

Big Data Konsepsiyası və Aktual Elmi-Nəzəri Problemlər

Rasim Əliquliyev¹, Məkrufə Hacırahimova², Aybəniz Əliyeva³

^{1,2,3}AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹rasim@science.az, ²makrufa@science.az, ³aybeniz63z@rambler.ru

Xülasə — Məqalədə “Big Data” konsepsiyası nəzərdən keçirilmişdir. Böyük verilənlərin tətbiqi sahəsində meydana çıxan aktual elmi-nəzəri problemlər təhlil olunmuşdur.

Açar sözlər— *big data, böyük verilənlərin analitikası, vizuallaşdırma, audio analiz, video analiz*

I. GİRİŞ

XXI əsrin əvvəllərindən başlayaraq texnika və texnologiyaların köməyi ilə yaradılan rəqəmsal verilənlər hər il həndəsi silsilə ilə artmaqdadır [1]. Belə ki, bəşəriyyətin mövcudluğundan 2003-cü ilə qədərki dövrdə dünyada cəmi 5 ekzabayt məlumat generasiya olunmuşdursa, 2015-ci ildə 8,1 zetabayt, növbəti hər il dünyada informasiyanın həcminin 40% artaraq 2020-ci ildə 44 zetabayta çatacağı proqnozlaşdırılır [2]. Nəticə etibarlı ilə verilənlərin emalı, saxlanması və istifadəsində yeni eranı əks etdirən “böyük verilənlər” (ing. big data) fenomeni meydana çıxmışdır [3]. Bu yeni termin həcm və mürəkkəblik baxımından mövcud idarəetmə metodları və intellektual analiz vasitələri ilə emal oluna bilməyən verilənləri təyin etmək üçündür.

Böyük verilənlər (BV) cəmiyyətin iqtisadi inkişafına təkan verməklə yanaşı elmi ictimaiyyəti bir çox problemlərlə üz-üzə qoymuş, yeni tədqiqat paradigması yaranmışdır [3]. BV-də bir çox elmi sahələrdə daha böyük nailiyyətlər və yeni elmi kəşflər, müəssisələrdə yüksək mənfəət əldə etmək üçün böyük potensial vardır. Bu potensialı reallaşdırmaq üçün əvvəlcə bir çox texniki, texnoloji problemlər həll olunmalıdır. Bir sıra problemlər də mövcuddur ki, onların həllində yeni elmi baxışlar, yanaşmalar, modelləşmə, riyazi metodlar, optimallaşdırma üsulları və s. tələb olunur. Böyük verilənlərin xarakterik xüsusiyyətlərindən (*böyük həcm, yüksək sürət, müxtəliflik*) irəli gələn problemlərin əsas istiqamətlərini isə miqyaslanma, saxlanma, ilkin emal, idarə etmə, analitika, analizin nəticələrinin interpretasiyası – vizuallaşdırma, təhlükəsizlik və s. təşkil edir. Bu problemlər tədqiqatçılardan real vaxtda böyük verilənlərdən lazım olan informasiyanı, biliyi, məlumatı əldə etmək üçün daha effektiv statistik və analiz üsulları, miqyaslılığı təmin edən arxitektura yanaşmalar, vahid infrastruktur və alətlər tələb edir. Bu mənada böyük verilənlərin elmi-nəzəri problemlərinin tədqiqi olduqca vacibdir və aktualdır.

