

UOT 004.9:351

Əliquliyev R.M.¹, Hacırəhimova M.Ş.², Əliyeva A.S.³

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

¹secretary@iit.ab.az, ²makrufa@science.az, ³aybeniz63z@rambler.ru

BIG DATA-NIN AKTUAL ELMİ-NƏZƏRİ PROBLEMLƏRİ

Son on ildə texniki və texnoloji inkişaf ilə əlaqədar olaraq dünyada çox böyük ölçüdə rəqəmsal verilənlər generasiya olunmuşdur. Nəticədə “big data” termini meydana gəlmiş, bu gün isə hər yerdə – qəzet, jurnal məqalələrində, bloqlarda və s. geniş müzakirə olunan əsas mövzuya çevrilmişdir. “Big data” müasir cəmiyyət üçün yeni imkanlar yaratmaqla yanaşı, elmi tədqiqatçılar üçün problemlər gətirmişdir. Məqalədə “big data” termininə kütləvi informasiya vasitələrindən, biznes sektorundan fərqli olaraq, elmi-tədqiqat obyektini kimi baxılır, “big data” konsepsiyasının qısa şərhi verilir. Onun tədqiqat istiqaməti kimi inkişafını şərtləndirən əsas amillər diqqətə çatdırılır. Həmçinin tədqiqatçıların diqqət mərkəzində olan aktual elmi-nəzəri problemlər təhlil olunur.

Açar sözlər: böyük verilənlər, böyük verilənlərin analitikası, audio analitika, video analitika, sosial media analitikası, vizuallaşdırma, təhlükəsizlik.

Giriş

XXI əsrin əvvəllərindən başlayaraq texnika və texnologiyalar – kompüterlər, mobil telefonlar, İnternet, sensor şəbəkələri, yerin süni peykləri, kosmik teleskoplar, bulud hesablamaları və s. vasitəsi ilə yaradılan rəqəmsal verilənlər hər il həndəsi silsilə ilə artmaqdadır. Bu vəziyyət cəmiyyətdə “informasiya partlayışı” kimi xarakterizə olunur. Belə ki, bəşəriyyətin mövcudluğundan 2003-cü ilə qədərki dövrdə dünyada cəmi 5 ekzabayt məlumat generasiya olunmuşdursa, 2012-ci ildə müvafiq göstərici 2,7 zetabayta bərabər olmuş və növbəti hər il informasiyanın həcmi 40% artaraq 2020-ci ildə 44 zetabayta çatacağı proqnozlaşdırılır [1]. Rəqəmsal verilənlərin sürətlə artması ilə verilənlərin emalı, saxlanması və istifadəsində yeni eranı əks etdirən “böyük verilənlər” (*ing. big data*) termini meydana çıxmışdır [2]. Bu termin həcm və mürəkkəblik baxımından mövcud idarəetmə metodları və intellektual analiz vasitələri ilə emal oluna bilməyən böyük verilənlər massivini müəyyən etmək üçündür [3].

Big data fenomenal hadisə olaraq tez bir zamanda cəmiyyətin hər bir segmentinin diqqətini cəlb etmişdir. Çünki böyük verilənlərdə (BV) idarəetmə və biznesdə inqilabi dəyişikliklər, müəssisələrdə yüksək mənfəət əldə etmək, bir çox sahələrdə elmi ideyaların inkişaf etdirmək və reallaşdırmaq üçün böyük potensial vardır [3,4]. Verilənlərin analizi və onlardan biliklərin və faydalı məlumatların əldə olunması yeni-yeni elmi kəşflərin edilməsində, təşkilatlarda əsaslandırılmış qərarların qəbul olunmasında, milli təhlükəsizlik, səhiyyə sahələrində çox önəmlidir. Lakin BV cəmiyyətin iqtisadi inkişafına təkan verməklə yanaşı, elmi ictimaiyyəti bir çox problemlərlə üz-üzə qoymuş, yeni tədqiqat paradigması yaratmışdır [5–11].

BV-dəki potensialı reallaşdırmaq üçün, ilk növbədə, bir çox texniki, texnoloji problemlər həll olunmalıdır. BV-nin xarakterik xüsusiyyətlərindən (böyük həcm, yüksək sürət, müxtəliflik) irəli gələn bir sıra problemlər (hesablama, əhatəlilik, saxlanma, düzgün olmayan korrelyasiyalar və s.) də mövcuddur ki, onların həllində yeni elmi baxışlar, yanaşmalar, modelləşmə, riyazi metodlar, optimallaşma üsulları və s. tələb olunur. Problemlərin əsas istiqamətlərini isə arxitektura, saxlanma, ilkin emal, idarəetmə, analitika, nəticələrin interpretasiyası – vizuallaşdırma, təhlükəsizlik və s. kimi məsələlər təşkil edir. Bu problemlər tədqiqatçılardan real vaxtda BV-dən lazım olan informasiyanı, biliyi, məlumatı əldə etmək üçün daha effektiv statistik və intellektual analiz üsulları, əhatəliliyi təmin edən arxitektura yanaşmalar, vahid infrastruktur və alətlər tələb edir. Bu baxımdan, *big data*-nın ən aktual elmi-nəzəri problemlərinin tədqiqi olduqca vacibdir.

