s. üsullarla onları mühafizə etmək mümkündür.

Effektiv sanitarizasiya metodlarından istifadə etməklə verilənlərin təmizlənməsi üçün onların üzərinə yeni bir qat əlavə olunur və ya onun konfidensiallığını təmin etməklə bu proses həyata keçirilir. İnformasiya yaddaş qurğusunda və ya sənəddə mövcud olduğu üçün qalıq məlumatlar mümkün qədər aradan götürülməlidir. Bu məqsədlə diski fiziki olaraq məhv etməklə və ya düzgün avtorizasiya olunmuş kriptoqrafik üsullardan istifadə etməklə verilənlərin təhlükəsiz şəkildə təmizlənməsi işi yerinə yetirilir [10,12,14,16]. Verilənlərin sanitarizasiyasını həyata keçirmək üçün aşağıdakı mövcud üsullardan istifadə olunur:

Verilənlərin maskalanması – verilənlərin müəyyən sahələrinin hər hansı maska simvolu ilə (məsələn, X) gizlədilməsidir. Bu zaman kontent effektiv formada gizlədilir. Verilənlərin başlanğıc və son hissələri saxlanılmaqla qalan hissə maska simvolu ilə gizlədilir, bunlar isə məlumatları mühafizə etməyə imkan verir [16, 17]. Məsələn, kredit kartı nömrələrinin sətirləri ilkin variantda aşağıdakı kimi görünə bilər:

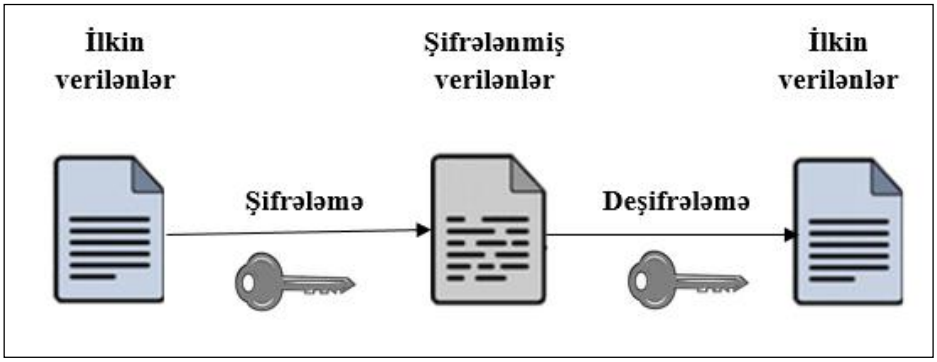
5447 6454 0020 5780
5392 9137 7315 5888
5197 8296 7496 8623

Maskalandıqdan sonra belə görünəcəkdir:

5447 XXXX XXXX 5780
5392 XXXX XXXX 5888
5197 XXXX XXXX 8623

Simvolların maskalanması həssas informasiyanın effektiv silinməsini təmin edir. Eyni zamanda, təhlükəsizlik üçün verilənlərin düzgün şəkildə qorunmasını da təşkil edir.

Şifrələmə/Deşifrələmə – kriptoqrafik alqoritmlərdən istifadə etməklə informasiyanı ötürən bir prosesdir. Asimmetrik şifrələmədən istifadə etməklə verilənlərə giriş müəyyən kodlar və ya şifrələrlə həyata keçirilir. Məlumat yalnız açar məlum olan şəxslərə açıq şəkildə təqdim edilir, digərləri üçün isə görünməz olaraq saxlanılır. Aşağıda məlumatların şifrələnməsi prosesi göstərilmişdir (Şəkil 3) [6].



Şəkil 3. Məlumatların şifrələnməsi prosesi

Şifrələmə verilənlər bazasının formatında və görüntüsündə dəyişiklik yaradır. Belə bazalar, adətən, binar verilənlər kimi görünür. Məlumdur ki, şifrələmənin dayanıqlılığı vacib şərtlərdən biridir. Məlumatın təhlükəsiz qalması şifrələmənin gücünə bağlıdır. Bu üsul verilənlərin sanitarizasiyasını həyata keçirmək üçün ən təhlükəsiz üsullardan biri hesab olunur.

Sıfırlama – bu üsul, sadəcə olaraq, verilənlər sətrinin "sıfır" qiymətləri ilə əvəz edilməsindən ibarətdir. Effektiv olmasına baxmayaraq, onun test bazalarına tətbiq edilməsi çox əlverişsizdir. Adətən, test komandalarının verilənlərin özü ilə, yaxud da ona ən yaxın forma ilə işləməsini nəzərdə tutur [18].