II. BİG DATA KONSEPSİYASI

“Big data” termini ilk dəfə 1998-ci ildə Silicon Graphics şirkətinin kompüter elmləri üzrə mütəxəssisi Con Meşi tərəfindən istifadə edilmişdir. Bu terminə bir qədər sonra, 2000-ci ildə, Pennsylvania universitetinin professoru, “big data” terminin əsas araşdırıcılarından olan Fransis Dieboldun akademik mühitdə dərc olunan tədqiqatında rast gəlinir [4]. Onun sonrakı tədqiqatlarında bu terminin bir qədər də möhkəməndiyini, artıq onun bir hadisə, bir fenomen deyil, ciddi tədqiqat istiqamətinə çevrildiyindən bəhs olunur. Lakin, termin professor Klifford Linçin (*Clifford Lynch*) 2008-ci ildə “Nature” jurnalının xüsusi nömrəsində nəşr olunmuş “Big Data: How do your data grow?” adlı məqaləsindən sonra akademik mühitdə populyarlıq qazanmışdır [1]. Vahid tərifin olmaması “big data” fenomeninin anlanmasında qeyri-müəyyənlik yaradır. Elm, sənaye və biznes kimi maraqlı tərəflərin müxtəlif yanaşmaları mövcuddur [5,6.]. McKinsey Global İnstitutunun hesabatında bu termin informasiyanın toplanması, saxlanması, idarə edilməsi, axtarışı və analizinin tipik verilənlər bazasının imkanları xaricində olan verilənlər dəsti kimi xarakterizə olunur [1,7].

Tətbiq olunduğu sahələrdən asılı olmayaraq böyük verilənləri təsvir etmək üçün ümumi xarakteristikalar mövcuddur. Bu xarakteristikalar böyük verilənlərin əsas problemlərini özündə əks etdirməklə üç əsas qrupa bölünür: həcm (*volume*), sürət (*velocity*) və müxtəliflik (*variety*). İngilis dilli mənbələrdə bunu «3V» də adlandırırlar. Bu parametrlərin konvergensiyası böyük verilənləri təyin etməyə və digər verilənlərdən fərqləndirməyə kömək edir. Bu model ilk dəfə 2001-ci ildə D.Laney tərəfindən verilmişdir [1]. O, “big data” terminini istifadə etməsə də, elektron kommersiyada bir tendensiyanı: verilənlərin idarə edilməsinin daha vacib, daha çətin olacağını əvvəlcədən xəbər vermiş və daha sonra verilənlərin idarə edilməsində verilənlərin ölçüsünü, ötürülmə sürətini və müxtəlifliyini əsas problem kimi təyin etmişdir. Bu xarakteristikalar isə ümumilikdə “big data” texnologiyalarının əsas konsepsiyasını təşkil edir. Bu konsepsiya çox böyük sürətlə və müxtəlif mənbələrdən toplanan çox böyük həcmdə verilənləri daha səmərəli istifadə etmək, saxlamaq, analiz edərək ondan daha qiymətli informasiyanı əldə etmək ideyasını özündə əks etdirir. [1].

Həcm (volume). Həcm böyük verilənlərin ən əsas xarakteristikasıdır. Hazırda BV-nin miqyası terabaytlardan zetabaytlara qədər həcm ilə xarakterizə olunur [1]. Həcm problemi ilk növbədə saxlama problemi yaradır ki, bu da genişmiqyaslı saxlama və paylanmış emal tələb edir.

Sürət (velocity). Burada iki hal nəzərdə tutulur. Birinci, yeni verilənlər böyük sürətlə generasiya olunur, mövcudlar yenilənir və toplanır. İkincisi, həcm artdıqca, emal üçün də çox böyük sürət tələb olunur. Sürət zaman problemi kimi dəyərləndirilir və mövcud emal texnologiyalarının verilənləri real vaxtda analiz etmək imkanına malik olması ilə izah olunur [1].

Müxtəliflik (variety). Müxtəliflik BV-nin təbii özəlliklərindəndir. Məlumatların əksəriyyəti müxtəlif mənbələrdən (e-poçt, sosial şəbəkələr, veb-saytlar, sensorlar və s.) müxtəlif formatlarda daxil olur və müxtəlif indeksləşmə sxemi istifadə olunur. Bunları sadəcə olaraq, bir araya yığmaq və birgə emal etmək və analiz üçün uyğun şəkllə salmaq çox çətin olur [1].

IBM verilənlərin həqiqiliyini əsas götürərək 4-cü “V” (*veracity*), Oracle isə BV-nin dəyərini qeyd etməklə 5-ci “V” (*value*)-ni daxil etmişdir [1]. Son zamanlar mütəxəssislər tərəfindən “v”-lərin sayı artmaqdadır.

