

AzScienceNet Şəbəkəsində İnternet-Resurslardan İstifadənin Monitoringi Problemləri

Yadigar İmamverdiyev

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
yadigar@lan.ab.az

Xülasə — İnternet-resurslardan istifadə edilməsi təşkilatlarda əməkdaşların əsas iş alətinə çevrilib. İş vaxtı ərzində əməkdaşlar çoxsaylı İnternet-resurslara müraciət edirlər və bəzən bu resursların heç də hamısı bilavasitə iş məqsədləri ilə əlaqədar olmur. Analizlər göstərir ki, İnternet-resurslardan istifadənin effektivliyi məsələsi global informasiya cəmiyyətində bütün təşkilatları maraqlandırır, bir sıra hüquqi, sosial və etik problemləri əhatə edir. İnternetdən istifadəyə müəyyən maneələr yaratmaqla bu məsələni effektiv həll etmək mümkün deyil və ən məqbul yanaşma mövcud qanunvericilik əsasında təşkilatın İnternetdən istifadə siyasətini müəyyən etmək və onun əsasında İnternet-resurslardan istifadənin monitorinqini həyata keçirməkdir. Bu işdə təşkilatlarda İnternet-resurslardan istifadə ilə bağlı problemlər və onların həllinə mövcud yanaşmalar analiz edilir, AzScienceNet şəbəkəsində İnternet-resurslardan istifadənin monitorinqi məsələləri üçün verilənlərin intellektual analizi metodlarının tətbiqi imkanları araşdırılır.

Açar sözlər — İnternet resursları; monitorinq; tematik profil; vebdən istifadənin intellektual analizi

I. GİRİŞ

İnternet-resurslardan istifadə edilməsi təhsil, elm, dövlət və kommersiya təşkilatlarının fəaliyyətində səmərəli iş alətinə çevrilib. Əksər hallarda İnternetdən istifadə zamanı təşkilat əməkdaşlarının fəaliyyətinin aparıcı motivləri yeni bilik əldə etmək və əsas məqsədləri isə informasiya axtarışı olur. Lakin bəzi əməkdaşlar müxtəlif şəxsi məqsədlər üçün İnternet “gəzintisində” iş vaxtının əhəmiyyətli hissəsini sərf edirlər, nəticədə təşkilatın ümumi əmək məhsuldarlığı aşağı düşür. Üstəlik, əməkdaşların İnternetdə “əyləncəsini” də təşkilat ödəyir. Dünyada iş vaxtı əməkdaşların İnternetdən işdənənar məqsədlərlə istifadə etməsi hadisəsi kifayət qədər geniş yayılıb və bu hadisə üçün hələ 2000-ci illərdən xüsusi termin də – kibersleking (ing. cyberslacking) işlədilir [1]. Eyni zamanda, əməkdaşların İnternetdən tənzimlənməyən istifadəsi ciddi informasiya təhlükəsizliyi təhdidləri də yarada bilər, o cümlədən qanunvericiliklə qadağan olunmuş İnternet-resurslardan istifadə edilməsi və istifadəçilərin arzuolunmaz kontentlə təmasları da baş verə bilər.

AzScienceNet elm-kompüter şəbəkəsi Azərbaycan Milli Elmlər Akademiyasının (AMEA) institut və təşkilatlarının fəaliyyətinin səmərəli təşkilinə, alim və mütəxəssislərin dünya elmi mühitinə inteqrasiya olunmasına, dünyanın tanınmış beynəlxalq elmi bazalarından geniş istifadəsinə imkan yaradılması məqsədilə fəaliyyət göstərir. AzScienceNet AMEA institut və təşkilatlarını elmi-tədqiqat, elmi-praktiki və tədris

məsələlərinin həyata keçirilməsi üçün zəruri olan müasir şəbəkə xidmətləri ilə təmin edir. Hazırda AzScienceNet şəbəkəsi vasitəsilə İnternetə qoşulan istifadəçilərin sayı 5000-dən çoxdur [2].

AzScienceNet-də şəbəkə xidmətlərinin keyfiyyət göstəricilərinin yaxşılaşdırılması, şəbəkə resurslarının daha səmərəli idarə edilməsi və istifadə olunması, informasiya təhlükəsizliyinin təmin olunması vəziyyətinə nəzarətin həyata keçirilməsi üçün istifadəçilərin İnternet-resurslardan istifadəsinin operativ monitorinqi həyata keçirilir.