Əvəzləmə – bu üsul vasitəsilə sütundakı informasiyanın tamamilə fərqli olan informasiya ilə əvəz edilməsi nəzərdə tutulur. Bu metod verilənlərin görüntüsünün saxlanılmasında effektiv hesab olunur [19]. Böyük həcmli verilənlərlə işləyərkən bu metodun yaratdığı əsas çətinlik yeni əvəzedicinin tapılması ilə bağlıdır. Məsələn, milyonlarla küçə adının sanitarizasiyası zamanı belə çətinliklər yarana bilər.

Yazıların qarışdırılması – qarışdırılmış məlumatlardan istifadənin effektivliyini daha aydın görmək üçün yüksək prioritetli məlumat bazalarında bu üsuldan istifadə olunur. Çünki bu üsul məlumat bazalarından istifadə zamanı məlumatların aşkarlanma riskini minimuma endirir. Praktik olaraq məlumatların qarışdırılması problemli və ya dəyişdirilmiş gizli məlumatların istifadəsi ilə bağlı təhlükələri aradan qaldırır. Təmizləmə üsullarının bütün mövcud xüsusiyyətlərini özündə saxlamaqla, məlumatların tamlığının pozulması riskini aradan qaldırır və bu prosesi daha effektiv şəkildə yerinə yetirir [19].

Ədədlərin variasiyası – bu üsul, əsasən, elektron resursların sanitarizasiyası zamanı istifadə edilir. Xüsusi alqoritm vasitəsilə rəqəmlər dəyişdirilir. Verilənləri müəyyən faizlə əvəz etməklə onlar gizlədilir [20].

Mənasız verilənlərin generasiyası – bu üsulun tətbiqində vacib şərt verilənlərə gələn bütün növ keçidləri, izləri silməkdir. Xüsusilə, gizlədilmə hər hansı açar sözlə, yaxud da sadə metodla həyata keçirildikdə izlərin itirilməsi daha vacibdir. Strukturlaşdırılmamış verilənlərin, məsələn, yaddaş qeydlərinin, məktubların sanitarizasiya edilməsi prosesi verilənlərin sanitarizasiyasının ən çətin üsullarından biri hesab edilir. Bu üsulun tətbiqi zamanı ən çox sözlərin hər hansı təsadüfi sözlərlə əvəz edilməsindən istifadə olnur [18].

Maqnitsizləşdirmə – bu proses xüsusi maqnit sahəsinə malik olan cihaz vasitəsilə həyata keçirilir [21]. Bu zaman yaddaş qurğusunun maqnitə həssas olan sektorlarına maqnit vasitəsilə təsir edərək sektorlardakı məlumatların üzərinə sıfır yazılır, yaddaş qurğusu güclü maqnit sahəsinə məruz qalaraq neytrallaşdırılır. Maqnitsizləşdirmə yalnız sərt disklərdə, bir çox maqnit lentlərində

tətbiq olunur və fiziki dağıdılma zamanı verilənlərin bərpa edilməsi qeyri-mümkün olur. Bu üsulun mənfi cəhəti ondadır ki, qurğular məhv edilərkən ekologiyanı korlayır. Bu üsulun tətbiqi zamanı istifadə olunan qurğu *degauster* adlanır [21].

Üzərinə yazmaq – verilənlərin sanitarizasiyasının ən ucuz və sürətli üsulu kimi qəbul edilmişdir. Amma sanitarizasiya zamanı qalıq bazada olan informasiyanın təhlükəsizliyi çox aşağı olur. Bununla da, bazada hansı dəyişikliklərin edildiyini müəyyən etmək çox asan olur [22].

Təmizləmə – verilənlərin sanitarizasiyasının elə silinmə üsuludur ki, belə silinməni yerinə yetirdikdən sonra heç bir iz qalmır və heç bir qalıq məlumat buraxılmır [23].