III. BIG DATANIN ELMİ-NƏZƏRİ PROBLEMLƏRİ

Bu gün “Big data” adlandırılan informasiya bolluğu həqiqətən də mövcuddur. Ancaq bu insanlarda məyusluq deyil, tam əksinə, ona təbii xammal, resurs kimi baxmaq lazımdır. Çünki bu xam verilənlərdə elmi kəşflərə səbəb ola biləcək dərin bilik toplanmışdır. Ancaq bu resursdan maksimum istifadə etməklə cəmiyyət və biznes sahəsində dəyər yaratmaq üçün yeni nəsill analitik texnologiyalara ehtiyac vardır. Bu baxımdan, böyük verilənlər mövzusu istər elmi tədqiqatçılar, istər dövlət qurumlarında qərar qəbul edən şəxslər və siyasətçilər, istərsə də biznes cəmiyyətləri və nümayəndələri tərəfindən böyük diqqət çəkmişdir. Qeyd edildiyi kimi böyük verilənlər mövzusu yeni tədqiqat istiqamətinə çevrilmişdir. Nüfuzlu beynəlxalq təşkilatlar, elmi qurumlar tərəfindən çox böyük həcmdə informasiyanın emalının müxtəlif aspektlərinə həsr olunmuş çoxsaylı konfranslar (IEEE International Congress on Big Data, International Conference on Data Science and Advanced Analytics, IEEE International Conference on Big Data and Cloud Computing, IEEE International Conference on Cloud and Big Data Computing, IEEE International Conference on Multimedia Big Data və s.), simpoziumlar (IEEE International Symposium on Big Data & Visual Analytics, IEEE Symposium on Large Data Analysis and Visualization, IEEE/ACM International Symposium on Big Data Computing və s.), seminarlar (Workshop on Big Data Analytics in CPS: Enabling the Move From IoT to Real-time Control), forumlar keçirilməkdədir. Bu tədbirlərin əsas müzakirə mövzularını: böyük verilənlərin arxitekturu (*Big Data Architecture*), idarə edilməsi (*big data management*), modelləşdirilməsi (*Big Data Modeling*), analitikası (*Big Data Analytics*), alətləri (*Big Data*

Toolkits), açıq platformalar (*Big Data Open Platforms*), “Big Data” xidmət kimi (*Big Data As a Service*), biznesin səmərəli idarə edilməsi (*Big Data in Business Performance Management*), e-dövlətdə və cəmiyyətdə böyük verilənlərin analitikası (*Big Data Analytics in e-Government and Society*), vizuallaşdırma (*Visualization*), təhlükəsizlik (*security*), böyük verilənlər üçün alqoritmlər və s. təşkil edir [1].

BV-nin problemlərinin həlli sahəsində dünyanın tanınmış elmi mərkəzləri (The Big Data Institute - University of Oxford, The Big Data Institute - University College London, Data Science Institute – Imperial College London, Cambridge Big Data University of Washington eScience Institute, Institute for Big Data Analytics - Dalhousie University, Institute for Data Science - University of Rochester, Big Data University – IBM) böyük verilənlərin arxitekturası, toplanması və emalı, saxlanması, analitikası, təhlükəsizliyi, vizuallaşdırılması və s. istiqamətlərdə elmi-tədqiqat işləri aparırlar [1].

Elmi və populyar jurnalların: Nature (2008), Science (2011), Computer (2013) və s. xüsusi nömrələri bu mövzuya həsr olunmuşdur. Son illərdə işiq üzü görmüş böyük verilənlərin elmi-nəzəri problemlərini işıqlandıran “Journal of Big Data”, “International Journal of Big Data” (IJBD), International Journal of Big Data Intelligence (IJBDI), Big Data Research, Big Data & Society, IEEE Transactions on Big Data və s. kimi akademik jurnallar nəşr olunmaqdadır.

Böyük verilənlərin analizi problemləri

Özündə son dərəcə faydalı informasiyanı daşıyan, adi relyasiya bazalarının emal edə bilmədiyi yüzlərlə terabayt və ekzabayt həcmində mətn, təsvirlər, audio-video və s. tip strukturlaşdırılmamış informasiyanın toplanması və idarə edilməsi, saxlanması, təhlükəsizliyi, axtarışı, analizi (*analitik hesabatların generasiyası və vizuallaşdırılması, proqnozlaşdırma*) və s. kimi məsələlərin həllində yeni elmi yanaşmalar, daha mükəmməl analiz üsulları tələb olunur [8].

