

# Viki-Mühitdə Big Data Problemləri və Onların Həlli Yolları

İradə Ələkbərova

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

*airada.09@gmail.com*

**Xülasə**– Məqalədə viki-mühitdə toplanan və durmadan artan böyük həcmli verilənlərin yaratdığı problemlər analiz edilmişdir. Viki-mühitdəki Big Data ilə işləmək üçün zəruri şərtlər göstərilmiş, problemin həllində bəzi mövcud yanaşmalardan birgə istifadə təklif edilmişdir.

**Açar sözlər**– viki-mühit, viki-texnologiya, Vikipediya, big data, xromoqram, intellektual analiz, Map-Reduce.

## I. GİRİŞ

Müasir dövrdə Big Data (BD) problemləri İnternet şəbəkəsinin genişlənməsi və yayılması ilə daha da aktual olmuşdur [1]. İnternetin nəhəng layihələrindən olan Facebook, Twitter kimi sosial şəbəkələrdə, Google, Yahoo, Yandex kimi axtarış sistemlərində, Wikipedia, WikiaMapia, WikiTravel kimi açıq ensiklopediyalarda, müxtəlif forumlarda, bloqlarda və s. sistemlərdə hər saniyə artan informasiyanın saxlanması, strukturlaşdırılması və emalı ilə bağlı problemlər bu gün İKT mütəxəssislərini düşündürməkdədir. Yuxarıda adı çəkilən layihələrdə kontent əsasən İnternet istifadəçiləri tərəfindən daxil edildiyi üçün virtual məkanda informasiyanın sürətlə artmasının qarşısını almaq mümkün deyil.

Viki-texnologiyaları ilə idarə olunan Vikipediya virtual ensiklopediyası və onun törəmə layihələri (Vikikitab, Vikimənbə, Commons, Vikinövlər və s.) ümumilikdə, viki-mühit təşkil edir və onun nəhəng məlumat bazası hər saniyə yeni məqalə, xəbər, kitab, foto, audio və video-fayllarla zənginləşir [2, 3]. Məqalədə əsas məqsəd viki-mühitdə BD ilə bağlı problemləri müəyyən etmək, Vikipediyanın verilənlər bazasında (VB) toplanmış məlumatlar əsasında bir çox məsələlərin həllinin mümkünliyünü və verilənlərin səmərəli emalı üçün düzgün həll yolunu göstərməkdir.

## II. VİKİ-MÜHİTDƏ BÖYÜK VERİLƏNLƏRLƏ BAĞLI PROBLEMLƏR

Viki-mühit özündə milyonlarla veb-səhifələri birləşdirir. 2015-cü ilin oktyabr ayına olan məlumata görə, yalnız Vikipediya layihəsi 58 milyondan artıq qeydiyyatdan keçmiş istifadəçiyə və 36 milyona yaxın ensiklopedik məqaləyə [4], 30 milyona yaxın şəkil, audio və video-fayla malikdir [5]. 4 milyondan artıq məqaləyə malik ingilis dilindəki Vikipediyada yalnız məqalələrin ümumi həcmi 1,7 terabaytdır [6]. Viki-mühitdə toplanan bütün kontent isə onlarla terabayt informasiya təşkil edir və bu informasiya durmadan

artmaqdadır. Göndərilən sorğuları, yaradılan şablon və bot-proqramları da əlavə etsək, informasiyanın həddən artıq çox olmasını təsəvvür etmək çətin deyil. İstifadəçilərin çox olması və yüksək aktivliyi qoyulan məsələlərin həllində BD texnologiyalarının aktuallığını önə çəkir. BD-də olduğu kimi, viki-mühitdə də problemlər, əsasən verilənlərin həddən artıq çox olması və strukturlaşmaması ilə bağlıdır. Əsas problemlər aşağıdakılardır:

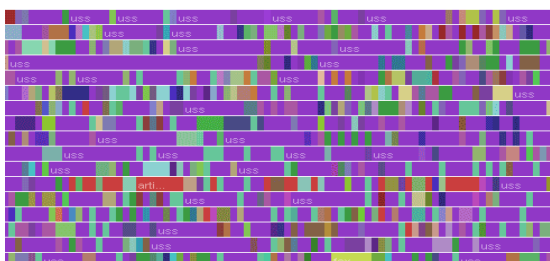
- *Verilənlərin ölçüsü.* Vikipediyanın VB-də verilənlərin həcmi baytlarla göstərilir. Hər bir viki-səhifədə, kateqoriyada, faylda, dilə və mövzuya görə ayrılmış layihələrdə toplanan verilənlərin həcmi onların BD kimi analiz olunub-olunmamasını müəyyən edir. Bu baxımdan, böyük həcmdə verilənlərin saxlanması üçün xüsusi şərtlər tələb olunur.
- *Verilənlərin müxtəlifliyi.* Viki-kontent müxtəlif tipdə ola bilər. Viki-səhifələrə mətn, səs, şəkil və video-fayllar yerləşdirmək mümkün olduğuna və viki-texnologiyaları əksər formatları dəstəklədiyinə görə verilənlər müxtəlif tipdə və strukturda toplanırlar. Bu verilənləri bir yerə toplayıb emal etmək mümkün deyil və analiz üçün əvvəlcə onları yararlı vəziyyətə gətirmək tələb olunur.
- *Verilənlərin sürəti.* Viki-serverlər brauzerdən daxil edilən bütün verilənləri (sorğular, mediafayllar, proqram kodları, mətn və s.) emal edərək yadda saxlayırlar. Nəzərə almaq lazımdır ki, Vikipediyaya daxil edilən hər bir simvol generasiya olunaraq bazada saxlanılır və veb-səhifələrdən silinsə belə, VB-da qalır və istənilən zaman istifadəçi tərəfindən bərpa edilir. Saniyədə milyonlarla informasiya saxlayan Vikipediyanın VB üçün heç bir informasiyanın silinməməsi zaman keçdikcə verilənlərin həddən artıq çoxalması deməkdir. Verilənlərin emalında sürət nə qədər çox olarsa, interaktivlik şərtləri daha yaxşı ödənilir.
- *Verilənlərin dəyişənliyi.* Viki-səhifələr açıqdır və hər an birbaşa brauzerdən dəyişdirilir. Baxılan səhifə yenilənmədən sonra artıq tam başqa strukturda və dizaynda təqdim oluna bilər. Səhifənin bütün köhnə versiyaları (yaranma tarixi, şərhlər və müəllif göstərilməklə) bazada saxlanılır. Bu faktorlar analiz zamanı verilənlərin idarə olunmasında müəyyən problemlər yaradır.
- *Mürəkkəbliik.* Verilənlər müxtəlif mənbələrdən daxil olurlar. Viki-səhifələr arasında mürəkkəb semantika mövcuddur. Səhifələr arasında linklər müxtəlif tipdə olurlar: layihədaxili linklər, dil faktoruna görə

(intervikilər) linklər, kateqoriyalara görə linklər, digər viki-layihələrə linklər və nəhayət, xarici layihələrə (İnternetin müxtəlif veb-saytlarına) linklər. Bu linklər hər an dəyişdirilə bilər. Belə bir mürəkkəblik vikimetrik tədqiqatları da mürəkkəbləşdirir.

Verilənlərin ənənəvi emalı zamanı onlar yoxlanılır, müəyyən formata gətirilir və yalnız ondan sonra sistemə yüklənir. Bu cür ardıcılıq viki-mühitdəki BD məsələlərində özünü doğrultmur. Böyük verilənlərin emalı prosesi verilənlər axınının vizuallaşdırılması, verilənlərin intellektual analizi, BD əsasında situasiyanın təsviri, analiz üçün müasir aparat və proqram təminatı və s. məsələlərin həllini nəzərdə tutur.