Big data konsepsiyası

Ümumiyyətlə, verilənlərin inkişaf tarixinə nəzər saldıqda görürük ki, hələ keçən əsrin 70-ci illərinin sonunda verilənlərin idarə olunması və analizində istifadə olunan texnologiya – “verilənlər bazası” idarəetmə sistemləri yaranmışdır [12]. Demək olar ki, bununla da “verilənlər bazası” konsepsiyasının əsası qoyulmuşdur. Ancaq verilənlərin həcmnin artması nəticəsində meynfreymlər (ümumi məqsədli universal elektrohesablama maşınları) saxlama və emal üçün adekvatlıq nümayiş etdirə bilməmişdir. Sonrakı illərdə problemin həllində “paralel verilənlər bazası sistemi” təklif olunmuşdur [13]. Bu sistemin arxitekturu klasterlərdən (klasterdə hər kompüter prosessor, yaddaş və diskdən ibarətdir) istifadəyə əsaslanır. Qeyd etmək lazımdır ki, ötən əsrin 90-cı illərinin sonuna kimi “paralel verilənlər bazası sistemi” çox populyar olmuşdur. Lakin İnternet xidmətlərinin çeşidləri artdıqca, böyük verilənlərin saxlama və emal problemi də çoxalmışdır. Problemin həllində hesablama arxitekturasında fundamental dəyişikliklər və genişlənmə bilən emal mexanizmləri tələb olunur. BV-nin problemləri ilə üz-üzə qalmış informasiya texnologiyaları (İT) sahəsinin nəhənglərindən olan *Google* şirkəti İnternet miqyasında verilənlərin idarə edilməsi və analizi üçün *File System Google* [14] və *MapReduce* [15] program-aparat platforması yaratmışdır. Açıq kodlu *Apache Hadoop* və *Hadoop Distributed File System* [16,17] program təminatları, *NoSQL* verilənlər bazası da işlənmişdir ki, bununla da *big data* texnologiyalarının əsası qoyulmuşdur. Bu texnologiya genişmiqyaslı verilənlər toplusunun saxlanması və idarə olunmasında ən düzgün seçim hesab olunur. Çünki bu texnologiya klasterləşmə məsələlərində, xüsusilə, veb-səhifələrin reytinginin müəyyən olunmasında daha effektivdir. Eyni zamanda, o, verilənlər xəzinəsi sistemlərinin (*data warehousing systems*) zəif cəhətlərini arxada qoymuş, daha dərin analitik vasitələrdən istifadə etməklə lazımi informasiyanın əldə olunmasını mümkün etmişdir [11].

Big data termininin tarixi və izahı bağlı bir çox tədqiqatlar mövcuddur [2,3,5–9,18-20]. Tədqiqatlar göstərir ki, *big data* tarixi dəqiq bəlli olan terminlərdəndir. Belə ki, *big data* termini ilk dəfə 1998-ci ildə ABŞ-ın *Silicon Graphics* kompüter şirkətinin kompüter elmləri üzrə mütəxəssisi Con Meşi tərəfindən istifadə edilmişdir. Bu terminə bir qədər sonra - 2000-ci ildə, Pensilvan universitetinin professoru, *big data* termininin əsas araşdırıcılarından olan Fransis Dieboldun akademik mühitdə dərc olunan tədqiqatında rast gəlinir [2]. Onun sonrakı tədqiqatında bu terminin bir qədər də möhkəmləndiyindən, artıq onun bir hadisə, bir fenomen deyil, ciddi tədqiqat istiqamətinə çevrildiyindən bəhs olunur [18]. Bütün bunlara baxmayaraq, termin Berkli universitetinin professoru Klifford Linçin elmi mühitdə dərc olunan tədqiqatından sonra populyarlıq qazanmışdır [21].