Bu işdə elmi ədəbiyyatın analizi əsasında təşkilatlarda İnternet-resurslardan istifadə edilməsinin monitorinqi məsələsinə mövcud yanaşmalar araşdırılır, monitorinqin effektivliyini artırmaq üçün veb-resurslara müraciətlərin loq-fayllarında toplanmış verilənlərə intellektual analiz üsullarının tətbiqi məsələsinə baxılır.

II. TƏŞKİLATLARDA İNTERNETİN İSTİFADƏSİ İLƏ BAĞLI PROBLEMLƏR

İnternetə idarə olunmayan, nəzarətsiz giriş kollektivdə əmək məhsuldarlığını xeyli aşağı salır, çünki iş vaxtının müəyyən hissəsi işdənənar resurslara sərf olunur [3]. Qeyd edək ki, iş yerlərində smartfonların həddindən artıq istifadəsi də əmək məhsuldarlığının aşağı düşməsinə səbəb olur [4].

Əməkdaşlar sürət böyük olduğu üçün işdə İnternetə evdəkindən daha tez-tez müraciət edirlər. IDC şirkətinin məlumatına görə, iş yerində İnternetdən istifadənin 30-40%-i işlə əlaqədar deyil [3] və təşkilatların əməkdaşları iş vaxtının təxminən 1/3-ni onların işi ilə əlaqədar olmayan məqsədlərlə İnternetdə keçirirlər (şəbəkə oyunları, internet mağazalar, informasiya axtarışı). Əməkdaşların 30%-i işdə idman yarışlarını onlayn izləyir, 24%-i onlayn alış-verişi işdə edirlər. Onlayn birja satışlarının 92%-i iş zamanı iş yerindən edilir.

İnternetdən nəzarətsiz istifadənin daha bir nəticəsi – şəbəkədə sürətin (buraxılış qabiliyyətinin) aşağı düşməsidir. Statistika görə, təşkilatların əməkdaşlarının 44%-i korporativ resurslardan videoya baxmaq, audioyazıları dinləmək, şəbəkə oyunları oynamaq, böyük həcmli faylları yükləmək (filmlər, musiqi, şəkillər və s.) üçün istifadə edirlər. Nəticədə şəbəkəyə əhəmiyyətli yük düşür.

İnternet zərərli proqramların (virus, troyan və s.) yayılmasının əsas mənbəyidir. Hazırda zərərli proqramların yayılması üçün İnternet (e-poçt, veb-saytlar, sosial şəbəkələr) geniş istifadə edilir.

İnternet təşkilatın lokal şəbəkələrinə, serverlərinə və kompüterlərinə hücumların həyata keçirilməsi kanalıdır. İnternet-resursların çoxunda müxtəlif proqram – JavaScript, Flash, ActiveX və s. kodları olur. Bədiyyətilər korporativ şəbəkələrə və istifadəçi kompüterlərinə hücumu təşkil etmək üçün bu kodlardan istifadə edə bilirlər [5].

İnternet konfidensial informasiyanın sızmasının əsas kanallarından biridir. Məsələn, əməkdaşların pulsuz e-poçt xidmətlərindən istifadə etməsi təşkilatın informasiya resurslarının konfidensiallığına ciddi təhdidlər yarada bilər. Statistika əsasən, əməkdaşların 80%-i pulsuz e-poçt xidmətlərindən istifadə edir, lakin bu halın təşkilatda konfidensiallığın təmin edilməsinə yaratdığı təhlükədən xəbərdar olsalar da, təşkilatların yalnız dördü biri pulsuz e-poçtdan istifadəni qadağan edir.

III. TƏŞKİLATLARDA İNTERNETDƏN İSTİFADƏ PROBLEMİNİN HƏLLİNƏ YANAŞMALAR

Veb-resursların sürətli inkişafının hələ lap başlanğıcında əməkdaşların İnternetdən istifadəsinin effektiv tənzimlənməsi təşkilatlarda problemə çevrilmişdi. Təşkilatlar müxtəlif nəzarət mexanizmlərindən istifadə etməyə çalışırdılar [3,6-9].