Məhv etmək – bu üsul fiziki olaraq kompüterlərin, sərt disklərin, smartfonların, printerlərin, noutbukların və digər informasiya daşıyıcılarının mexaniki vasitələrlə parçalanmasıdır. Bu metod, xüsusilə, yüksək həssas məlumatlarla zəngin bazalarla işləyərkən daha əhəmiyyətlidir. Məlumatlar məhv edildikdən sonra verilənlərin yenidən əldə olunması mümkünsüz olur [24]. Bu proses qurğuları hissələrə ayırmaqla müxtəlif maqnitsizləşdiricilərdən istifadə etməklə yerinə yetirilir.

Sanitarizasiya üsullarından düzgün istifadə edilmədikdə informasiya sıza və konfidensial məlumatlar geniş auditoriyaya yayılaraq müəyyən problemlərin yaranmasına səbəb ola bilər.

Verilənlərin sanitarizasiyasının elmi-nəzəri problemləri

Müxtəlif informasiya sistemlərində saxlanılan, emal olunan və ötürülən məlumatlarda mövcud olan konfidensial, dövlət əhəmiyyətli informasiyanın gizlədilməsi üçün bu məlumatların "təmizlənməsinə" ehtiyac vardır. İnformasiyanın emaldan öncə aqreqatlaşdırılması, strukturlaşdırılması və qiymətləndirilməsi prosesləri də "təmizlənməyə" aid edilir. Kompüter şəbəkələri əsasında mürəkkəb məsələlərin həlli üçün paylanmış hesablama sistemlərinin yaradılmasında bulud texnologiyalarından geniş istifadə olunur. [25].

[26]-də Tayvanın Milli *Chung Cheng* Universitetinin mütəxəssisləri Cheng-Yuan Ku və Yu-Siang Chiu tərəfindən buludda saxlanılan məlumatların təmizlənməsi məsələsinin həlli üçün infrastruktur təklif edilmişdir. Bu mexanizm monitorinq agentləri mexanizmi vasitəsilə proqramın bütün həyat dövrü ərzində məlumatların istifadəsini izləmək, informasiyanı strukturlaşdırmaq və əldə olunan verilənlər əsasında qərar qəbul edərək onların sanitarizasiyasını təmin etməyi nəzərdə tutur. Təhlükəsizlik məsələsinin həllinə yönəlmiş verilənlərin sanitarizasiyası yalnız qərar qəbulundan sonra həyata keçirilir. Bu üsul bulud texnologiyaları istifadəçiləri üçün təhlükəsizlik və məhsuldarlıq tələblərini yerinə yetirmiş olur.

Elektron sənədlərin təmizlənməsi prosesi həssas məlumatların müəyyən edilməsi və bu məlumatların sənəddən çıxarılması və ya gizlədilməsidir. Burada iki mərhələdən istifadə olunur. Birinci mərhələdə sənəddəki həssaslıq şərtləri müəyyənləşdirilir. Şərtlər qərar qəbuletmə səlahiyyətinə malik olan şəxs və ya təşkilat tərəfindən müəyyənləşdirilir. İkinci mərhələdə isə müəyyənləşdirilmiş şərtə uyğun olaraq həssas terminlər silinir.

Adətən, mətn sənədinin sanitarizasiyası ənənəvi olaraq ixtisaslı rəyçilər tərəfindən əl ilə aparılırdı. Məlumatların həcmi artdıqca, bu üsuldan istifadə çətinləşir. Yəni avtomatik üsullar təklif olunur. Sənəddən həssas terminlərin müəyyən edilməsi və aradan qaldırılması üçün istifadə olunan müxtəlif üsullar müzakirə olunur. [15]-də sənədlərin sanitarizasiyası üçün yeni metod təklif edilmişdir. Təklif olunan metod vasitəsilə, ilk növbədə, həssas terminlər aşkarlandıqdan sonra onlar mətn daxilindən çıxarılır.

[27]-da Erase silmə alqoritmindən istifadə edərək həssas və ya konkret məlumatları əhatə edən sənədlərin gizliliyinin əldə olunması məqsədilə sənədləşdirmə sisteminin avtomatlaşdırılması üçün metod təklif olunur.

Data Mining texnologiyalarından istifadə sosial şəbəkələrdə gizli biliklərin aşkarlanması ilə yanaşı, dünyada gedən proseslərin izlənməsinə, ayrı-ayrı dövlətlərdə baş verən hadisələrdən xəbərdar olmağa imkan yaradır. Sosial şəbəkələrdən əldə olunan məlumatlar emal olunaraq, oradakı gizli informasiya müəyyən olunur. Bunlara qarşı da verilənlərin sanitarizasiyası metodlarının işlənməsi aktualdır.