Tətbiq olunduğu sahələrdən asılı olmayaraq, *big data*-ni təsvir etmək üçün ümumi xüsusiyyətlər mövcuddur. Bu xüsusiyyətlər BV-nin əsas problemlərini özündə əks etdirməklə üç əsas qrupa bölünür: həcm (*volume*), sürət (*velocity*) və müxtəliflik (*variety*). İngilis dilli mənbələrdə bunu “3V” də adlandırırlar. Bu xüsusiyyətlər bir sıra elmi mənbələrdə geniş şərh olunmuşdur [3, 5–9, 20, 22, 23]. BV-ni təyin etməyə və digər verilənlərdən fərqləndirməyə imkan verən ilk model 2001-ci ildə *Gartner* şirkətinin analitiki Doung Laney tərəfindən verilmişdir [24]. O, elektron kommersiada bir tendensiyanı: verilənlərin idarə edilməsinin daha vacib, daha çətin olacağını əvvəlcədən xəbər vermiş və daha sonra verilənlərin idarə edilməsində onların ölçüsünü, ötürülmə sürətini və müxtəlifliyini əsas problem kimi müəyyən etmişdir. Qeyd olunan xüsusiyyətlər 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.

[3]-də bu xarakteristikalar aşağıdakı kimi şərh olunmuşdur:

Həcm (volume). Həcm BV-nin əsas xüsusiyyətidir və verilənlərin kəmiyyət göstəricisidir. Hazırda bu göstərici terabaytlardan zətabaytlara qədər həcm ilə ölçülür. 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, mövcud olanlar 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.

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 asan məsələ deyildir.

IBM şirkəti verilənlərin həqiqiliyini əsas götürərək 4-cü “V” (*veracity*), *Oracle* şirkəti isə BV-nin dəyərini qeyd etməklə 5-ci “V”ni (*value*) daxil etmişdir.

Həqiqilik (veracity). Həqiqilik dedikdə, verilənlərin keyfiyyəti (tam, natamam, ziddiyyətli və s.) başa düşülür. Verilənlərin keyfiyyəti böyük ölçüdə dəyişə bilər ki, bu dəyişkənlik də analizin nəticəsinə təsir edə bilər.

Dəyər (value). Verilənlər dəyər yaratmaq xüsusiyyətinə malik olmalıdır. BV dəyər yaratmırsa, “məlumat zibilliyi”nə çevrilir. *Big data*-nın biznes sektoru tərəfindən daim diqqətdə saxlanması əlavə dəyər yaratma xüsusiyyətinə görədir. Buna görə də bu amil marketinq xüsusiyyəti kimi qiymətləndirilir [1, 4]. Çünki informasiyanın dəyəri bizim onu necə istifadə etməyimizlə müəyyən olunur.

Onu da qeyd etmək lazımdır ki, son zamanlar mütəxəssislər tərəfindən “v”lərin sayı artırılmaqdadır.

***Big data*-nın tədqiqat paradigması kimi formalaşması amilləri**

Bu gün *big data* adlandırılan informasiya bolluğu, həqiqətən də, mövcuddur. Ancaq bu insanlarda məyusluq yaratmamalıdır, əksinə, ona təbii xammal, resurs kimi baxmaq lazımdır. Çünki bu xam verilənlərdə elmi kəşflərə səbəb ola biləcək dərin bilik toplanmışdır. Bu resursdan maksimum istifadə etməklə cəmiyyət və biznes sahəsində dəyər yaratmaq üçün yeni nəsil analitik texnologiyalara ehtiyac vardır. Bu baxımdan, BV mövzusu istər dövlət qurumlarında qərar qəbul edən şəxslər və siyasətçilər, istər biznes sektoru, istərsə də elmi tədqiqatçılar tərəfindən böyük diqqət çəkmiş və qeyd edildiyi kimi, yeni tədqiqat istiqamətinə çevrilmişdir. Bunu şərtləndirən bir çox amillər vardır.

İlk növbədə, qeyd etmək lazımdır ki, nüfuzlu beynəlxalq təşkilatlar, elmi qurumlar tərəfindən çox böyük həcmdə informasiyanın emalının müxtəlif aspektlərinə həsr olunmuş bir çox konfranslar, simpoziumlar, seminarlar, forumlar keçirilməkdədir. Bu tədbirlərin əsas müzakirə mövzularını: BV-nin 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 [3].

Elmi və populyar jurnalların xüsusi nömrələrinin *big data* mövzusunda həsr olunması və 2014-cü ildən başlayaraq BV-nin elmi-nəzəri problemlərini işıqlandıran yeni akademik jurnalların nəşri bu sahənin çox ciddi elmi istiqamət olduğunu göstərən əsas amillərdəndir.