BV-nin analizi onun əsas problemlərindəndir. Bu problem BV-nin xarakteristikaları, mövcud analiz model və metodları və verilənlərin emal sistemlərinin məhdudiyətləri ilə bağlıdır [8]. Verilənlərin həcmnin kəskin sürətdə artması və real zamanda onların analizinə olan tələbat BV-nin ən əsas problemlərindən sayılan böyük verilənlərin analitikasının (*Big Data Analytics*) yaranmasına gətirib çıxardı. “Big Data Analytics” daha böyük və mürəkkəb massivlərə tətbiq edildiyindən kəşf edən analitika (*Discovery Analytics*) və izah edən analitika (*Exploratory Analytics*) terminlərindən də istifadə edilir. Necə adlandırılmasından asılı olmayaraq, mahiyyət eynidir – qərar qəbul edən şəxsləri müxtəlif proseslər haqqında məlumatlarla təmin edən əks əlaqəni yaratmaq [1].

Ümumiyyətlə, BV-dən ideyaların çıxarılması prosesi verilənlərin idarə edilməsi və analizi kimi iki alt prosesdən ibarətdir. Məlumdur ki, analiz müxtəlif parametrlər, xarakteristikalar, hadisələr və s. arasındakı korrelyasiyanı tapmağa, təsnifatlandırma və analitik hesabatlar və bunun

əsasında proqnozların verilməsinə imkan verir. Bu aspektdən müasir texnologiyalar verilənlərdəki informasiyanın yeni biliklərə çevrilməsinə və ya biliklərin əldə edilməsinə imkan verməlidirlər. BV-in saxlanması, emalı və analizi üçün böyük hesablama gücü, miqyaslılığı təmin edən arxitektura yanaşmalar, vahid infrastruktur tələb olunur [1].

Əlbəttə verilənlərin çox da sadə olmayan analiz üsulları mövcuddur, lakin mətn, video və audio kimi strukturlaşdırılmamış informasiyaların analizi ciddi problemdir.

Mətn analizi (text mining) – “data mining” sinfindən olan üsulların köməyi ilə təbii dildə mənlərdən əvvəlcədən məlum olmayan əlaqələri və korrelyasiyaları aşkarlamaqla bilikləri çıxarır [9]. Təsnifatlandırma (*classification*) və klasterləşdirmə (*clustering*), informasiyanın çıxarılması (*information extraction*), avtomatik referatlaşdırma (summarization) isə text mining-in həll etdiyi əsas problemlərdir. Mahiyyət etibarilə *text mining* mənləri daha dərin analiz etmək üçün linqivistik, statistik üsulların, informasiya axtarışı, həmçinin məşin təlimi alqoritmlərini istifadə edir [10,11].

Video analitika – video axınının monitorinqi, analizi və onlardan faydalı informasiyanın çıxarılmasını özündə birləşdirir. Video analitika verilənlərin digər analiz növləri ilə müqayisədə başlanğıc mərhələsində olmasına baxmayaraq, onların real-vaxt rejimində emalı üçün iki yanaşma mövcuddur: “*server-based*” və “*edge-based*” [10].

Video məlumatların analizi hazırda tədqiqatçılar qarşısında duran çox çətin məsələ olaraq mövcuddur. Burada əsas problem kadrların tezliyi və təsvirlərin dəqiqliyinin azalması nəticəsində informasiya itkisinin yaranmasıdır [10].

Audio analitika – strukturlaşmamış audio verilənlərin analizi və onlardan faydalı məlumatın çıxarılması üsuludur. Bu üsullar müştərilərə göstərilən xidmətlərin keyfiyyətini yüksəltməyə, məxfilik və təhlükəsizlik kimi müxtəlif məsələlərin yerinə yetirilməsinə nəzarət etməyə və s. kömək edir. Onun əsas problemləri isə nitqin tanınması, küy və s. ilə bağlıdır [10].

