Elektron Kitabxanalarda Verilənlərin Sanitarizasiyası Üsulları

Sabirə Allahverdiyeva
AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
allahverdiyevasabira@gmail.com

***Xülasə* — Məqalədə elektron kitabxanalardan və elektron məlumat bazalarından istifadə edərkən qarşıya çıxan problemlər və onların həlli yolları araşdırılır. Sənəd və ya məlumat bazalarında mövcud olan konfidensial məlumatların sanitarizasiya metodları vasitəsilə gizlədilməsi və bu məqsədlə istifadə olunan müxtəlif üsullar şərh edilir.**

Açar sözlər – verilənlərin sanitarizasiyası, sanitarizasiya üsulları, konfidensial məlumatlar, metaverilənlər.

I. GİRİŞ

İnformasiya tələbatının operativ və daha dolğun ödənməsində elektron nəşr texnologiyalarının və elektron kitabxanaların rolu böyükdür. Elektron kitabxanaların ümumi istifadə üçün açıq olması və müxtəlif məsafələrdə yerləşən paylanmış informasiya ehtiyatlarına malik olması, informasiyadan istifadə edərkən rəqəmsal formada hazırlanmış müxtəlif formatlı sənədlərin mövcudluğu istifadəçiyə daha tez məlumat əldə edilməsinə şərait yaradır.

İnternet texnologiyaları dövründə informasiya qıtlığı demək olar ki, yoxdur. Lakin elektron nəşrlərin olması məlumatı rahat əldə etməkdən başqa müxtəlif problemlərə də yol açır. Əgər informasiya hər hansı bir şəxsin kimliyini ortaya çıxarırsa və ya konfidensial məlumatları deşifrə edirsə, bu zaman informasiya şəxsi təhlükəsizlik üçün bir təhdidə çevrilir [1]. Beləliklə, sənədləri dərc etməzdən əvvəl konfidensial məlumatlar üzərində müxtəlif "təmizləmə" əməliyyatlarının (verilənlərin sanitarizasiyası) aparılması zərurəti yaranır. Bu məqsədlə məlumatların silinməsi və ya onların maskalanmasından istifadə etmək lazım gəlir.

Həssas məlumatların aşkarlanaraq oradakı məzmunun silinib, maskalanmasını həyata keçirən bir çox avtomatik metodların mövcud olması məlumatın təhlükəsizliyini təmin edir. Beləliklə, bir çox şəxs üçün faydalı olan sənəd daxilindəki məzmunun təhlükəsizliyi qorunur və daha geniş kütlə üçün əlçatan olur.

Bu məqsədlərlə həm geniş kütlə üçün nəzərdə tutulmayan məlumatların qorunması, həm də ziyanlı informasiyanın bilavasitə azyaşlı İnternet istifadəçisinə yönəlmiş təhlükələrin qarşısının alınması üçün müxtəlif texnoloji üsullardan, mexanizmlərdən və ya proqram vasitələrindən istifadə olunur.

Sənədə daxil olan həssas məlumatların gizli saxlanması çox vacib olduğundan, müxtəlif üsullardan istifadə zərurəti yaranır. Gizliliyin saxlanması üçün gizli məlumatın silinməsi, yaxud gizlədilməsi onun yararlılığına təsir göstərməməlidir. Bu isə, verilənlərin sanitarizasiyasının başlıca məqsədidir.

II. METAVERİLƏNLƏR

Metaverilənlər hər hansı bir sənəd daxilindəki informasiyanın mənbəyi haqqında məlumatları özündə saxlayır. Onlar vasitəsilə sənədin kim tərəfindən harada yazılması və s. barədə geniş məlumat almaq mümkündür. Çox zaman metaverilənlər "informasiya haqqında məlumat" kimi adlandırılır [2].

Sənəd daxilindəki metaverilənlərin verdiyi məlumatları 3 kateqoriyaya ayırmaq olar :

1. Açıqlayıcı metaverilənlər – sənəd daxilindəki informasiya mənbəyi haqqında məlumat verir və sənədin yazılma məqsədini açıqlayır.
2. Struktur metaverilənlər – sənəd daxilindəki informasiyanın strukturu haqqında məlumatdır (məs., sənədin mətn və qrafik məzmunu haqqında).
3. Administrativ metaverilənlər – sənədin nə zaman, hansı fayl formatında yaradıldığı (.pdf, .docx, .ppt vəs.) və sənədə kimlərin baxış icazəsinin olması kimi məlumatlardır.

Müxtəlif metaverilənlər nümunəsində sənədin kim tərəfindən yazılması, yazılma məqsədi, fayl formatı, hansı dildə yazılması, sənədin İnternetdə yerləşdiyi ünvan kimi bir çox "həssas" məlumatlar verilə bilər. İnformasiyanın təhlükəsizliyi baxımından belə "həssas" məlumatların qorunması məntiqi nəticə olaraq qarşımıza çıxır.