Bəzi təşkilatlarda İnternetdən işdənkənar istifadə üçün cərimələr tətbiq edilirdi. Cərimələrin məbləği əməkdaşın şəxsi məqsədlər üçün İnternetdə sərf etdiyi vaxtdan və şirkətin əməkdaşa ödədiyi iş vaxtının dəyərindən asılı idi. Əməkdaşların İnternetdən istifadəsinin detallı uçotu aparılır (hansı kontentə baxılıb, hansı məlumat, hansı istifadəçi tərəfindən, nə zaman yüklənib və s.) və bunun əsasında rəhbərlik işdən çıxarılmada da daxil olmaqla inzibati tənbeh üsullarından istifadə edirdi [3].

KLegal hüquq şirkəti ilə Personnel Today jurnalı tərəfindən 2002-ci ildə aparılmış tədqiqat göstərmişdi ki, İnternetin və e-poçtun işdənkənar məqsədlər üçün istifadə edilməsi əməkdaşların işdən çıxarılmasının ən geniş yayılmış səbəbidir. Tədqiqatlar Böyük Britaniyada 212 şirkəti əhatə edirdi. Hər on şirkətdən birində rəhbərlik İnternetə çıxışı tamamilə bağlamağa məcbur olmuşdu, təşkilatların 30 %-ində iş zamanı İnternetin istifadəsi ciddi məhdudlaşdırılmışdı [10].

Aydındır ki, cərimələr və işdən çıxarılmalar heç də ən effektiv tədbir deyil (texniki dəstək xidmətindən alınmış siyahıların daim nəzərdən keçirilməsi zərurəti də rəhbərliyə əlavə yükə çevrilir və tədricən lazımsız olur).

Coker B. tərəfindən 2011-ci ildə təsadüfi seçilmiş 700 ofis əməkdaşı arasında aparılmış sorğuda bəzi maraqlı qanunauyğunluqlar aşkarlanmışdı. Sorğunun nəticələrinə görə, İnternetdən istifadəyə iş vaxtının 12%-dən çoxu sərf edilməkdə İnternetdən istifadə əmək məhsuldarlığı ilə müsbət korrelyasiyaya malik olur. İş yerində İnternetdən istifadə edən əməkdaşların əmək məhsuldarlığı iş yerində İnternetdən istifadə etməyən həmkarları ilə müqayisədə 9% artır [11]. Lakin vebdən həddindən artıq istifadə əmək məhsuldarlığının aşağı düşməsinə səbəb olur, çünki əməkdaşlar tapşırılmış işi planlaşdırılmış vaxt ərzində başa çatdırma bilmirlər.

Analizlər göstərir ki, təşkilatlarda İnternetdən təhlükəsiz və məhsuldar istifadə probleminin həllinə iki yanaşma mövcuddur:

- zərurət olmadan İnternetin istifadəsinin qadağan edilməsi (“radikal” yanaşma);
- istifadəçiyə “qadağan edilməyən hər şeyə icazə verilib” prinsipi ilə hərəkət etməyə icazə verilməsi. İstifadəçi İnternet-resurslardan istifadə edir, lakin onun hərəkətlərinə nəzarət (monitorinq) edilir.

Beləliklə, təşkilatlarda İnternet-resurslardan effektiv istifadə edilməsi probleminin həllinə kompleks yanaşma vacibdir: normativ-hüquqi sənədlərin işlənməsi və qəbul edilməsi; istifadəçilərin maarifləndirilməsi, texniki nəzarət və inzibati tədbirlər. Məsələnin texniki həlli – İnternetdən istifadə üzrə reqlamentin müəyyən edilməsi, internetə çıxışın tənzimlənməsi, hər bir əməkdaşın şəbəkə resurslarından istifadəsinin monitorinq edilməsindən ibarətdir. Nəzarətin konkret reallaşdırılması təşkilatın daxili siyasətindən və korporativ mədəniyyətindən asılıdır:

- iş yerindən İnternet şəbəkəsinə giriş məhdudlaşdırılır, yalnız fasilə zamanı və işdən sonra əməkdaşlara İnternetə sərbəst girişinə icazə verilə bilər;
- iş vaxtında idman, xəbər və əyləncə saytlarına giriş qadağan olunur;
- müəyyən veb-səhifələrə və xidmətlərə girişə icazə verilir;
- kənar e-poçt sistemlərinə girişə məhdudiyət qoyulur;
- qeyri-məhdud girişə icazə verilir, lakin monitorinq aparılır;
- hər bir istifadəçi üçün konkret giriş növünə icazə verilir;
- gün ərzində istifadə edilən trafikənin həcminə məhdudiyət qoyulur.