### III. PROBLEMİN HƏLLİ ÜÇÜN BƏZİ YANAŞMALAR

IBM şirkəti Vikipediya istifadəçilərin fəaliyyətini və hadisələr arasında qanunauyğunluqları müəyyən etmək üçün xüsusi alqoritm işləmişdir. 2007-ci ildən başlayaraq tətbiq edilən alqoritm verilənlərin vizuallaşdırılmasına əsaslanır. Xromoqram şəklində təsvir və xromoqramdakı rənglər mətndəki sözlərə verilən rənglərdən asılıdır. Xromoqramdan istifadə etməklə istifadəçinin hansı mövzuya maraq göstərməsini, vandalizm hallarını, konfliktli situasiyaları və digər maraqlı məsələləri analiz etmək mümkündür [7]. Məsələn, aşağıda göstərilən xromoqram hərbi-dəniz qüvvələrinin tarixi ilə bağlı məqalələrdə edilən düzəlişlər əsasında təşkil olunmuşdur. Xromoqramda üstünlük təşkil edən bənövşəyi rəng hərbi-dəniz gəmilərinin adlarını göstərir. Yəni, xromoqram göstərir ki, istifadəçi müxtəlif mövzularda müraciət etsə də, hərbi gəmilərə olan maraq üstünlük təşkil edir.



Şəkil 1. Viki-mühitdə istifadəçinin maraq dairəsini göstərən xromoqram

Vikipediya böyük verilənlərin analizini nəzərdə tutan digər sistem SGI UV 2 sistemidir [8]. SGI şirkəti və İllinoys Universitetinin əməkdaşı K. Liptau (Kalev H. Leetaru) tərəfindən yaradılan bu sistem Vikipediya tammətini kontentlərin zaman və məkana görə xronoloji kartoqrafiyasını yaradır və axtarışını həyata keçirir. Sistem ingilis dilindəki Vikipediya toplanmış verilənləri analiz etməklə dünya tarixinin inkişaf mərhələlərini müəyyən etmiş və müasir tarixin vizual təsvirini yaratmışdır. Sistem həmçinin, Vikipediya verilənlərdən istifadə etməklə son 2 ərddə dünyanın necə vizuallaşmasını göstərmişdir. Sistem viki-səhifələrin məzmununu müəyyən etmiş, səhifələr arasındakı semantik əlaqələri dəqiqləşdirmiş və bu əlaqələr əsasında böyük şəbəkə yaratmışdır (şəkil 2).



Şəkil 2. Dünyanın SGI UV 2 sistemi vasitəsi ilə qurulmuş 1900-cü ildə vizual tarixi xəritəsi

SGI şirkətinin marketing üzrə direktoru Frans Aman (Franz Aman) SGI UV 2 sistemini Google Earth ilə müqayisə edərək bildirmişdir: “Google Earth-də xəritənin ölçülərini azaltmaqla ümumi vəziyyətə baxmaq bizim xoşumuza gəlir və bu konsepsiyayı Vikipediya böyük verilənlərə tətbiq etməklə, biz ümumi təsvir əldə etmişik” [8]. Sistemdə verilənlərin emalı əsas yaddaşda toplanan verilənlərin intellektual analizi (in-memory data-mining) vasitəsi ilə həyata keçirilir. İntellektual analiz alqoritmləri, ilk növbədə, mətnin xüsusiyyətini, məzmununu, səhifələr arasındakı əlaqələri üzə çıxarır. Nəticələr metaverilənlərdə saxlanılır və sonrakı analizlərdə istifadə olunur.

Bəzi yanaşmalarda BD problemləri ənənəvi 3V (volume, variety, velocity) [9] ilə deyil 3C parametrləri ilə göstərilir: cardinality (həddən artıq böyüklük), continuity (fasiləsizlik) və complexity (mürəkkəblik) [10]. Müəlliflərin fikrincə məsələnin həllində riyazi və statistik analiz üsullarını 3C parametrlərinə görə qurmaq daha sadədir. Nəzərə almaq lazımdır ki, 3V şərtindəki verilənlərin sürəti Vikipediya bir çox hallarda özünü doğrultmur. Belə ki, viki-mühitdə kontentin daxil edilməsində və dəyişdirilməsində fərqli fasilələrin yaranması ehtimalı vardır.