III. SANİTARİZASİYA METODU

Təhlükədən qorunma metodlarından biri sanitarizasiya (sanitization) metodudur. Bu metodun mahiyyəti həssas məlumatların geniş ictimaiyyətdən gizlədilməsidir [3]. Adətən çap olunmuş materiallar üzərində həyata keçirilir. Eyni zamanda gündəlik onlayn mediada və kompüterdə də istifadə edilir. Məxfiliyin qorunması zamanı bu metod əsasən məxfi olmayan məlumatların konfidensiallıq səviyyəsini azaldır [4].

Qeyd etmək lazımdır ki, müasir dövrdə texnologiyanın yüksək inkişafının bir nəticəsi kimi elektron məlumatların konfidensiallığının tam qorunmasına zəmanət verilmir. Beləliklə, metaverilənlərin sənəd daxilindən silinməsi həyata keçirilir. Bunun üçün bir çox proqram təminatları "Metadata Removal" adı altında mövcuddur. Bu proqramlar sənədin yaradılma formatına görə fərqlənirlər, yəni şəkil daxilindəki metaverilənlər və .pdf formatlı sənəd daxilindəki metaverilənlər fərqli proqramlar vasitəsilə silinir, lakin yenə də məlumatın tam silinməsinə zəmanət verilmir. Beləliklə, növbəti mərhələlərdə artıq "məlumatın məhv edilməsi" ilə konfidensiallığın tam qorunması labüd olur.

Verilənlərin sanitarizasiyasının həyata keçirilmə yolları müxtəlifdir. Bir çox kompüter proqramlarından istifadə edilir

və bu proqramlara "metadata removal tool" (metaverilənlərin silinməsi aləti) deyilir. Onlar aşağıdakı kimi qruplaşdırılır:

- İnteqral metaverilənlərin silinməsi aləti – onlar proqramların öz daxilində olur (məs., Microsoft Office 2007 Document Inspector).
- Toplam metaverilənlərin silinməsi aləti – bir neçə faylın eyni anda sanitarizasiyası üçün nəzərdə tutulur.
- E-poçt kliyenti əlavələri – elektron ünvanlardan gərəksiz məlumatların silinməsi üçün yararlıdır.
- Server əsaslı sistemlər – şəbəkədən metaverilənləri silir.

IV. SANİTARİZASİYANIN HƏYATA KEÇİRİLMƏSİ ÜSULLARI

Sanitarizasiya üsulları qurğulardan informasiyanın müstəqil, daimi və bərpaolunmaz şəkildə silinməsi və ya məhv edilməsi üçün istifadə olunur [5]. Sanitarizasiya olunmuş qurğuda olan informasiya hər hansı proqramla yenidən bərpa oluna bilmir və ya heç bir işə yaramır. Sanitarizasiya prosesi özündə verilənin hər hansı bir proqramla silinməsini, yaxud sanitarizasiya olunacaq qurğunun başqa qurğu ilə birləşdirilməsini ya da fiziki olaraq qurğunun məhvini ehtiva edir ki, bu yol ilə həmin verilənlər bərpa oluna bilmir.

Sanitarizasiya subyektlərinə DVD-lər, CD-lər, smartfonlar və s. aid edilir. Verilənlərin sanitarizasiyası aşağıdakı üsullarla həyata keçirilir [6,7]:

Sıfırlama (NULL`ing out) – Bu üsul sadəcə olaraq verilənlər sətrin NULL (sıfır) qiymətləri ilə əvəz edilməsindən ibarətdir. Bu metod effektiv olmasına baxmayaraq, test bazalarına tətbiq edilməsi çox əlverişsizdir. Adətən test komandalarının verilənlərin özü ilə, yaxud da ona ən yaxın forma ilə işləməsi nəzərdə tutulur.

Verilənlərin maskalanması (Masking Data) – Verilənlərin maskalanması dedikdə, veriləndə olan müəyyən sahələrin hər hansı maska simvolla (məsələn, X) gizlədilməsi nəzərdə tutulur. Bu verilənlərin kontentin effektiv formada gizlədərək, verilənlərin başlanğıc və son hissələrini saxlamaqla hesabatları mühafizə etməyə imkan verir. Məsələn, kredit kartı nömrələrinin sətirləri aşağıdakı kimi görünə bilər:

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

Maskalanma həyata keçirildikdən sonra isə o belə görünəcəkdir:

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

Simvolların maskalanması həssas informasiyanın effektiv silinməsini təşkil edir, lakin bununla bərabər təhlükəsizlik üçün verilənlərin düzgün şəkildə qorunmasını təşkil edir.

QEYD: Əgər verilənlər xüsusi dəyişilməyən formatdadırsa, maskalanma verilənin sanitarizasiyası üçün güclü imkandır.

Əvəzləmə (Substitution) – bu üsul sütundakı informasiyanı tamamilə fərqli olan informasiya ilə əvəz edir. Bu metod verilənlərin görüntüsünün saxlanılmasında effektivdir.