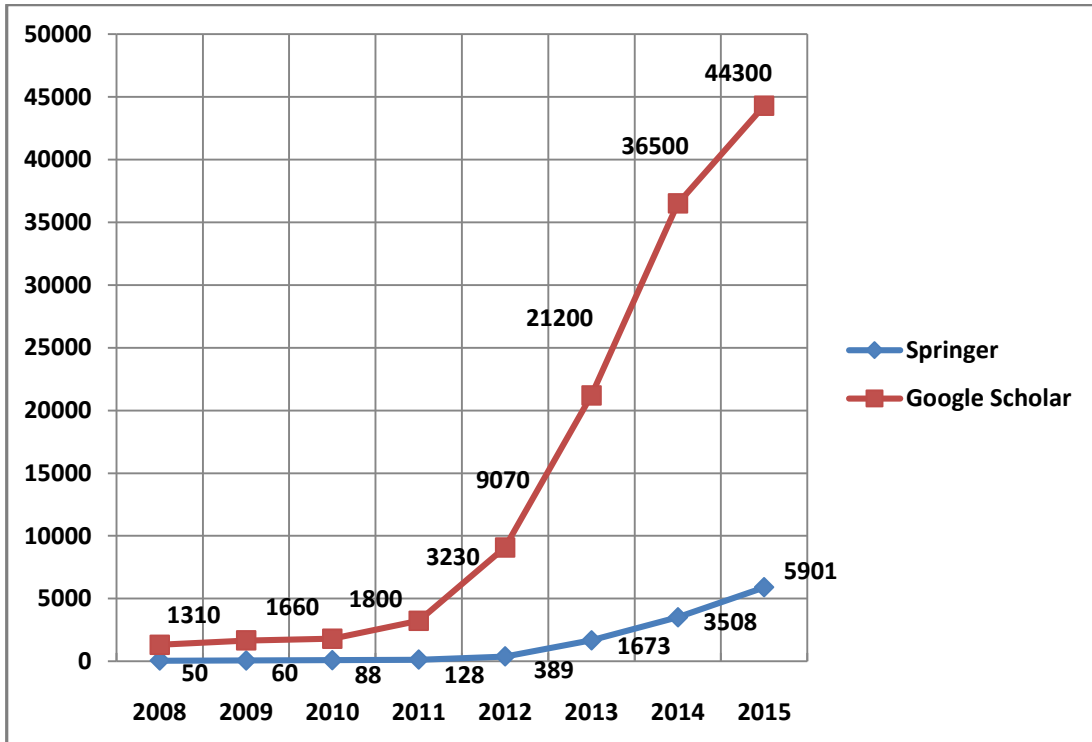
Dünyanın tanınmış elm mərkəzlərində BV-nin toplanması və emalı, saxlanması, arxitekturu, analitikası, təhlükəsizliyi, vizuallaşdırılması və s. istiqamətlərdə fundamental elmi-tədqiqat işləri aparılmaqdadır [3].

2013-cü ildən başlayaraq dünyanın bir çox aparıcı universitetlərində: Dandi (Şotlandiya), Oklend (Yeni Zelandiya), Cənubi Kaliforniya, Vaşinqton, Berkli, Nyu-York, London İmperial Kolleci, və s. “verilənlər haqqında elm” (*data science*) akademik fənn kimi bakalavr, magistr və doktorluq təhsil pilləsində tədris olunmağa başlanmışdır [6]. Belə proqramlar böyük həcmli verilənlərin potensialından maksimum istifadə etmək üçün hesablama modelləri,

modelləşdirmənin və proqnozlaşdırmanın riyazi metodları, arxitektura, müasir proqramlaşdırma metodları, verilənlərin əldə edilməsi, saxlanması və analizi kimi fundamental hazırlığı təmin etməyi nəzərdə tutur. Hazırda bu universitetlərin sayı getdikcə artmaqdadır. Türkiyədə Sabançı Universitetində, Rusiyada Ali İqtisadiyyat Məktəbində və s. magistr pilləsi üzrə proqram tədris olunmağa başlanmışdır. Qeyd edək ki, “*data scientist*” (“verilənlər elmi”) ixtisası son illərdə dəbdə olan perspektivli ixtisaslardandır. *Big data* texnologiyaları və müvafiq fundamental tədqiqatlar bu tədris mərkəzlərinin əsas elmi tədqiqat istiqamətlərinə çevrilmişdir.