Assosiativ qaydaların aşkarlanması ilə hücum metodu Data Mining texnologiyalarında ən geniş istifadə olunan və İnternetdə ciddi problemlərə səbəb olan əməliyyatdır. Bu metodun məqsədi daha yüksək səviyyədə qorunması lazım olan faktların müəyyənləşdirilməsi üçün məlumatları müəyyən təhlükəsizlik səviyyəsində birləşdirməkdir [28].

[29]-da sosial şəbəkə verilənlərindən gizli informasiya əldə etmək üçün istifadə edilən məntiqi çıxarış hücumlarına qarşı bəzi alqoritm və metodlar təqdim edilmişdir. Klasterləşmədən istifadə etməklə hücum alqoritmi tətbiq edən fərdlərin gizliliyi təmin edilir. İnformasiya mənbəyi kimi sosial şəbəkələrdə, açıq kodlu sistemlərdə və mənbələrdə olan məlumatlardan istifadə olunur.

Verilənlərin sanitarizasiyasının perspektivləri

Məlumatların sanitarizasiyası inkişaf etməkdə olan tədqiqat sahəsi olduğundan burada bir sıra problemlər vardır. Əsas problemlərdən biri gizlilik və təhlükəsizlik tədbirlərinin daha geniş və səmərəli şəkildə yerinə yetirilməsində xətaların olması, sənədlərdən müəyyən identifikatorların təmizlənməsi və informasiyanın qurğulardan tamamilə silinməsi məsələlərinin həllinin reallaşdırılmasıdır.

Bundan əlavə, məlumatın strukturunda olan dəyişikliklər yeni informasiya sisteminə keçdikdə və ya birdən çox məlumat mənbəyinə inteqrasiya edildikdə, informasiya auditoriyaya fərqli formada təqdim olunur. Yaranmış problemlərin həlli məlumatların strukturunda, təqdimatında və ya məzmununda hər hansı dəyişikliklərin edilməsi ilə yanaşı, onların təhlükəsizliyi məsələsini də təmin edə bilər. Verilənlərin sanitarizasiyası metodlarının tətbiqi ilə bağlı bir sıra problemlər mövcuddur. Onlardan bir neçəsini qeyd edək:

- verilənləri yüksək səviyyədə sanitarizasiya edən proqram vasitələrinin sayının az olması;
- verilənlərin sanitarizasiyası və onun ayrı-ayrı metodlarının geniş tədqiq olunmaması;
- qurğularda verilənlərin sanitarizasiyası metodlarının yeni üsullarının az olması;
- mobil və ya şəbəkə qurğularında məlumatların məhv edilməsi ilə bağlı problemlərin olması;
- verilənlərin sanitarizasiyası üçün hüquqi tələblərə dair qanunların, normativ sənədlərin kifayət qədər olmaması;
- verilənlərin sanitarizasiyası metodlarının tətbiqi zamanı onun ətrafında olan əlaqəli məlumatlarla (küylərlə) bağlı problemin mövcudluğu;
- silinmə prosesi düzgün yerinə yetirilmədiyi zaman informasiya sızmalarının olması;
- qurğulardan informasiyanın çıxarılması problemi;
- sanitarizasiya prosesinə casus proqramlarının müdaxiləsi problemi;
- sanitarizasiya olunan sənədlərdə dil (linqvistika) problemi və morfoloji analiz problemi;
- başqa dillərdə istifadə edilən sanitarizasiya metodlarının Azərbaycan dilinə uyğun olmaması problemi;
- multimedia fayllarının sanitarizasiyasında mürəkkəbliklər, xətalar və s. olması;

Azərbaycan dilinə uyğun müəyyən sanitarizasiya siyasətinin həyata keçirilməsi də çox əhəmiyyətlidir. Dilin həm müəyyən söz və ifadələrdən təmizlənməsi linqivistika problemi ilə bilavasitə əlaqəlidir. Bu baxımdan, dilin müəyyən zərərli məzmundan (vulqar sözlər, zorakılıq xarakterli ifadələr, yaş qrupları üçün uyğun olmayan mətnlər, sözlər, səslər, videotəsvirlər və s.) təmizlənməsi problemi həlli olunmalıdır.