Böyük həcmli verilənlər işlənilərkən bu metodun yaratdığı əsas çətinlik yeni əvəzedicinin tapılması ilə bağlıdır. Məsələn, milyonlarla küçə adının sanitarizasiyası zamanı belə çətinlik yarana bilər.

Yazıların qarışdırılması (Shuffling Records) – bu üsul əvəzetmə üsuluna bənzəyir, amma əsas fərqi onun sütunun daxilindəki məlumatların bir-biri ilə yerini dəyişməsini nəzərdə tutmasıdır. Əsas çatışmayan cəhətinin biri odur ki, müəyyən təhlükəsizlik problemləri yaradır. Məsələn, əgər kiməsə sadəcə iştirakçıların adı (ardıcıllığı yox) maraqlıdırsa, bu üsuldan istifadə zamanı ehtiyacı olan informasiyanı əldə edə bilər.

Ədədlərin variasiyası (Number Variance) – rəqəm dəyişdirilməsi üsulu əsasən rəqəmsal verilənlərin sanitarizasiyası üçün istfadə edilir. Xüsusi alqoritm vasitəsilə rəqəmləri müəyyən faizlə əvəz edərək onları gizlədir.

Mənasız verilənlərin generasiyası (Gibberish Generation) – Mənasız verilənlərin generasiyası verilənlərin sanitarizasiyası zamanı əsas məsələ verilənlərə bütün növ keçidləri, izləri silməkdir. Xüsusilə, gizlədilmə hər hansı açar sözlə, yaxud da sadə metodla həyata keçirildikdə, izlərin itirilməsi vacibdir.

Formasız verilənlərin məsələn, yaddaş qeydlərinin, məktubların sanitarizasiyası sanitarizasiyanın ən çətin üsullarından biri hesab edilir. Bu zaman ən çox istifadə edilən üsul isə bu tip mətnlərin, sözlərin hər hansı təsadüfi sözlərlə əvəz edilməsidir.

Şifrələmə/Deşifrələmə (Encryption/Decryption) **–** bu üsuldan istifadə zamanı verilənlərə giriş müəyyən kodlar və ya şifrlərlə həyata keçirilir.

Maqnitsizləşdirmə (Degaussing) **–** Xüsusi maqnit sahəsinə malik cihaz vasitəsilə həyata keçirilir, yaddaş qurğusunun maqnitə həssas olan sektorlarına maqnit vasitəsilə təsir edərək sektorlardakı məlumat sıfırlaşdırılır. Bu üsulu həyata keçirərkən diqqətli olmaq tələb olunur, həddindən artıq maqnit təsiri daxilində həssas məlumatlar olan xarici yaddaş qurğusunu sıradan çıxara bilər. Uğurlu şəkildə həyata keçirilmiş maqnitlə təsir seansından sonra yaddaş daşıyıcılarından yenidən istifadə edərək məlumatlar əlavə etmək mümkündür. Maqnitlə təsir üsulundan plastik kartlarda istifadə tövsiyə olunmur.

NƏTİCƏ

Verilənlərin sanitarizasiyası elektron resurslardan istifadə zamanı, verilənlərin analizi və bizneslə bağlı qərarların qəbulu kimi fəaliyyətlər üçün cox əhəmiyyətlidir. Yuxarıda müəyyən üsulları tətbiq etməklə yaranacaq problemlərin və risklərin aradan qaldırılmasını qeyd etdik. Lakin məlumatların sanitarizasiyası inkişaf etməkdə olan tədqiqat sahəsi olduğundan, burada gizlilik və təhlükəsizlik tədbirləri mühüm rol oynayır.

Çoxlu sayda sənədlərin dərc olunduğu və paylaşıldığı sürətlə inkişaf edən informasiya əsrində bu mövzunun araşdırılmasına zərurət hiss olunur. Bir sıra sualların cavablandırılması və yaranmış problemlərin qarşısının alınması üçün mövzu daha dərindən tədqiq edilməlidir.

ƏDƏBİYYAT

[1] V.Vasudevan, A.John. "A Review on Text Sanitization", International Journal of Computer Applications, Vol. 95, No.25, pp. 14-17.
[2] http://www.lawpro.ca/LawPRO/Metaverilən.pdf
[3] Wikipedia. Sanitization (classified information) —wikipedia, the free encyclopedia, 2015.
[4] S.S. Allahverdiyeva. "Uşaqların İnternetdə təhlükəsizliyinin təmin edilməsi problemləri". Bakı, 2016. - 91 s.
[5] A.O.Olajide, A.T. Olarewaju, A. Ondo State. "Application of Data Masking in Achieving Information Privacy", Innovative Systems Design and Engineering, Vol.5, No.1, 2014, pp.27-35.
[6] A.Krishna Kumar, M.Suriya. "Preventing Private Information Leakage on Social mining". International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Special Issue 1, March 2014, pp. 3530-3534.
[7] V.T. Chakaravarthy, H. Gupta, P. Roy and M.K. Mohania, "Efficient techniques for document sanitization", In Proceeding of the 17th ACM Conference on Information and Knowledge Mining (CIKM), pp. 843– 852, 2008.