### IV. PROBLEMİN HƏLLİ YOLLARI

BD probleminin həlli ilə bağlı bəzi mövcud yanaşmaların analizi belə deməyə əsas verir ki, problemin həllində elə optimal həll yolunu seçmək lazımdır ki, alqoritm əvvəlki təcrübələrdən topladığı verilənlərdən istifadə edərək gələcəkdə viki-mühitdə hadisələrin inkişafını qabaqcadan avtomatik göstərsin. Belə ki, ənənəvi verilənlər xəzinəsi ilə işləyərək müəyyən məsələlərin həlli zamanı, məsələn, verilənlər arasında uyğunsuzluğun təyin edilməsi və s. məsələlərdə əvvəlcədən verilənlər xəzinəsində toplanmış və aqreqatlaşdırılmış verilənlərdən istifadə olunur. BD ilə işləyərək məsələnin bu cür həlli problemlər yaradır. İnformasiyanın emal prosesinə hazırlanması üçün verilənlər əvvəlcədən hazırlanmış serverlərə köçürülməlidirlər. Viki-mühitdəki BD bu yanaşmanı da mümkünsüz edir. Digər problem ondadır ki, viki-mühitdə verilənlər paylanmış şəkildədir və kontent yaratmaqla məşğul olan istifadəçilərin istəyinə uyğun şəkildə strukturlaşdırılır və formata gətirilir. Məsələn, verilənlər regional serverlərdə toplanır və hər bir serverdəki verilənlərin strukturlaşdırılması müxtəlif olur.

Verilənlərin analizi baxımından bu verilənləri əraziyə görə deyil, zamana görə serverlərə paylamaq daha düzgün olardı və hər bir server müəyyən zaman intervalına cavab verərdi. Lakin viki-serverlərin imkanları buna yol vermir. Bunları nəzərə alsaq viki-mühitdəki BD-nin analizi üçün verilənləri aşağıdakı növlərə bölmək olar:

- Yaranma zamanına görə (köhnə verilənlər/yeni dəyişdirilmiş verilənlər);
- Əldə olunma üsuluna görə (kompüterlərdən/mobil telefonlardan daxil edilən verilənlər);
- Tipinə görə (mətn/video/səs/şəkil tipli verilənlər).

BD ilə işləyərək verilənlər axını girişə verilir və bu zaman çox sürətli analiz tələb olunur. Məsələn, viki-mühitdə gizli sosial şəbəkələrin aşkarlanması, konfliktli situasiyaların qarşısının alınması, viki-səhifələrin və istifadəçilərin monitorinqinin aparılması, sürətlə yayılan informasiyanın və onun mənbəyinin müəyyən edilməsi və s. kimi məsələlər tez bir zamanda həll olunmalıdır. Lakin informasiya həddən artıq çox olduğundan biz burada sürətli emal imkanına malik intellektual analiz alqoritmlərinə ehtiyac görürük. Alqoritmlər həm də viki-səhifələrdəki mətnləri, istifadəçilərin müzakirələrini, daxil etdikləri şəkil, video və audio-faylları emal etməlidirlər. Odur ki, viki-mühitdə toplanmış BD ilə işləmək üçün aşağıdakı şərtlər ödənməlidir:

1. Viki-mühitdəki verilənlərin analizində istifadə olunacaq metod müəyyən edilməlidir;
2. Qısa zamanda böyük verilənlərin analizini həyata keçirə biləcək çox güclü intellektual sistem olmalıdır.