Sosial media analitika – sosial media kanallarının strukturlaşdırılmış və strukturlaşdırılmamış verilənlərinin analizini həyata keçirir. Sosial medialarda məzmunlar çox vaxt böyük həcmli, küylü və dinamik olur. Ona görə də mətn, audio və video analitikada qeyd olunan problemlər, həmçinin, BV-nin ötürülməsi onun əsas problemlərindəndir [10,12].

Proqnostik analitika – tarixi və cari məlumatlar əsasında gələcək nəticələri proqnozlaşdırmağa imkan verir və müxtəlif statistik metodlara əsaslanır. Lakin böyük verilənlərin xarakterik xüsusiyyətlərindən irəli gələn verilənlərin müxtəlifliyi, küyün yaranması, yanlış korrelyasiya və təsadüfi endogenlik kimi problemlər ənənəvi statistik metodların tətbiqinə əngəl törədir və BV üçün yeni statistik metodların işlənilməsinə tələb edir [3,10,13,14].

Müxtəlif tip verilənlərin inteqrasiyası və həcmi, miqyaslanması, təhlükəsizliyi, uyğunsuzluğu və s. analitikanın əsas problemləridir [15].

A. Böyük Verilənlərin Arxitekturası

Hal-hazırda BV-nin analitikası üçün geniş və hamı tərəfindən qəbul edilmiş arxitektura yoxdur. Böyük verilənlərin arxitekturasının əsas funksional komponentlərinə faydalı verilənlərin çıxarılması (Data Extraction), verilənlər axınının emalı (Stream Processing), informasiyanın çıxarılması (Information Extraction), verilənlərin keyfiyyətinin/qeyri-müəyyənliyin idarə olunması (Manage Data Quality /Uncertainty), verilənlərin inteqrasiyası (Data Integration), verilənlərin analizi (Data Analysis), verilənlərin paylanması (Data Distribution), verilənlərin saxlanması (Data Storage), meta verilənlərin idarə olunması (Metadata Management), verilənlərin həyat dövrünün idarə olunması (Data Lifecycle Management), məxfilik (Privacy) daxildir [16].

BV-nin emal sisteminin enerji səmərəliliyinin qiymətləndirilməsi və optimallaşdırılması böyük elmi-tədqiqat problemlərindəndir [8].

B. Böyük Verilənlərin emalı və saxlanması problemləri

Verilənlərin idarə edilməsinə informasiyanın əldə edilməsi, saxlanması, analiz üçün hazırlanması və emal prosesləri daxildir.

BV-nin emalında aşağıdakı yanaşmalar mövcuddur:

Package: paket emalı; **Stream Computing:** Real-vaxt rejimində daim yenilənən verilənlərin analitik emalını təmin edir və proqnozlar tərtib etmək, daha sürətli analiz və qərar qəbul etmək imkanı verir [15]; **Data-intensive computing:** terabayt və petabaytlarla verilənlərin emalında paralel hesablamaların aparılmasına yönəlmişdir. Petabaytlarla verilənlər massivlərinin emalının zəruriliyi “data-intensive computing” yanaşmasını ortaya çıxarmışdır ki, bu “hesablama deyil, verilənlər daha əsas sərvətdir” anlamına gəlir [15,17,18].

BV emalı və saxlanması üsullarını, saxlanma qurğularını, verilənlərin saxlanması arxitekturasını, verilənlərə giriş mexanizmlərini köklü surətdə dəyişmişdir. Saxlanma qurğuları böyük həcmli verilənlərin əlyətərliyini və operativ analizini təmin etmək imkanına malik olmalıdır [15]. Hazırda saxlama məsələsinin həllində informasiyanın qurğular arasında miqyasını həyata keçirmək imkanına malik olan DAS (Direct-Attach-Storage), NAS (Network Attached Storage), SAN (Storage Area Networks), HSM (Hierarchical Storage Management), ILM (Information Life-cycle Management) kimi texnologiyalar geniş istifadə olunur [1,8]. Lakin irimiqyaslı paylanmış sistemlər səviyyəsinə gəldikdə isə verilənlərin bu saxlanma arxitekturalarının çatışmazlıqları və məhdudiyətləri ortaya çıxır. Son zamanlar isə saxlama qurğularının yaddaş tutumunun artması, çoxsaylı kompüterlərin (server, kompüter və s.) hesablama və yaddaş resurslarının klasterləşməsi və virtuallaşdırılmasını həyata