İnternet-resurslardan istifadəyə çevik nəzarəti təmin etmək üçün təşkilatda İnternet-resurslardan istifadə siyasətinin işlənməsi və həyata keçirilməsi zəruridir. Bu siyasətin həyata keçirilməsi istifadəçilərin fəaliyyətinin monitorinqini tələb edir və kontent-filtrasiya sistemlərinin istifadəsi bu məsələnin həlli üçün effektiv vasitə hesab edilir. Bu sistemlər istifadəçilərin İnternet-resurslara tematik girişini təmin etməyə yönəlir və İnternet-resursların analizinə və kateqoriyalar üzrə təsnif edilməsinə əsaslanır (URL-filtrasiya). Kateqoriyalar üzrə təsnif edilmiş resursların keyfiyyətli bazası əsasında bu sistemlər ingilis dilli resurslarla yaxşı işləyə bilər. Digər dillər üçün belə sistemlərdə URL-bazasının natamamlığı və operativ yenilənməməsinə görə problemlər yarana bilər.

Qeyd edək ki, AzScienceNet şəbəkəsində istifadəçilərin İnternet-resurslardan istifadəsinə hər hansı bir məhdudiyət yoxdur. AzScienceNet şəbəkəsinin əsas vəzifəsi günün 24 saati ərzində istifadəçilərinin global İnternet şəbəkəsinə yüksək sürətli, təhlükəsiz çıxışını təmin etməkdir və bu vəzifənin səmərəli həyata keçirilməsi üçün “AzScienc

eNet şəbəkəsindən istifadə siyasəti” [12] və “AzScienceNet şəbəkəsində informasiya təhlükəsizliyi siyasəti” [13] qəbul edilmişdir. Bu siyasətlərin həyata keçirilməsi vəziyyətini qiymətləndirmək üçün AzScienceNet Monitoring və İnformasiya Təhlükəsizliyi Xidməti tərəfindən vaxtaşırı monitoring və analiz işləri aparılır.

IV. İNTERNET-RESURLARDAN İSTİFADƏNİN MONİTORİNQİ ÜÇÜN ALƏTLƏR

İnternet-resurslardan istifadə vəziyyətini izləməyə və bir sıra statistik hesabatlar almağa imkan verən xüsusi proqram təminatı istifadə edilir. Fərdi kompüterlər (NetPolice, iCencor və s.) və təşkilatın korporativ şəbəkəsi üçün həllər və proqram təminatı (UserGate Web Filter, ICS-Ideco, Internet Access Monitor, Websense Web Security Gateway) vardır. Veb-brauzerlərdə də kontent-filtrasiya imkanları mövcuddur.

İnternet-resurslardan istifadənin monitoringi sistemlərində aşağıdakı avtomatik məhdudiyətlər də ola bilər:

- müəyyən saytlara girişin qadağan edilməsi;
- müəyyən İnternet-protokolların qadağan edilməsi;
- İnternet-trafikin antivirusla yoxlanması və viruslu faylı qəbul etməkdən avtomatik imtina;
- faylların müəyyən tiplərinin yüklənməsinin qadağan edilməsi;
- günlük, həftəlik və aylıq limitlərə çatdıqda istifadəçilərin və (ya) istifadəçi qruplarının ödənişli resurslardan avtomatik ayrılması.

Bir sıra həllərdə ISS (Internet Security Systems) şirkətinin təqdim etdiyi İnternet-resursların kateqoriyalar bazası istifadə edilir, bu baza daha tam və aktual URL-baza hesab edilir. URL-bazada milyonlarla veb-saytın ünvanı var və hər gün yenilənir. Veb-saytın kateqoriyası kontent-filtrasiyanın ilkin mərhələlərində müəyyən edilir və çox vaxt tələb etmir, kifayət qədər effektivdir və verilənlərin emal vaxtını azaldır.