BD-dan əldə olunmuş massivlər heç də həmişə təsadüfi seçmədən əldə olunmur. Bəzən statistik verilənlər daha böyük əhəmiyyət kəsb edir. Burada əsas problem ondan ibarətdir ki, böyük, mürəkkəb, geterogen verilənlərin statistik analizi yalan və uyğun olmayan dəyişənlərin və alqoritmlərin seçilməsinə səbəb ola bilər. Odur ki, böyük verilənlərdən istifadə zamanı əhatəlilikdə, seçimdə, ölçmədə və əldə olunan cavablarda yanlışlıq ola bilməsi riskləri nəzərə alınmalıdır [11].

Böyük verilənlərin analizində yalnız proqram deyil, həm də xüsusi aparat təminatına tələbin olmasını nəzərə alaraq bu sahədə yanaşmaların xüsusi kompüter modelləşmələri ilə birgə tədqiqi məsələsi önə çəkməlidir. Məsələn, buna misal olaraq agent modelləşdirməsini göstərmək olar. Agent modelləşdirmə desentralizə olunmuş verilənlərin davranışını tədqiq edir və bu verilənlərin davranışı bütün sistemin davranışını təyin edir [12]. Eyni zamanda faktor və klaster analiz kimi çoxölçülü analiz metodlarından istifadə etməklə verilənlərin gizli stukturunu tədqiq etmək mümkündür [13]. Böyük verilənlərin klasterləşdirilməsində Hadoop Distributed File Systems (HDFS) [14] modelindən, Hive database və ZooKeeper metodlarından [15], Cloud texnologiyalarından [16] birgə istifadə daha səmərəli nəticə əldə etməyə imkan verə bilər.

HDFS verilənlərin xüsusi klasterlərdə saxlanması və təşkilini təmin edir. ZooKeeper – yüksək əlyətərliyə malik əlaqələndirmə servisidir. Paylanmış əlavələrin qurulmasında istifadə olunur. Hive – paylanmış verilənlər xəzinəsidir. Hive

xüsusi sorgu dili olan HiveQL dilində işləyir və HDFS-də toplanmış verilənləri idarə edir. HiveQL sorguları avtomatik olaraq Map-Reduce məsələlərinə tətbiq edirlər.

BD problemlərinin həlli ilə bağlı yanaşmaların qısa təhlili göstərir ki, verilənlərin paylanmış şəkildə emalı böyük verilənlərin tez bir zamanda analizini aparmağa kömək edə bilər. Böyük verilənlərin səmərəli emalı paylanmış fayl sistemləri texnologiyalarına əsaslanır. Paylanmış sistemlərdə verilənlər bir fayl sistemində deyil, bir neçə serverlərdə yerləşən verilənlər xəzinəsində saxlanılır və indeksasiya edilir. Bu baxımdan, böyük verilənlərin emalında paralel və paylanmış verilənlər bazası idarəetmə sistemlərindən (VBİS) istifadə və verilənləri yerləşdikləri qovşaqlarda emal etmək daha düzgündür. Bu halda paralel VBİS-lərin iş keyfiyyəti verilənlərin düzgün paylanmasından asılıdır. Analitik VB üçün verilənlərin paylanması sorguların yerinə yetirilməsi zamanı qovşaqlar arasında informasiya mübadilələrinin sayının minimum olmasına gətirib çıxaracaqdır.