Dünyanın bir sıra aparıcı elmi bazalarında bibliometrik təhlillər də BV-nin tədqiqat mövzusu kimi inkişafını sübut edən əsas faktorlardandır [8-11, 25]. Bu halda mövzunun inkişafına zaman, məkan, elm sahələri, çap olunmuş tədqiqatların sayı, tipi (kitab, məqalə, konfrans materialları və s.) və s. kimi bibliometrik göstəricilər kontekstində baxmaq mümkündür. Məsələn, [10, 25]-də *Scopus* bazasında “*big data*” açar sözü üzrə aparılmış axtarışın nəticəsi olaraq 2008-ci ildə çap olunmuş yalnız bir elmi-tədqiqat işinə rast gəlinirdisə, 2012-ci ildən başlayaraq elmi-tədqiqat materiallarının sayının dəfələrlə artdığı görünür. Tədqiqatların coğrafi paylanması isə ilk yerləri ABŞ, Çin, Hindistan, Böyük Britaniya, Cənubi Koreya və s. ölkələr bölüşürlər.

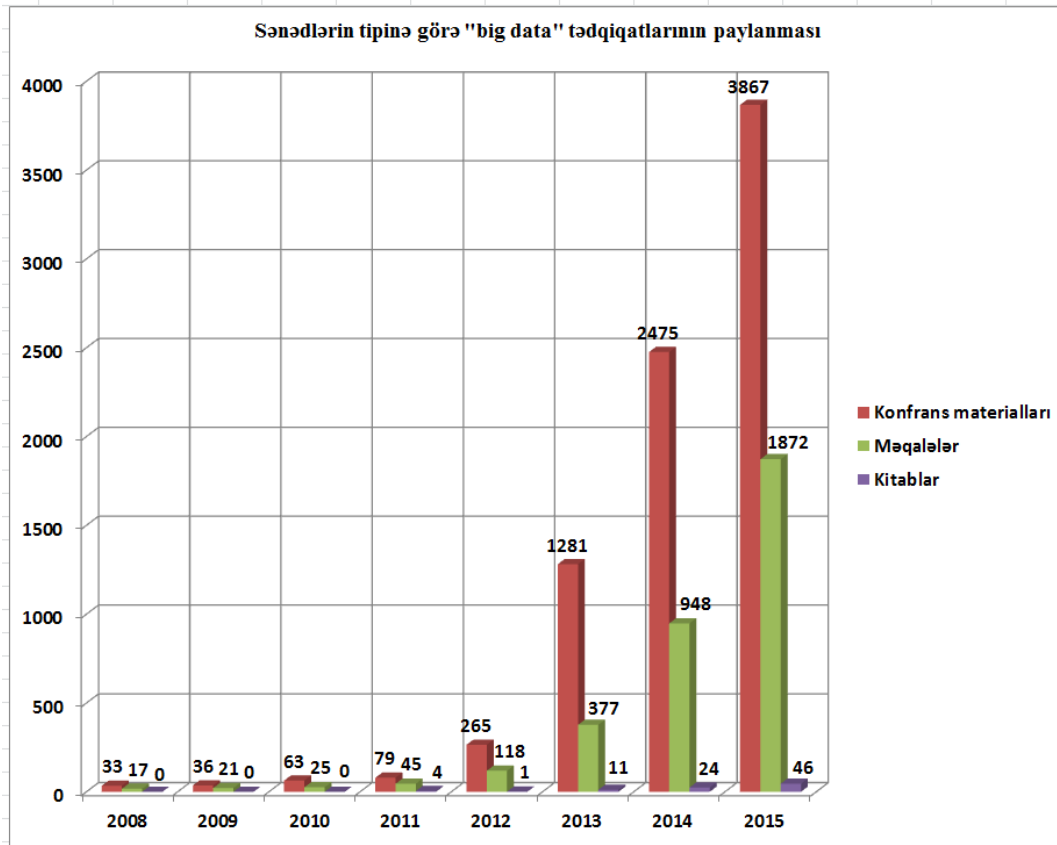
Google Scholar və *Springer* bazasında apardığımız araşdırmanın nəticəsi də [10, 25]-ə çox oxşardır. Adı çəkilən bazalarda tədqiqatların illər üzrə dinamikası aşağıdakı qrafikdə verilir (Şəkil 1).