Yuxarıda qeyd edilənləri nəzərə alaraq problemlərin aradan qaldırılması üçün aşağıdakı təklifləri irəli sürmək olar:

- göstərilən problemlərin həllində yaranacaq çətinliklərin və risklərin aradan qaldırılması üçün geniş elmi-tədqiqat işlərinin aparılmasına ehtiyac vardır;
- sosial şəbəkə yazılarında, mətn sənədlərində, audio və video fayllarda süzülmə və klassifikasiya işlərini yerinə yetirən riyazi yanaşmalara yer verilməli, məlumat bazalarında həssas məlumatların qorunması üçün yeni metodlar işlənilməlidir;

- onlayn mühitdə "Vulqar sözlər bazası" yaradılmalıdır;
- mətnlər çap edilərkən sözlər kompüterdə və ya onlayn mühitdə əvvəlcədən yaradılmış "Vulqar sözlər bazası"na daxil edilmiş sözlərlə müqayisə olunmalıdır (siqnal verilir, bloklanır və s.);
- televiziya kanallarında canlı yayımda danışarkən insanın dilindən çıxan vulqar söz və ifadələrin işlənməsi zamanı önləmə tədbirləri görülməli və ya bloklanmalıdır;
- sanitarizasiya işinin yerinə yetirilməsi məsələlərinin nəzəri həllinə müxtəlif ixtisaslardan olan mütəxəssislər (dilçi, filoloq və s.) cəlb edilməlidir;
- verilənlərin sanitarizasiyası metodlarının tətbiqi ilə İnternetdə mövcud olan qanunazidd və təhlükəli kontentdən qorunmaq üçün müxtəlif mexanizmlər işlənilməlidir.

Nəticə

Sürətlə inkişaf edən informasiya əsrində İnternet resusrlarından geniş istifadə olunduğu, çoxlu sayda sənədlərin dərc edildiyi və paylaşıldığı bir vaxtda bu mövzunun araşdırılması və tədqiq olunması labüddür. Müəyyən üsulları tətbiq etməklə yaranacaq problemlərin və risklərin aradan qaldırılması mümkündür. Lakin məlumatların sanitarizasiyası yeni və inkişaf etməkdə olan tədqiqat sahəsidir. Burada gizlilik və təhlükəsizlik tədbirləri mühüm rol oynayır. Bu sahədə bir sıra mövcud sualların cavablandırılması və yaranmış problemlərin qarşısının alınması məqsədilə mövzunun daha dərindən tədqiq edilməsinə ehtiyac vardır.

Ədəbiyyat

1. İmamverdiyev Y.N. E-dövlətin informasiya təhlükəsizliyinin idarə olunmasının konseptual modeli // İnformasiya cəmiyyəti problemləri, 2013, №1, s.20-25.
2. Azərbaycan Respublikasında informasiya cəmiyyətinin inkişafına dair 2014-2020-ci illər üçün Milli Strategiya, 2 aprel 2014-cü il, http://www.president.az
3. "Uşaqların zərərli informasiyadan qorunması haqqında" Azərbaycan Respublikasının Qanunu, http://www.president.az
4. Əliquliyev R.M. Ocaqverdiyeva S.S. Uşaqların İnternetdə informasiya təhlükəsizliyini təmin edən sistemin konseptual modeli / İnformasiya təhlükəsizliyinin aktual problemləri III respublika elmi-praktiki seminarı, Bakı, 2017, s. 84-87.
5. Crawford R., Bishop M., Bhumiratana B., Clark L., Levitt K. Sanitization Models and their Limitations / Proceedings of the 2006 Workshop on New Security Paradigms, 2006, pp. 41-56.
6. Spacey J., Edqar D. Data Sanitization Techniques, https://www.orafaq.com/papers/data_ sanitization.pdf
7. Riley J. Understanding metadata what is metadata, and what is it for? www.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf
8. Coyle K. Understanding Metadata and Its Purpose// The Journal of Academic Librarianship, Vol. 31, No. 2, pp. 160-163.
9. Shamir A. How to share a secret // Communications of the ACM, 1979, vol.22, no.11, pp. 612-613.
10. Stanley R. M. Oliveira, Osmar R. Zaiane. A Unified Framework for Protecting Sensitive Association Rules in Business Collaboration // International Journal of Business Intelligence and Data Mining, 2006, vol.1, issue 3, pp.247-287.
11. Hughes G., Coughlin T. Tutorial on Disk Drive Data Sanitization, 2006, https://www.researchgate.net/publication/229003088
12. Chakaravarthy V.T., Gupta H., Roy P., Mohania M. Efficient techniques for document sanitization /Proceeding of the 17th ACM Conference on Information and Knowledge Mining (CIKM), 2008, pp.843-852.
13. Ivashcu M. Data erasure on magnetic storage / International conference of scientific paper afases, 2011, pp. 614-617.

