

Elektron Kitabxana Sistemində Toplanan Loq-Verilənlərin Analizi

Fərhad Yusifov

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
farhad.yusifov@gmail.com

Xülasə — Elektron kitabxana sistemlərinin səmərəliliyinin qiymətləndirilməsində loq-faylların analizi mühüm əhəmiyyət kəsb edir. Məqalədə elektron kitabxana sistemində toplanan loq-verilənlərin analizi məsələsi araşdırılır. Serverdə toplanan loq-fayllar sistemin fəaliyyəti, sistem informasiyası, sorğular və istifadəçilərin davranışları haqqında məlumatı özündə əks etdirir. Loq-fayllar analiz olunmaqla sorğuların, istifadəçilərin profilinin öyrənilməsi və eləcə də, kəmiyyət və keyfiyyət xarakteristikalarının qiymətləndirilməsi üsulları tədqiq olunur.

Açar sözlər — elektron kitabxana; loq-fayl; loq-verilənlər; istifadəçi profili

I. GİRİŞ

Son illər elektron kitabxanalar (e-kitabxanalar) biliyin formalaşdırılması və yayılması sisteminin mühüm tərkib hissəsinə çevrilmişdir. E-kitabxana xidmətlərinin inkişafı, eləcə də, müxtəlif sahələrdə aparılan tədqiqatların nəticələrinə çıxış imkanının yaradılması, əlyətərlilik elmin gələcək inkişafının təminatçısıdır. Təbiidir ki, elektron resursların (e-resursların) yaradılmasına böyük maliyyə vəsaiti, vaxt və intellektual əmək sərf olunur. Bununla yanaşı, e-kitabxanaların gələcək inkişafının göstərilən xidmətlərdən, e-resursların populyarlığından (müraciətlərin çoxluğundan) birbaşa asılılığı ön plana çəkilir [1, 2].

Təcrübədə e-kitabxanaların strukturunu şərh etmək üçün onu bir informasiya sistemi kimi xarakterizə edirlər [1]. Son on illikdə yaradılmış istifadəçi-mərkəzli informasiya sistemlərində istifadəçilərin e-resurslara çıxış imkanları, həmçinin e-resursların formalaşması prosesində onların aktiv iştirakı xüsusilə qeyd oluna bilər. E-kitabxanaların strukturunu ümumi formada analiz etsək əsasını istifadəçi, resurs və sistem təşkil etməklə informasiya sistemi olduğunu göstərmək olar. E-kitabxananın yaradılması prosesində onun strukturunu təşkil edən elementlər kimi isə texniki, riyazi və proqram, informasiya, təşkilatı və hüquqi təminatları qeyd olunmalıdır.

E-kitabxana sistemində toplanan loq-verilənlər müraciət olunan kontent, sorğular və istifadəçilərin davranışları haqqında əsas informasiya mənbəyi hesab olunur [2]. Hal-hazırda loq-verilənlərin analizi, resursların və istifadəçi profillərinin analizi üçün vasitələrin (alətlərin) yaradılmasına xüsusi diqqət yetirilir [1, 3, 4]. Hazırda e-kitabxana sistemində toplanan loq-faylların müfəssəl analizi və istifadəçi profillərinin öyrənilməsi sahəsində effektiv həllərin, metod və alqoritmlərin işlənilməsinə dair aparılan tədqiqatlar

artmaqdadır. Bu isə e-kitabxana xidmətlərinin səmərəliliyinin qiymətləndirilməsinə və eləcə də istifadəçi profillərini öyrənməklə əks əlaqə mexanizmlərinin işlənməsi baxımından böyük əhəmiyyət kəsb edir.

Məqalədə e-kitabxana sistemində toplanan loq-verilənlərin analizi məsələsinə baxılır. Loq-verilənlər analiz olunmaqla sorğuların, istifadəçilərin profilinin öyrənilməsi və eləcə də, kəmiyyət və keyfiyyət xarakteristikalarının qiymətləndirilməsi üsulları araşdırılır.

II. ELEKTRON KİTABXANA SİSTEMİ VƏ LOQ-VERİLƏNLƏR

E-kitabxana sisteminin yaradılması ideologiyası özündə İKT-nin tətbiqinin elmi-nəzəri məsələlərini cəmləşdirərək istifadəçiyönümlü informasiya mübadiləsi sisteminin yaradılması ideyasını təqdim edir. Bu ideyanın həyata keçirilməsi isə arxitektura prinsiplərin müəyyənləşdirilməsini ön plana çəkir. Bu prinsiplər kimi e-kitabxana sisteminin digər informasiya sistemləri ilə inteqrasiya etməsi imkanı, standartların müəyyənləşdirilməsi, resursların formatları və yaradılması göstərilə bilər.