keçirməklə, verilənlərin emalı və yadda saxlanılmasına xidmət edən “grid” və “cloud computing” texnologiyalarının tətbiqi saxlama sahəsindəki problemləri demək olar ki, aradan qaldıra bilmişdir [1, 15]. Bulud hesablamaları (**cloud computing**) saxlanma, böyük hesablamaların aparılmasında son dərəcə müvəffəqiyyətli yanaşmalardandır. Digər tərəfdən bulud saxlaması verilənlərin təhlükəsizliyinə nəzarət baxımından verilənlərin təhlükəsizliyi problemi yaradır [15].

C. Böyük Verilənlərin Vizuallaşması Problemləri

BV-nin analizində əsas məsələlərdən biri də nəticələrin təqdim olunması – vizuallaşdırılmasıdır. Böyük ölçüyə və həcmə malik böyük verilənlərin vizuallaşdırılması xüsusilə çətinlikdir. BV-nin xarakteristikaları nəzərə alınmaqla onların vizuallaşması zamanı yaranan əsas problemlər aşağıdakılardır [19]:

- vizual küy (*visual noise*);
- böyük təsvirin anlaşılması (*large image reception*);
- informasiya itkisi (*information loss*);
- yüksək məhsuldarlıq tələbləri (*high performance requirements*);
- təsvirlərin dəyişməsinin yüksək sürəti (*high rate of image change*) və s.

Bu problemlərin aradan qaldırılması üçün yeni üsulların və texnologiyaların işlənməsi, eləcə də ixtisaslı kadrların olması əsas məsələlərdəndir.

D. Böyük verilənlərin təhlükəsizlik problemləri

BV meydana gəlməsiylə informasiya təhlükəsizliyi baxımından yeni problemlər yarandı. “Big data” texnologiyalarının tətbiqi mövcud təhlükəsizlik modellərinin köhnəliyini üzə çıxarmışdır. 15 il əvvəl istifadə olunan təhlükəsizlik yanaşması bu gün üçün adekvat deyildir. Eyni zamanda böyük verilənlərin həcm, müxtəliflik və sürət kimi xarakteristikaları isə təhlükəsizlik və gizlilik problemini daha da kəskinləşdirir [1]. Geniş miqyaslı “bulud” infrastrukturunu, verilənlərin mənbələrinin müxtəlifliyi, axın şəklində informasiyanın toplanması və böyük həcmdə informasiyanın “buludlarda” miqrasiyası təhlükəsizlik sistemlərinin “zəif” cəhətlərini üzə çıxarmışdır. Belə ki, BV-nin genişlənməsi ilə ənənəvi təhlükəsizlik mexanizmləri kifayət deyildir. Eyni zamanda verilənlərin axını çox çevik və sürətli təhlükəsizlik həlləri tələb edir [15,17]. Böyük verilənlərin tətbiqiylə bağlı verilənlərin təhlükəsizlik problemləri bir neçə səbəblərə görə daha da mürəkkəbləşmişdir. Birincisi, böyük verilənlərin ölçüsü mövcud mühafizə yanaşmalarına münasibətdə çox böyükdür. Bu da öz növbəsində təhlükəsizlik sahəsində işləri bir qədər də çətinləşdirir. Digər tərəfdən böyük verilənlər paylanmış şəkildə saxlanılır və şəbəkələrdəki təhdidlər bu təhlükələri artırır bilirlər [15].