14. Data Sanitization Terminology and Definitions, https://www.datasanitization.org/data-sanitization-terminology
15. Krishna A.K, Suriya M. Preventing Private Information Leakage on Social mining // International Journal of Innovative Research in Computer and Communication Engineering, 2014, vol.2, issue 1, pp. 3530-3534.
16. Wiseman L., Gordon V. Data protection: Guidelines for the use of personal data in system testing, 2009. https://www.shop.bsigroup.com/upload/Shop/Download/Books/BIP0002sample.pdf
17. Sarada G., Abitha N., Manikandan G., Sairam N. A few new approaches for data masking / International Conference on Circuits, Power and Computing, India, 2015, pp. 1-4.
18. About Data Sanitization, Redaction & Cleansing Methods, https://www.forensicsware.com/blog/data-sanitization.html
19. Edqar D. Data Sanitization Techniques, https://www.orafaq.com/papers/data_sanitization.pdf
20. Ch. Deepika, Bansal P. Study on need of data sanitization and sanitization techniques for memory devices // Access International Journal of Science& Engineering, 2017, vol. 2, Issue 11, pp. 6-16.
21. Degaussing: An Introduction, https://www.veritysystems.com/degaussing-an-introduction/
22. Michael W., Laura M. Grupp, Frederick E. Spadat, Steven S. Reliably Erasing Data From Flash-Based Solid State Drives / FAST'11 Proceedings of the 9th USENIX conference on File and stroage technologies, 2011, pp. 8-21.
23. Overview of Data Masking Methods, https://www.smartbridge.com/overview-data-masking-methods
24. Violino B. The in-depth guide to data destruction, CSO Online, 2012, http://www.csoonline.com/article/2130822/it-audit/the-in-depth-guide-to-data-destruction.html
25. R.Q. Ələkbərov, M.A. Həşimov. Bulud texnologiyaları: xidmətlər, problemlər və tətbiq sahələri // İnformasiya texnologiyaları problemləri, 2016, №.1, s.3-10.
26. Cheng-Yuan Ku, Yu-Siang Chiu. A novel infrastructure for data sanitization in cloud computing / Proceedings of Technology Innovation and Industrial Management, 2013, pp. 25-28.
27. Chakaravarthy V. T., H. Gupta, P. Roy, Mohania M. K. Efficient techniques for document sanitization / Proceeding of the 17th ACM Conference on Information and Knowledge Mining (CIKM), 2008, pp. 843-852.
28. Dasseni E., Verykios V. S., Elmagarmid A. K., Bertino E. Hiding Association Rules by Using Confidence and Support / Proc. of the 4th Information Hiding Workshop, 2001, pp. 369-383.
29. Chandra D, Antony Roosevelt L. Sanitations To Prevent Inference Attack On Social Network Data // International Journal of Innovative Research in Science, International Conference on Engineering, 2014, vol.3, special issue 1, pp.1265-1269.

УДК 04:37

Оджагвердиева Сабира С.
Институт Информационных Технологий НАНА, Баку, Азербайджан
allahverdiyevasabira@gmail.com

О некоторых актуальных проблемах санитаризации данных

В статье описываются санитаризация данных, ее цели и задачи, сферы применения, область применения и перспективы. Проанализированы технологии санитаризации данных, существующие методы, некоторые научно-теоретические и актуальные проблемы. Проделана работа, направленная на решение этой проблемы, и предложены соответствующие рекомендации и предложения.

Ключевые слова: *санитаризация данных, очистка данных, конфиденциальность, личная информация.*

Sabira S. Ojagverdiyeva
Institute of Information Technology of ANAS, Baku, Azerbaijan
allahverdiyevasabira@gmail.com

Some actual problems of data sanitization

The article explores the goals and objectives of data sanitization, its scope and application areas and prospects. Data sanitization technologies, existing methods, and some scientific-theoretical and actual problems in this field are analyzed. The implementations for the solution of this problem are analyzed, and relevant recommendations and suggestions are put forward.

Keywords: *data sanitization, data cleaning, confidentiality, personal data.*