E-kitabxana sistemləri mürəkkəb strukturlu sistemlər olmaqla, əsasən xidmət yönümlü arxitektura əsaslanır və böyük həcmdə e-resursu idarə etməyə imkan verir [2]. Qabaqcıl texnologiyalara əsaslanan ilk e-kitabxana sistemlərinin yaradıldığı vaxtdan müraciət verilənlərinin (loq-verilənləri) toplanması və analizi məsələləri daim diqqət mərkəzində saxlanılırdı. Toplanan loq-verilənləri sistemin özünün idarə olunması və xüsusilə istifadəçilər tərəfindən axtarış sistemlərinin istifadəsinin monitorinqini aparmağa imkan verirdi. Kataloq verilənlərinin axtarışı və onlara çıxış imkanının təmin edilməsi elektron kataloq (Online Public Access Catalogue) adlandırılır [1, 2].

Ümumilikdə qeyd etmək lazımdır ki, müxtəlif kateqoriyalı e-kitabxana sisteminin istifadəçilərinin hamısı üçün göstərilən xidmətlərin səmərəliliyi və e-resurslardan istifadə mühüm əhəmiyyət kəsb edir. Bu baxımdan loq-verilənlər e-kitabxana sisteminin səmərəliliyini və göstərilən xidmətlərin qarşılıqlı əlaqəsinin effektivliyinin qiymətləndirilməsində mühüm rol oynayır. Nəzərə almaq lazımdır ki, loq-faylların analizi fonunda istifadəçi profilləri və e-xidmətlər haqqında təfərrüatlı məlumatın alınması prosesində əsas mərkəzi mövqedə istifadəçiyönümlü xarakteristikalar yer alır. İstifadəçi profillərinin analizi nəticəsində əldə olunan bilik əks əlaqə

vasitəsi kimi elə istifadəçiləri özləri üçün həllərin tapılmasında naviqasiya rolunu oynayır [5,6].

Son onillikdə loq-faylların analizi axtarış sistemlərinin istifadəçilərini identifikasiya etməyə imkan verən əsas tədqiqat istiqamətlərindən birinə çevrilmişdir [6,7]. Loq-fayllar sistemin fəaliyyətini, sistem informasiyasını və istifadəçilərin davranışları haqqında informasiyanı özündə birləşdirir: müraciət olunan kontent, müraciət tarixi və vaxtı, istifadəçinin kompüterinin IP-ünvanı, istifadəçinin brauzerinin adı, müraciət olunan səhifənin URL ünvanı və istifadəçi istinadları və s. Loq-fayllar analiz olunmaqla veb-səhifələrin və onlara müraciət edən istifadəçilərin identifikasiyası, kəmiyyət və keyfiyyət xarakteristikalarının müəyyən edilməsi və eləcə də, müxtəlif meyarlara görə qiymətləndirmə mexanizmləri işlənib bilər.

III. LOQ-FAYLLARIN ANALİZİ

İstifadəçilərin müraciət xarakteristikalarının avtomatik öyrənilməsi prosesi əsasən assosiativ qaydaların axtarılmasını, naviqasiya yollarını, identifikasiya, klassifikasiya və s. ehtiva edir [7]. Bu məsələlərin həll edilməsi üçün loq-fayllarda saxlanılan məlumatlardan istifadə etmək olar. Serverlərdə avtomatik yaradılan və jurnallarda qeyd olunan məlumatlar (loq-verilənlər) toplanılır. Hər bir səhifəyə müraciəti (link) olan istinad jurnalları, brauzer jurnalları, istifadəçilərin qeydiyyat və anket məlumatları da (onlarda CGI (*Common Gateway Interface*) skriptlər vasitəsilə toplanır) informasiya mənbələri hesab olunur [4].

Loq-fayl müxtəlif kombinasiyaları (formatları) olsa da onlar arasında ən önəmlisi müraciət loqları (access log) hesab olunur [4, 8]. Belə ki, bu loq-fayllarda hər bir müraciətin HTTP-sorğuları saxlanılır habelə, qrafik təsvirə, CGI proqrama, audioklipə və ya digər obyektlərə müraciətin uğurlu olub-olmamasından asılı olmayaraq bu jurnalda saxlanılır. Lakin HTTP-sorğu və veb-səhifəyə sorğu eyni bir mənanı ifadə etmir. Demək olar ki, bilavasitə veb-səhifənin özü bir neçə qrafik təsvirə malik ola bilər. Bu halda istifadəçi belə bir səhifəyə baxanda veb-server yalnız həmin səhifənin HTTP-sorğusunun deyil, həm də təsvirlərinin HTTP-sorğularına xidmət edir. Beləliklə, səhifəyə müraciət zamanı müraciət jurnalında bir yox, qrafik təsvirlərin sayı qədər qeydiyyat yazısı əlavə olunacaq.