NƏTİCƏ

Biz böyük verilənlər erasına daxil olmuşuq. BV artıq elmi tədqiqat istiqamətinə çevrilmişdir. Elmi tədqiqatların əsas istiqamətlərini isə böyük verilənlərin konsepsiyasının əsasını təşkil edən, BV-nin həcm, sürət və mürəkkəblilik kimi təbii

xüsusiyyətlərindən doğan elmi tutumlu məsələlər təşkil edir. Məsələlərin həllində data mining sinfinin üsulları (assosiativ qaydalar, rəqressiya, klassifikasiya, klasterizasiya və s.), süni neyron şəbəkələr, maşın təlimi, optimallaşma, o cümlədən genetik alqoritmlər, obrazların tanınması, prediktiv analitika, imitasiya modelləşməsi, statistik analiz, analitik verilənlərin vizuallaşdırılması kimi metodların tətbiqi vacibdir. Bunun üçün yüzlərlə terabayt və ekzabayt həcmində mətn, təsvirlər, audio-video və s. tip strukturlaşdırılmamış informasiyanın emalı və idarə edilməsi, saxlanması, təhlükəsizliyi, axtarışı, analizi və s. bağlı problemlər öz elmi həllini tapmalıdır.

ƏDƏBİYYAT

- [1] Əliquliyev R.M., Hacırahimova M. Ş. "Big Data" fenomen: problemlər və imkanlar, İnformasiya texnologiyaları problemləri, 2014, №2, səh. 3-16.
- [2] The digital universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Study report, IDC, December 2012. www.emc.com/leadership/digital-universe/
- [3] Fan J., Han F. & Liu H. Challenges of Big Data analysis, National Science Review, 2014, vol. 1, no 2, pp. 293–314.
- [4] Diebold, F.X. "Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting", Discussion Read to the Eighth World Congress of the Econometric Society, 2000.
- [5] NIST Big Data Working Group (NBD-WG). <http://bigdatawg.nist.gov/home.php>.
- [6] Madden S. From Databases to Big Data, IEEE Internet Computing, 2012, vol.16, no 3, pp.4-6.
- [7] Big data: The next frontier for innovation, competition, and productivity. Analyst report, McKinsey Global Institute, May 2011. <http://www.mckinsey.com/>
- [8] Jina X., Benjamin W. Waha, Chenga X., Wanga Y. Significance and Challenges of Big Data Research, Big Data Research, 2015, Vol. 2, no 2, pp. 59–64.
- [9] Alguliyev R., Aliguliyev R., Hajirahimova M. Multi-document summarization model based on integer linear programming, Intelligent Control and Automation, 2010, vol.1, no.1, pp.105-111.
- [10] Gandomi A., Haider M. Beyond the hype: Big data concepts, methods, and analytics TedInternational Journal of Information Management, 2015, vol. 35, pp. 137–144.
- [11] Jiang, J. Information extraction from text. In C. C. Aggarwal, & C. Zhai (Eds.), Mining text data (pp. 11–41). United States: Springer, 2012.
- [12] Barbier, G., & Liu, H. Data mining in social media. In C. C. Aggarwal (Ed.), Social network data analytics (pp. 327–352). United States: Springer, 2011.
- [13] Kambatla K., Kollias G., Kumar V., Grama A. Trends big data analytics, Parallel and Distributed Computing, 2014, vol.74, no.7, pp.2561-2573.
- [14] Hajirahimova M. Sh., Aliyeva A.S., Review of statistical analysis methods of high-dimensional data, Eastern-European Journal of Enterprise Technologies, Kharkov, 2015, no 5, pp. 23-30. <http://journals.urau.ua/eejet>
- [15] Philip C.L., Zhang C-Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Information Sciences, 2014, vol. 275, pp.314–347.
- [16] Maier M, Towards a Big Data Reference Architecture, fevral, 2013 - www.win.tue.nl/~gfletche/Maier_MSc_thesis.pdf
- [17] Labrinidis, A., & Jagadish, H. V. Challenges and opportunities with big data. Proceedings of the VLDB Endowment, 2012, vol. 5, no12, pp. 2032–2033.
- [18] Jagadish H.V. Big Data and Science: Myths and Reality // Big Data Research, 2015, vol. 2, no 2, pp. 49–52.
- [19] Gorodov E., Gubarev V., Analytical Review of Data Visualization Methods in Application to Big Data, Journal of Electrical and Computer Engineering, 2013, pp.7 - <http://www.hindawi.com>