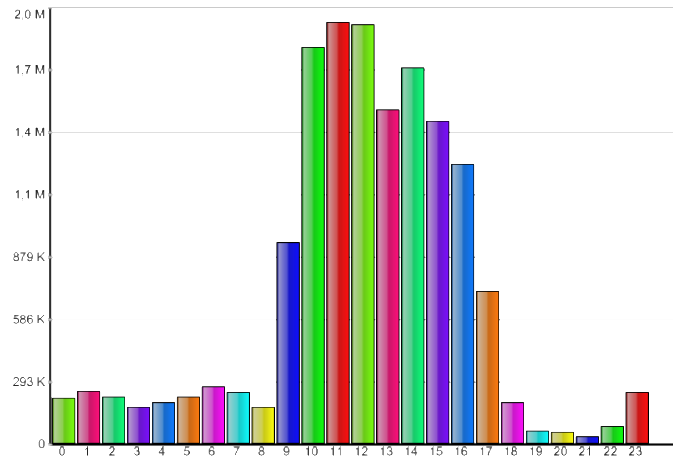
Müasir təşkilatların İnternetə çıxışının tipik üsulu xüsusi-proqramlardan – proksi-serverlərdən istifadə olunmasıdır. Proksi-server İnternet bağlantısını bütün əməkdaşlar arasında bölməyə imkan verir. Bu proqramın yaratdığı loq-faylı analiz etməklə kimin nə vaxt, hansı sayta müraciət etməsi haqqında məlumat əldə etmək olar.

AzScienceNet şəbəkəsində istifadəçilərin veb saytlara müraciətlərinin statistikasını əldə etmək üçün Squid proksi-serverinin loq fayllarının analizi üçün WebSpy Vantage proqram təminatı istifadə edilir. Proqram aşağıdakı hesabatları almağa imkan verir:

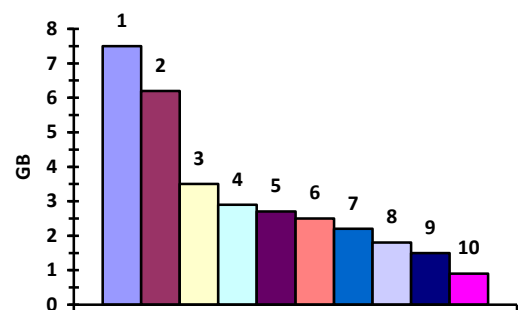
- trafikə verilənlərin tipinə görə (şəkil, video, audio, mətn) paylanması;
- trafikə veb-saytların kateqoriyaları üzrə paylanması;
- trafikə istifadəçilər, istifadəçilər qrupu (təşkilatlar), protokollar üzrə paylanması;

- trafikə saatlar, həftənin günləri, aylar və tarixlər üzrə paylanması;
- trafikə istifadəçinin müraciət etdiyi saytlar üzrə paylanması;
- istifadəçinin İnternetə giriş və çıxış zamanları.

Göründüyü ki, proqram təminatı loq-faylı əsasında sadə statistik hesabatlar almağa imkan verir. Şəkil 1-də AMEA təşkilatlarından birində istifadəçilərin müraciətlərinin saatlar üzrə paylanması göstərilir (ordinat oxunda K – min, M – milyon bildirir, müraciətlərin sayı olaraq hitlərin sayı götürülüb). Şəkil 2-də isə həmin təşkilatda bəzi istifadəçilərin bir həftə ərzində İnternetdən yüklədikləri informasiyanın həcmi (qıqabaytlarla –GB) göstərilmişdir, istifadəçilər adsızlaşdırılıb və rəqəmlərlə göstərilir. Loq-fayllarda toplanmış informasiyanın daha dərinəndən analizi üçün verilənlərin intellektual analizi metodlarından istifadə edilməlidir. Bu məqsədlə növbəti bölmədə vebdən istifadənin intellektual analizi metodlarının qısa icmalına baxılır. Qeyd edək ki, bu işdə veb-saytların kontentinin analizinə baxılmaz (Deep Content Inspection).