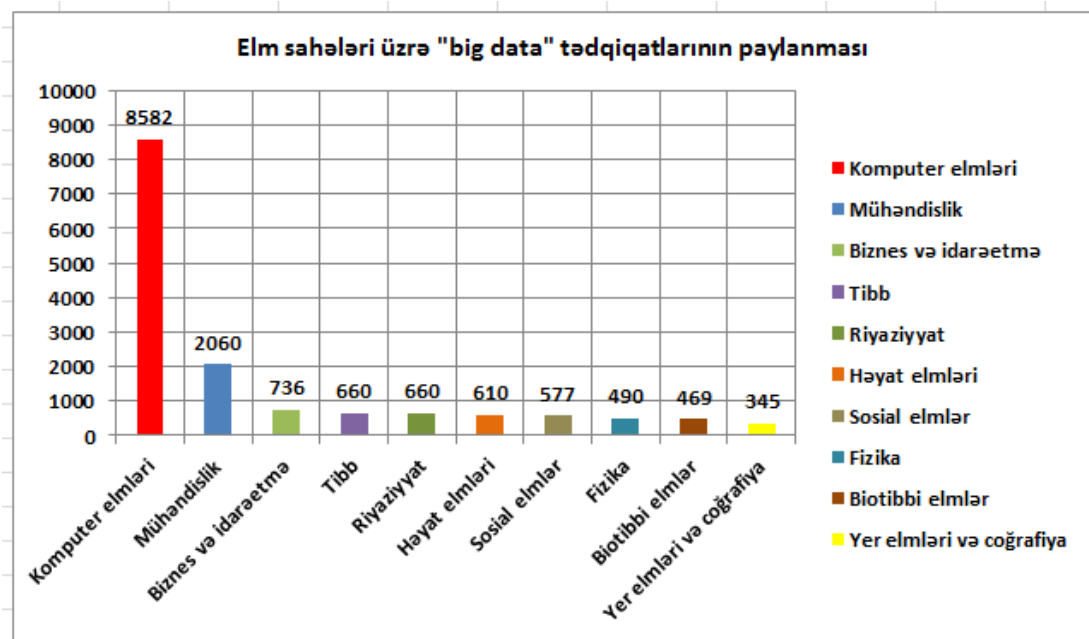
Şəkil 1. *Google Scholar* və *Springer* bazalarında *big data* üzrə tədqiqatların paylanması

Springer bazasında sənədlərin tipinə və elm sahələrinə görə paylanması isə uyğun olaraq Şəkil 2 və Şəkil 3-də təqdim olunmuşdur.

Qrafikdən də göründüyü kimi, materiallarda kompüter elmləri, mühəndislik, idarəetmə, tibb, riyaziyyat, sosial elmlər kimi sahələrlə bağlı tədqiqatlar geniş yer tutur. Bibliometrik göstəricilər müxtəlif elm sahələrində verilənlərin artan əhəmiyyətini və *big data*-nın elmi istiqamət kimi formalaşdığını bir daha sübut edir.



Şəkil 2. Springer bazasında *big data* üzrə sənədlərin tipinə görə tədqiqatların paylanması



Şəkil 3. Springer bazasında *big data* üzrə tədqiqatların paylanması

Ümumiyyətlə, *big data* elmi-tədqiqatların metodologiyasında, elmi düşüncə və metodlarda inqilabi dəyişikliyə səbəb olmuşdur. Verilənlərdən yeni biliklərin alınması bu resursun intensiv emalına əsaslanan elmi tədqiqatların əsasını qoymuşdur [6,11]. 2007-ci ildə Turing mükafatı laureatı Cim Qrey dörd elmi paradigmanı qeyd etmişdi. Belə ki, min il əvvəl elm empirik (*empirical*) idi, son bir neçə yüz il nəzəri (*theoretical*), son bir neçə onilliklər hesablayıcı

(*computational*), bu gün isə verilənlərin intensiv emalına əsaslanan elmi, hesablama elmindən ayıran elektron elmi (*data exploration – e-Science*) dördüncü paradigma adlandırmışdır [26]. C.Qrey hesab edirdi ki, bu gün bəşəriyyətin qarşılaşdığı mürəkkəb qlobal çağırışların bir çoxunun həlli üçün yalnız elmi tədqiqatların dördüncü paradigması sistemli yol ola bilər. Bu paradigmaya uyğun olaraq verilənlərin intensiv emalına əsaslanan elm sahələri inkişaf etməyə, elmi tədqiqatlar isə müasir elmi kəşflərin əsas mənbəyi kimi çıxış edən verilənlərdən daha da asılı olmağa başlamışdır [26]. Dördüncü paradigmanın inkişafı böyük həcmli verilənlərlə stimullaşdırılır, bir çox elm sahələri yüksək səviyyədə verilənlərlə idarə olunur [9]. Məsələn, astronomiya [27], sosial hesablamlar [28], bioinformatika [29] və hesablama biologiyası, meteorologiya və s. kimi elm sahələri böyük həcmli verilənlərin intensiv emalına əsaslanır. İnkişaf etməkdə olan fənlərarası sahə adlanan “*data science*” (“verilənlər elmi”) tədricən öz yerini tutmaqdadır [6, 30, 31]. Bu sahənin tədqiqat obyektini isə böyük verilənlərdir və verilənlərdən ümumiləşdirilmiş biliyin çıxarılmasına istiqamətlənmişdir. Verilənlər elmi informatika, riyaziyyat, sosial elmlər, şəbəkə elmi, iqtisadiyyat və s. daxil olmaqla bir çox fənlərlə çulğalaşmışdır. Odur ki, BV-nin emalı və analitikasının fundamental əsaslarını riyaziyyat, statistika, informatika ilə yanaşı müxtəlif metod və nəzəriyyələr, o cümlədən, ehtimal nəzəriyyəsi, obrazların tanınması, neyron şəbəkələr, intellektual analiz [32,33], maşın təlimi [34], siqnalların emalı, təbii dilin emalı, proqnoz analiz, vizuallaşdırma [35,36], optimallaşdırma, statistik metodlar [37], proqramlaşdırma, mühəndislik, qeyri-müəyyənliklərin modelləşdirilməsi, yüksəkməhsuldarlıqlı hesablamlar və s. təşkil edir.