Nəzərə almaq lazımdır ki, ənənəvi verilənlər xəzinəsində bütün verilənlər onların konvertasiyası, təmizlənməsi, yoxlanması və yüklənməsi işlərinə cavabdeh olan bir məntiqi blokdan verilsə, BD-nin emalı zamanı bir məntiqi blokdan istifadə səmərəli deyildir. Ənənəvi verilənlər xəzinəsində istifadə olunan alətlər: çoxölçülü analiz (OLAP), regressiya, təsnifatlandırma, klasterləşmə və qanunauyğunluqların axtarışıdır. Bu gün SAP HANA, Greenplum Chorus, Aster Data Cluster kimi sistemlər bu metodları BD-nin analizi üçün tətbiq etməyə imkan yaradır. BD ilə işləmək üçün ən populyar texnologiyalara misal olaraq 2004-cü ildə Google şirkəti tərəfindən təklif olunan Map-Reduce modelini və Yahoo şirkəti tərəfindən təklif olunan açıq kodlu Hadoop proqramını göstərmək olar. Bu sistemlərdə verilənləri cədvəllərə yığmaq və ciddi iyerarxiyaya tabe olmaq lazım gəlmir. Analizdən öncə verilənləri növlərə bölmək kifayətdir. Map-Reduce modeli verilənlərin düzgün paylanması işində səmərəliliyi təmin edir. Map-Reduce – paylanmış verilənlərin hesablanması modelidir. Böyük verilənlər üzərində paralel hesablamalar üçün istifadə olunur. Giriş verilənləri ilkin emal üçün nəzərdə tutulmuş xəritə (Map) işçi qovşaqlara (individual nodes) paylanmış verilənlərin harada yerləşməsinə göstərir, emalda istifadə olunan alqoritmlər və nəticə haqqında informasiya saxlayır. Əməliyyatların sonunda nəticələr toplanır (Reduced). Məsələn, alqoritm yekun cəmi tapmaq üçün paralel olaraq paylanmış fayl sisteminin hər bir qovşaqlarda aralıq cəmləri hesablayacaq və sonrakı mərhələdə bu aralıq qiymətləri toplayacaqdır.

Verilənlərin effektiv paylanması üçün verilmiş qrafın müəyyən xüsusiyyətli altqraflara ayrılması məsələlərinə də baxmaq lazımdır. Burada əsas şərt altqrafları birləşdirən qovşaqların sayının minimum olmasıdır. Bu, “NP-completeness” məsələsidir. Yəni, məsələni həll etmək üçün onu NP (non-deterministic polynomial) sinifindəki məsələlərə bölmək daha düzgündür. Əgər bölünmüş məsələlərdən hər hansı biri üçün həll alqoritm tapılırsa, o zaman NP sinifindən olan digər məsələlərin həllində də bu alqoritmədən istifadə

etmək mümkündür. Məsələnin həllində intellektual analiz metodlarından və hibrid alqoritmlərdən istifadə daha səmərəli nəticələr əldə etməyə imkan yaradır. BD-nin emalı üçün yalnız verilənlərin paylanmış emalını həyata keçirən konkret texnologiyalardan istifadə ilə yekunlaşmaq olmaz. Emal zamanı BD üçün xarakterik olan vacib parametrlər də nəzərə alınmalıdır. Məsələn, şəbəkədə qarşılıqlı əlaqənin intensivliyi, verilənlərin həcmi və s. Viki-mühitdə də, digər sahələrdə olduğu kimi BD problemləri yeni suallar yaradır ki, onlara cavab vermək lazım gəlir. Məsələn, BD ilə bağlı yanaşmaların və modellərin verifikasiyası və validasiyası vacib məsələlərdəndir. Bu ondan irəli gəlir ki, verilənlər həddən artıq böyük olduqda xətlərin ölçülməsi və ölçmə zamanı yaranan xətlərin sayı qarşılıqlı əlaqəlidir.

### NƏTİCƏ

Tədqiqat nəticəsində məlum olmuşdur ki, viki-mühitdə böyük verilənlərin toplanması bu verilənlərin emalında və saxlanmasında problemlər yaradır və məsələlərin həllində yeni yanaşmaların tətbiqini tələb edir. Viki-mühitdəki böyük verilənlərin vizuallaşdırılmasına aid mövcud sistemlər göstərdi ki, Vikipediya toplanmış verilənlər özlərində BD-nin xüsusiyyətlərini saxlayırlar: ölçüləri böyükdür – onların səmərəli strukturlaşdırılması və saxlanması üçün xüsusi şərtlər tələb olunur, müxtəlifdir – onları ənənəvi üsullarla bir VB-yə toplayıb emal etmək mümkün deyil. Həmçinin viki-səhifələr arasında mürəkkəb semantika mövcuddur ki, bu da onların emalını çətinləşdirir. Viki-mühitdəki verilənlərin müxtəlifliyi və dəyişkənliyi onların analizi və təsvirində kompüter qrafikasının yeni texnologiyalarından istifadəni qaçılmaz edir.