Şəkil 1. İstifadəçilərin müraciətlərinin saatlar üzrə paylanması



Şəkil 2. İnternetdən ən çox informasiya yükləyən istifadəçilər

V. VEB-DƏN İSTİFADƏNİN İNTELLEKTUAL ANALİZİ METODLARI

Veb-də toplanmış verilənlərin intellektual analizi metodları (ing. Web Mining) 1990-cı illərin ortalarından inkişaf edir və hazırda aktual tədqiqat istiqamətlərindən biridir [14,15]. Web Mining üç istiqamətə ayrılır: veb kontentin, veb strukturun və vebdən istifadənin intellektual analizi. Vebdən istifadənin intellektual analizi (Web Usage Mining, WUM) vebdən istifadə şablonlarını aşkarlamaq üçün istifadəçinin vebə qarşılıqlı əlaqəsi zamanı toplanmış verilənlərə intellektual analiz üsullarının tətbiqidir [16]. Analiz üçün veb məlumatları əsasən veb-serverin və ya proksi-serverin (bəzən veb-brauzerin) loq-fayllarından alınır. Vebdən istifadənin intellektual analizi veb-saytların dizaynını təkmilləşdirmək, istifadəçiləri cəlb etmək və daimi istifadəçilərə fərdi və adaptiv xidmətlər təklif etmək üçün istifadəçilərin davranışını aşkarlamağa və proqnozlaşdırmağa çalışır [17,18,19]. WUM yüklənmənin balanslaşdırılması və veb-trafikin effektiv idarə edilməsi, kiber-hücumlarla əlaqədar şübhəli aktivliyin aşkarlanması [20], istifadəçilərin müxtəlif qrupları üçün populyar mövzuların aşkarlanması, elm-tədqiqat prosesləri baxımından veb-saytların qiymətləndirilməsi üçün də istifadə edilə bilər.

AMEA İnformasiya Texnologiyaları İnstitutunda vebdən istifadənin intellektual analizi üzrə bir sıra tədqiqat işləri aparılmışdır. Məsələn, [21]-də veb-trafikin intellektual analizi metodlarına baxılır, trafik klaster analizi nəticəsində veb-istifadəçilərin aktivliyinin modelləşdirilməsi məsələsi həll edilir. [22]-də veb-istifadəçilərin maraqlarının avtomatik identifikasiyası məsələsinə baxılır. [23,24]-də veb-serverlərin loq-fayllarının analizi əsasında veb-istifadəçilərin klasterizasiyası modeli təklif edilir.

[25]-də istifadəçilərin müraciət etdikləri İnternet-resursları haqqında verilənlərin analizi əsasında qurulmuş tematik profillərin klasterizasiyası metodu təklif olunur. Məqalədə ilkin verilənlər, əlamətlərin qiymətləndirilməsi və seçilməsi prosesləri, aqlomerativ iyerarxik klasterizasiya metodu və k-means metodu istifadə edilməklə aparılmış eksperimentlər ətrafı təsvir olunur.

VI. TƏKLİF EDİLMİŞ YANAŞMANIN MAHİYYƏTİ

Vebdən istifadənin intellektual analizi üç mərhələdən ibarətdir: verilənlərin ilkin emalı, şablonun aşkarlanması və şablonun analizi.

Verilənlərin ilkin emalı mərhələsində proksi-serverin loq-faylından alınmış verilənlər əsasında hər bir istifadəçi üçün əlamətlər vektoru formalaşdırılır. Proksi-serverin loq-faylında hər bir müraciətin zamanı, müddəti, istifadəçinin IP-ünvanı, müraciət edilən veb-resursun IP-ünvanı və ya DNS-adı, İnternetdən alınmış informasiyanın həcmi haqqında məlumat yazılır. Beləliklə, loq-faylıdan intellektual analiz üçün zəruri olan verilənləri almaq olar. Şəkil 3-də Squid proksi-serverinin loq-faylından bir fraqment göstərilib.

```
1444776131.155 515 10.100.80.23 TCP_MISS/403 810 GET
http://data.icecentury.com/product/wifi/ - HIER_DIRECT/54.169.165.185
text/html1444776131.187 547 10.100.80.23 TCP_MISS/403 813 GET
http://data.icecentury.com/product/cleaner/ -
HIER_DIRECT/54.169.165.185 text/html1444776131.198 560
10.100.80.23 TCP_MISS/301 568 GET http://v.icecentury.com/upgrade? -
HIER_DIRECT/54.169.165.185 text/html 1444776131.244 528
10.100.80.23 TCP_MISS/302 250 GET
http://market.voga360.com/frontend/checkpc.htm? -
HIER_DIRECT/54.255.133.240 -
```

Şəkil 3. Squid loq-faylından bir fraqment

Əlamətlər vektorunu formalaşdırmaq üçün istifadəçinin hitlər axını seanslara bölünür. Seans – İnternetdə fasiləsiz aktivlik intervallarıdır. Seansları müəyyən etmək üçün bir sıra yanaşmalar var, bu işdə aşağıdakı iki sadə qayda istifadə edilir:

1) əgər iki hit arasındakı interval 20 dəqiqədən çox deyilsə, onda onlar bir seansa aiddir.

2) Seans axırınıcı hitdən sonra 20 dəqiqə davam edir.