Big data-nın elmi-nəzəri problemləri

Əlbəttə, son illər bəşəriyyəti əhatə edən sirlərin açılmasıyla bağlı aparılan elmi-tədqiqatların nailiyyətləri məhz BV-nin imkanlarından səmərəli istifadənin nəticəsidir [9]. Ancaq BV-dən gizli biliklərin aşkarlanması və bu biliklər əsasında qərarların qəbul edilməsi verilənlərin təşkili, emalı baxımından mürəkkəb məsələ – “*big data computing*” adlanan yeni paradigmadır [11]. Bu yeni paradigma genişmiqyaslı saxlama, emal və hesablama metod və modellərini özündə birləşdirir. Yəni 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 açıq məsələlərin həllində yeni elmi yanaşmalar, daha mükəmməl analiz üsulları tələb olunur [9].

[38]-də BV-yə xas olan problemlər üç qrupa bölünür.

1. *Verilənlər problemi* həcm, sürət, müxtəliflik, həqiqiliklə əlaqədardır. Belə ki, verilənlərin həcmi əvvəlki dövrlərlə müqayisədə daim artır və mövcud alətlər onları emal etmək imkanında deyildir. Verilənlər müxtəlif tipə (mətn, sensor, audio, video, qrafik və b.) malikdir. Verilənlər daim axın şəklində generasiya olunur və real vaxt rejimində onlardan faydalı informasiya əldə etmək lazım gəlir. Verilənlər toplusundakı uyğunsuzluq onların emal və idarə olunması prosesini çətinləşdirə bilər. Verilənlərin keyfiyyəti dəyişə bilər və bu dəyişiklik analizin nəticəsinə təsir edə bilər.
2. *Emal probleminə* analiz üçün məqbul hesab olunan məlumatların toplanması, müxtəlif mənbələrdən həll oxşarlıqlarının tapılması daxildir. Həmçinin buraya analizin özü və çıxış təqdimatı, yəni nəticələrin insanın anlaması üçün ən uyğun bir formada vizuallaşdırılması daxildir.
3. *Verilənlərin idarə edilməsi* təklif olunan təsnifatın sonuncusudur. İdarəetmə məsələsi, əsasən, verilənlərin analiz üçün hazırlanması, saxlanması, əldə edilməsi, məxfiliyi, təhlükəsizliyinin təmin olunması, ümumiyyətlə, həyat dövrünün idarə edilməsi ilə əlaqədardır. Yəni verilənlərin düzgün istifadəsini təmin etməkdir. Buna da dövlət və beynəlxalq səviyyədə nəzərdə tutulmuş informasiya təhlükəsizliyi siyasəti və qaydaları əsasında nəzarət olunur.

Böyük verilənlərin analizi problemləri. Ümumiyyətlə, verilənlərin həcmnin kəskin surətdə artması və real zaman rejimində onların analizinə dair tələbat BV-nin ən əsas problemlərindən sayılan böyük verilənlərin analitikasının (*Big Data Analytics*) yaranmasına gətirib çıxartdı. O, optimal (ən yaxşı) qərarlar qəbul etmək üçün böyük həcmli verilənlərdə gizli qanunauyğunluqların, məlum olmayan korrelyasiyaların və digər faydalı informasiyanın aşkarlanması prosesidir. *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, analitikanın mahiyyəti birdir – qərar qəbul edən şəxsləri müxtəlif proseslər haqqında məlumatlarla təmin edən əks əlaqəni yaratmaqdır [3].

[39]-da BV analitikasının üç tipi qeyd olunur: təsviri analitika (*big data descriptive analytics*), proqnostik analitika (*big data predictive analytics*), göstərən analitika (*big data prescriptive analytics*).

Big Data descriptive analytics “nə baş vermişdir?”, “nə üçün baş vermişdir?” kimi suallara cavab verir.

Big Data predicative analytics “nə baş verəcəkdir?” sualına cavab verir.