BD ilə bağlı məsələlərin həllində araşdırılan yanaşmalar belə deməyə əsas verir ki, verilənlərin paralel və paylanmış şəkildə emalı daha səmərəli nəticə əldə etməyə imkan yaradır.

### ƏDƏBİYYAT

- [1]. R.M. Əliquliyev, Hacırahimova M.Ş. “Big Data” fenomeni: problemlər və imkanlar // İnformasiya texnologiyaları problemləri. 2014, №2, s. 3–16.
- [2]. I.Y. Alakbarova. Some Approaches to the Development of Information Influence and Hidden Communications Detection Systems in Wiki-Environment / Proceedings of the 4th International Conference “Problems of Cybernetics and Informatics” (PCİ), vol. I, Sept. 12-14, 2012, pp. 119-120

- [3]. P.M. Алгулиев, P.M. Алыгулиев, И.Я. Алекперова. Викиметрические исследования: современное состояние и перспективы // “Телекоммуникации”, М. 2014, №5, стр. 15-31
- [4]. [https://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias](https://meta.wikimedia.org/wiki/List_of_Wikipedias)
- [5]. [https://commons.wikimedia.org/wiki/Main\\_Page](https://commons.wikimedia.org/wiki/Main_Page)
- [6]. <https://www.openhub.net/p/mediawiki>
- [7]. M. Wattenberg, F. B. Viégas, Hollenbach K. Visualizing Activity on Wikipedia with Chromograms // Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction, 10 Sept. 2007, Berlin, vol. 4663, pp. 272–287
- [8]. <http://www.sgi.com/go/wikipedia/>
- [9]. R. Alguliyev, Y. Imamverdiyev. Big Data: Big Promises for Information Security / Proceedings of the 8th International Conference on Application of Information and Communication Technologies (AICT), 15-17 Oct. 2014, IEEE, Astana, pp. 1–4
- [10]. M. Hilbert. Big Data for Development: From Information- to Knowledge Societies, 2013, Available at SSRN: <http://ssrn.com/abstract=2205145>
- [11]. M. Couper. Is the sky falling? New technology, changing media, and the future of surveys // Survey Research Methods, 2013, vol. 7, no. 3, pp. 145–156.
- [12]. S. Suthaharan. Big Data Classification: Problems and Challenges in Network Intrusion Prediction with Machine Learning // ACM Sigmetrics Performance Evaluation Review archive, 2014, vol. 41, no 4, pp. 70–73
- [13]. R.M. Alguliyev, R.M. Aliguliyev and I.Ya. Alekperova. Cluster approach to the efficient use of multimedia resources in information warfare in Wikimedia // Automatic Control and Computer Sciences, 2014, vol.48, no 2, pp. 97–108
- [14]. A. Thusoo, J.S. Sarma, N. Jain, Z. Shao, P. Chakka, N Zhang, A. Antony, H. Liu and R. Murthy. Hive – A Petabyte Scale Data Warehouse Using Hadoop // Proceedings of the 26th International Conference on Data Engineering (ICDE), 2010 IEEE, 1-6 March, 2010, Long Beach, pp. 996 – 1005
- [15]. P. Hunt, M. Konar, F.P. Junqueira and B. Reed. ZooKeeper: wait-free coordination for internet-scale systems // Proceedings of the 2010 USENIX conference on USENIX annual technical conference. Berkeley, CA, USA: USENIX Association, 2010, pp. 11–12
- [16]. S. Carlin and K. Curran. Cloud Computing Technologies // International Journal of Cloud Computing and Services Science (IJ-CLOSER), 2012, vol. 1, no. 2, pp. 59-65