Alınmış seanslar əsasında istifadəçilərin veb-resurslara müraciətlər xronologiyasını matrisə çevirmək olar. Matrisin sətirlərində istifadəçilər, sütunlarında əlamətlər verilir. Əlamətlər kimi veb-saytların tematik-kateqoriyaları götürülür (məsələn, ISS-in URL bazasına əsasən).

Konkret istifadəçi üçün əlamətin qiyməti olaraq bu əlamətin rast gəldiyi istifadəçi seanslarının sayı götürülür. Əlamətlərin qiymətini istifadəçinin hər bir tematik kateqoriyaya daxil olan veb saytlara müraciət tezlikləri və daxil olan trafik həcmi əsasında da hesablamaq olar. Modelin keyfiyyətini yaxşılaşdırmaq üçün əlamətlərin qiymətləri [0...1] intervalında normallaşdırılır.

Əlamətlərin qiymətləri hesablandıqdan sonra (feature engineering) modelin qurulması üçün ən informativ və statistik etibarlı əlamətlər seçilməlidir. Bu emal edilən informasiyanın həcmi azaltmağa, izafi təlimdən qaçmağa və bütövlükdə modelin keyfiyyətini yaxşılaşdırmağa xidmət edir. Yalnız bir istifadəçidə rast gəlinən əlamətlər qeyri-informativ hesab olunur.

Nəticədə yalnız informativ əlamətlərdən ibarət olan “istifadəçi/kateqoriya” matrisi alınır. Adətən, bu matrisin ölçüsü böyük olur, həm də bu matris seyrək də ola bilər (sparse matrix). Buna görə belə matrislərlə işləyə bilən maşın təlimi alqoritmlərindən istifadə edilməlidir. Alınmış matris klasterizasiya metodları ilə emal olunur [26]. Eksperimentlər üçün WEKA proqram təminatı istifadə edilir [27].

İstifadəçilərin tematik profilləri müəyyən edildikdən sonra veb-saytların tematik kateqoriyalarına uyğun çəkilər verməklə istifadəçiləri tematik profillərinə görə rəqləşdirməyə də mümkündür.

NƏTİCƏ

İstifadəçilərin veb-resurslara müraciətləri haqqında loq-fayllarda toplanmış verilənlərin intellektual analizi istifadəçilərin və istifadəçi qruplarının maraqları və seçimləri haqqında fikir yürütməyə, təşkilatda İnternet-resurslardan nə dərəcədə məqsədyönlü istifadə edilməsi barədə nəticə çıxarmağa imkan verir. İstifadəçi seçimlərinin klasterləri haqqında informasiya əməkdaşların işinin effektivliyini qiymətləndirməyə, istifadəçi qruplarının informasiya tələbatlarını müəyyən etməyə və İT-infrastrukturun inkişafına investisiyaları əsaslandırmağa imkan verir.