Qeyd edək ki, loq-fayllardan biliklərin aşkarlanması üsulları son illərdə daha çox populyarlaşmağa başlamışdır. Bu sahədə yaxşı göstərici olaraq elmi-tədqiqat işlərinin sayının artmasını göstərmək olar. E-kitabxana infrastrukturunun sürətli inkişafı, təklif olunan xidmətlərin, istifadəçilərin sayının kəskin artmasını nəzərə alsaq müvafiq proqram vasitələrinə böyük tələbat vardır. Tələbatı ödəmək və bu sahədə vacib problemləri həll etmək üçün intellektual analiz metodları bir sıra məsələləri həll etməyə kömək edə bilər. Belə məsələlərə istifadəçi profilinin öyrənilməsini, sorğuların analizi, tranzaksiyaların identifikasiyasını və s. nümunə göstərmək olar [6, 9, 10].

Qeyd edək ki, Web mining metodları tətbiq olunmaqla şəbəkədə lazımi informasiyanın axtarışı və əldə olunmuş resursların emalı üçün çox güclü instrumental vasitələr yaradıla bilər [8, 10]. Bu nöqteyi-nəzərdən də, müxtəlif təyinatlı intellektual sistemlərin yaradılması məqsədəuyğun hesab olunur. Məsələn, loq-faylların analizi veb-istifadəçilər haqqında verilənlərin analizi istifadəçinin profilinin

öyrənilməsi və onların daha çox hansı resurslara maraq göstərdiklərini təhlil etməyə imkan verir [3, 7].

İstifadəçilərin identifikasiyası üçün 3 əsas üsul istifadə olunur [4, 7].

- İstifadəçinin kompüterinin IP-ünvanına görə. Bu üsul digərləri ilə müqayisədə maksimal çatışmazlıqlara malikdir. Bu çatışmazlıqlar hər şeydən əvvəl onunla müəyyən olunurlar ki, veb-səhifəyə eyni bir IP-ünvanla bir neçə istifadəçi gələ bilər. Məsələn, proksi-server vasitəsilə müraciətlər;
- *Cookies* fayllarına görə. Brauzer vasitəsi ilə müraciət zamanı brauzerlə server arasında mübadilə verilənləri istifadəçinin kompüterində saxlanılır. Bu üsulun tətbiqi zamanı əsas çatışmazlıqlar ona görə yaranır ki, *cookies* faylları məhz konkret insanı deyil, istifadəçinin brauzerini identifikasiya edir;
- İstifadəçilərin məcburi qeydiyyatı zamanı. Bu halda veb-səhifəyə müraciət və ya onun xidmətlərindən birini istifadə etdikdə istifadəçi öz adını və parolunu daxil edir və sistem bütün seans müddətində birqiymətli olaraq onu identifikasiya edə bilər. Bu üsul istifadəçilər və onların təkrar müraciətləri hesablandıqda özündə daha az xəta daşıyır, amma, təəssüf ki, ancaq ayrı-ayrı xidmətlərə tətbiq oluna bilər.

Sayta müraciət edən istifadəçilərin özlərini aparması, davranışı haqqında informasiyanın yığılmasının əsas üsulları analizatorlar və serverin loq-fayllarının analizində alınan statistikadan istifadədir [6, 7, 9].

Aşağıda veb-səhifəyə müraciətlərə görə statistik informasiyanın yığılması və analizi sisteminin hansı məlumatları yığılmalı olduğu göstərilmişdir [6, 9]:

- Bütövlükdə sayta görə, eləcə də hər bir səhifəyə görə hostların (unikal IP-ünvanların) və hitlərin (səhifələrin yüklənməsinin ümumi sayı) miqdarı.
- Müraciət edənlərin istifadə etdikləri əməliyyat sistemlərinə və brauzerlərə görə statistika.
- IP-ünvanlar üzrə informasiya və hər bir istifadəçinin sayta sonuncu müraciət vaxtı.
- Sayta axtarış sistemləri (Rambler, Google, Yahoo və s.) ilə istifadəçilərin sayı haqqında informasiya.
- Sayta istinadların yerləşdirildiyi səhifələrdən sayta müraciətlərin sayı.