ƏDƏBİYYAT

- [1] W. Block, “Cyberslacking, business ethics and managerial economics,” *Journal of Business Ethics*, Vol. 33, No. 3, pp. 225-231, 2001.
- [2] AzScienceNet-in veb-saytı <http://azscience.net.az/az>
- [3] M. Anandarajan, C.A.Simmers. *Managing Web Usage in the Workplace: A Social, Ethical and Legal Perspective*. IGI Publishing Hershey, 2002.
- [4] T. Pitichat, “Smartphones in the workplace: Changing organizational behaviour, transforming the future,” *A Journal of Trans-disciplinary Writing and Research from Claremont Graduate University*, Vol. 3, No. 1, pp.1-11, 2013.
- [5] O. Слепов, А. Отт, “Контроль использования интернет-ресурсов,” *Информационный Бюллетень*, № 2 (141), 2005.
- [6] L. Qinyu, L. Xin, G. Anil, L. Long, “Workplace Management And Employee Misuse: Does Punishment Matter?” *Journal of Computer Information Systems*, Vol. 50, No. 2, pp. 49-59, 2009.
- [7] J. C. Ugrin, J. M. Pearson, “Exploring Internet Abuse in the Workplace: How Can We Maximize Deterrence Efforts?” *Review of Business*, Vol. 28, No. 2, pp.29-40, 2008.
- [8] H. Jiang, A. Tsohou, “Dual nature of personal web usage in workplace: Impacts, antecedents and regulating policies,” *Proc. of the 22nd European Conference on Information Systems (ECIS)*, pp.1-16, 2014.
- [9] P. Mahatanankoon, M. Anandarajan, M. Igbaria, “Development of a Measure of Personal Web Usage in the Workplace,” *Cyberpsychology & Behavior*, Vol 7, No 1, pp. 93-104, 2004.
- [10] Internet abuse is top reason for workplace discipline. <http://www.computerweekly.com/news/2240047296/Internet-abuse-is-top-reason-for-workplace-discipline>
- [11] B. L. Coker, “Freedom to surf: the positive effects of workplace Internet leisure browsing,” *New Technology, Work & Employment*, Vol. 26, No. 3, pp. 238-247, 2011.
- [12] R.Ələkbərov, “AzScienceNet şəbəkəsinin informasiya təhlükəsizliyi siyasəti haqqında,” Azərbaycan xalqının ümummilli lideri Heydər Əliyevin 90 illik yubileyinə həsr olunmuş “İnformasiya təhlükəsizliyi problemləri üzrə I respublika elmi-praktiki konfransı”nın əsərləri, səh. 126-127, 2015.
- [13] B.Nəbiyev, “AzScienceNet şəbəkəsində informasiya təhlükəsizliyinin monitoringi sistemi,” Azərbaycan xalqının ümummilli lideri Heydər Əliyevin 90 illik yubileyinə həsr olunmuş “İnformasiya təhlükəsizliyi problemləri üzrə I respublika elmi-praktiki konfransı”nın əsərləri, səh. 128-130, 2015.
- [14] Q. Zhang, R. Segall, “Web mining: a survey of current research, techniques and software,” *International Journal of Information Technology & Decision Making*, Vol. 7, No. 4, pp. 683-720, 2008.
- [15] K. Sharma, G. Shrivastava, V. Kumar, “Web mining: Today and tomorrow,” *Proc. Of the 3rd International Conference on Electronics Computer Technology (ICECT)*, Vol. 1, pp.399-403, 2011.
- [16] J. Srivastava, R. Cooley, M. Deshpande, P.-N. Tan, “Web usage mining: discovery and applications of usage patterns from Web data,” *ACM SIGKDD Explorations Newsletter*, Vol. 1, No. 2, pp. 12-23, 2000.
- [17] J. Pei, J. Han, B. Mortazavi, H. Zhu, “Mining Access Patterns efficiently from Web Logs,” *Proc. of the 4th Pacific-Asia Conf. Knowledge discovery and Data Mining (PAKDD'00)*, pp. 396-407, 2000.
- [18] B. Berendt, M. Spiliopolou, “Analysis of navigation behaviour in web sites integrating multiple information systems,” *VLDB Journal*, Vol. 9, No. 1, pp. 56-75, 2000.
- [19] M. Eirinaki, M. Vazirgiannis, “Web mining for web personalization,” *ACM Transactions on Internet Technology (TOIT)*, Vol. 3, No. 1, pp. 1-27, 2003.
- [20] D. Stevanovic, N. Vljajic, A. An, “Unsupervised Clustering of Web Sessions to Detect Malicious and Non-malicious Website Users,” *Procedia Computer Science*, Vol. 5, pp. 123-131, 2011.
- [21] P.M. Алыгулиев, Ф.Ф. Юсифов, Разработка модели для изучения активности пользователей Web с помощью анализа Web-трафика,” *Автоматика и вычислительная техника*, No. 4, С. 73-77, 2007.
- [22] P.M. Алгулиев, P.M. Алыгулиев, Ф.Ф. Юсифов, “Автоматическая идентификация интересов Web-пользователя,” *Автоматика и вычислительная техника*, Vol.41, No.6, с.32-47, 2007.
- [23] P.M. Алгулиев, Ф.Ф. Юсифов, “Кластеризация Web-пользователей на основе анализа посещаемости Web-сайтов,” *Телекоммуникации*, No: 2, С. 13-17, 2008.
- [24] Ф.Ф. Юсифов, “Извлечение знаний из Интернет с использованием лог-файлов,” *Телекоммуникации*, №8, с. 13–18, 2006.
- [25] К.Г., Паутов, Ф.А. Попов, “Метод кластеризации тематических профилей пользователей и его применение для анализа Интернет-трафика,” *Фундаментальные Исследования*, № 7 (часть 4), с.765-769, 2015.
- [26] J. Han, M. Kamber, *Data Mining: Concepts and Techniques*. 2nd edition. Morgan Kaufmann, 2006.
- [27] Weka 3: Data Mining Software in Java. <http://www.cs.waikato.ac.nz/ml/weka/>