Serverdə toplanan loq-verilənlərin analizi e-kitabxananın fəaliyyəti ilə bağlı məlumatlar verməklə kəmiyyət və keyfiyyət xarakteristikalarının formalaşdırılmasına imkan verir. E-kitabxana sistemində e-kataloq mühüm rol oynayır. Bununla yanaşı, dizayn məsələləri, struktur, kontent, materialın təqdim olunması forması, təqdim olunan materiallar və xidmətlər, naviqasiya, məlumatların operativ yenilənməsi, veb-səhifələrə, kataloqlara istifadəçilərin əlyətərliyi və s. amillərdə nəzərə alınmalıdır. Təcrübədə e-xidmətlərdən istifadə edən istifadəçilər üçün çoxvariantlı testləşdirmədən istifadə oluna bilər ki, bu da real zamanda istifadəçilərin davranışları, profili haqqında məlumat almağa imkan verir.

Beləliklə loq-faylların analizi, veb-analitika metodlarından istifadə e-kitabxanaların fəaliyyətini qiymətləndirməklə yanaşı, onların maraq dairələri haqqında müxtəlif statistikalar toplamaqla istifadəçilərin müraciətlərinin sayını izləməyə, istifadəçi auditoriyasını müəyyənəlməyə, yeni istifadəçilərin cəlb olunmasına, göstərilən e-xidmətlərin səmərəliliyinin artırılmasına imkan verir.

IV. NƏTİCƏ

Hazırda e-kitabxana sistemlərinin səmərəliliyinin qiymətləndirilməsində loq-faylların istifadəsi mühüm tədqiqat istiqamətlərindən biri hesab olunur. Qeyd etmək lazımdır ki, e-kitabxana sistemlərində toplanan loq-verilənlərin analizi yalnız informasiya resurslarının yaradılmasının və istifadəsinin kəmiyyət göstəriciləri ilə məhdudlaşmır, eləcə də, keyfiyyət xarakteristikalarına diqqət yetirilməli və bu sahədə intellektual analiz vasitələri, veb-analitik həllər inkişaf etdirilməlidir.

Məqalədə e-kitabxana sistemində toplanan loq-faylların analizi məsələsi araşdırılır. Loq-fayllar analiz olunmaqla sorğuların, istifadəçilərin profilinin öyrənilməsi və eləcə də, müxtəlif meyarlara görə e-kitabxanaların fəaliyyətinin qiymətləndirilməsi üsulları tədqiq olunur.

ƏDƏBİYYAT

- [1] H. Han, W. Jeong and D. Wolfram, “Log analysis of academic digital library: user query patterns,” Proc. of the iConference, 2014, pp. 1002-1008.
- [2] M. Agosti, F. Crivellari, G. M. Di Nunzio, “Evaluation of digital library services using complementary logs,” Proc. of the SIGIR'09, 2009, pp. 32-35.
- [3] R. Əliquliyev, F. Yusifov, “Web-serverlərdə toplanan statistik verilənləri analiz etməklə istifadəçi profilinin yaradılması,” AMEA-nın xəbərləri, fizika-riyaziyyat və texnika elmləri seriyası, 2007, №2, s. 144-148.
- [4] Ф. Юсифов, “Извлечение знаний из Интернет с использованием лог-файлов,” Телекоммуникации, 2006, №8, с. 13-18.
- [5] X. Wanga, A. Abraham, K. A. Smitha, “Intelligent web traffic mining and analysis,” Journal of network and computer applications, 2005, vol. 28, pp. 147-165.
- [6] R.M. Alguliyev, R.M. Aliguliyev, F.F. Yusifov, “Automatic identification of the interests of web users,” Automatic control and computer sciences, 2007, vol. 41, No 6, pp. 320-331.
- [7] H. Liu, V. Keselj, “Combined mining of web server logs and web contents for classifying user navigation patterns and predicting users’ future requests,” Data and knowledge engineering, vol. 61, no 2, 2007, pp. 304-330.
- [8] R. Iváncsy, I. Vajk, “Different aspects of web log mining,” Proc. of the 6th International symposium of hungarian researchers on computational intelligence, 2005, pp. 413-424.
- [9] A. Abraham, “Business intelligence from web usage mining”, Journal of information & knowledge management, 2003, vol. 2, pp. 375-390.
- [10] J. Srivastava, P. Desikan, V. Kumar, “Web mining-accomplishments and future directions,” Proc. of the NSF workshop on next generation data mining (NGDM), 2002, pp. 51-61.