Big Data prescriptive analytics isə yalnız “nə baş verəcəyini”, “nə vaxt baş verəcəyini” deyil, həm də “nə üçün baş verəcəyini” göstərir.

Ümumiyyətlə, BV-nin analitikası onun əsas problemlərindəndir. Bu problem BV-nin xüsusiyyətləri, mövcud analiz model və metodları və verilənlərin emal sistemlərinin məhdudluqları ilə bağlıdır [9]. [6, 40]-da analitikanın əsas problemləri kimi müxtəlif tip verilənlərin inteqrasiyası, həcmi, miqyaslanması, təhlükəsizliyi, uyğunsuzluğu və s. göstərilmişdir. Əlbəttə, verilənlərin çox da sadə olmayan analiz üsulları mövcuddur. Lakin verilənlərin ənənəvi analiz metodlarının əksəriyyəti müxtəlif situasiyalara dinamik uyğunlaşmaq qabiliyyətinə, miqyaslanma xarakterinə malik deyil, paralel hesablama mühitində işləmirlər. Mətn, video və audio kimi strukturlaşdırılmamış informasiyanın analizi və müxtəlif mənbələrdən informasiyanın əldə olunması və onların inteqrasiyası ciddi problemdir. Yəni mövcud metodların əksəriyyəti böyük və mürəkkəb verilənlər üçün yetərli deyildir [41]. Məlumdur ki, analiz müxtəlif parametrlər, xüsusiyyətlər, hadisələr və s. arasındakı korrelyasiyanı tapmağa, təsnifatlandırma, analitik hesabatlar və bunların əsasında proqnozların verilməsinə imkan verir. Bu aspektdən müasir texnologiyalar verilənlərdəki informasiyanın yeni biliklərə çevrilməsinə və ya biliklərin əldə edilməsinə imkan verməlidirlər. BV-nin saxlanması, emalı və analizi üçün böyük hesablama gücü, miqyaslılığı təmin edən arxitektura yanaşmalar, vahid infrastruktur tələb olunur. Elə *big data mining* metodları işlənməlidir ki, onlar, ilk növbədə, verilənlərdə dəyişikliyi aşkarlaya bilsinlər. Bir sözlə, mövcud model və metodların paylanmış versiyasını yaratmaq və ya yenilərinin işlənməsi üçün praktiki və nəzəri tədqiqatların aparılmasına ehtiyac vardır.

BV analitikasının əsas istiqamətlərini mətn, video, audio və sosial media analitikası təşkil edir [8].

Mətn analizi (text analytics) – *data mining* sinfindən olan üsulların köməyi ilə təbii dildə mətnlərdən əvvəlcədən məlum olmayan əlaqələri və korrelyasiyaları aşkarlamaqla bilikləri çıxarır [32, 33]. Təsnifatlandırma (*classification*) və klasterləşdirmə (*clustering*), informasiyanın çıxarılması (*information extraction*), avtomatik referatlaşdırma (*summarization*) isə *text mining*-in həll etdiyi əsas problemlərdir. Mahiyyət etibarilə, *text mining* mətnləri daha dərin analiz etmək üçün linqvistik, statistik üsulların, informasiya axtarışı, həmçinin maşın təlimi alqoritmlərini istifadə edir [8, 42].

Video analitika (video analytics) – video axınının monitorinqi, analizi və onlardan faydalı informasiyanın çıxarılmasını özündə birləşdirir və *big data* baxımından çox ciddi problemdir. Video-məlumatlar indi rəqəmsal informasiyanın, müşahidələrin əsas formasıdır və böyük ölçüyə malikdir. Bu tip yazıların analitikası verilənlərin digər analiz növləri ilə müqayisədə başlanğıc mərhələsindədir. 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 [8].

Audio-analitika (audio analytics) – 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 [8].

Sosial media analitikası (social media analytics) – sosial media kanallarının strukturlaşdırılmış və strukturlaşdırılmamış verilənlərinin analizini həyata keçirir, istifadəçilərin davranışını proqnozlaşdırmaq üçün istifadə olunur. Sosial medialarda kontent çox vaxt böyük həcmli, küylü və dinamik olur. Ona görə də mətn, audio və video analitikasında qeyd olunan problemlər, həmçinin BV-nin ötürülməsi onun əsas problemlərindəndir [8, 40, 43, 44].

Böyük verilənlərin arxitektura problemləri. 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*) və məxfilik (*Privacy*) daxildir. [19, 45]-də böyük həcmli verilənlərin analizi üçün *Hadoop*, *MapReduce*, *NoSQL* və s. kimi mövcud etalon arxitektura və platformaların icmalı verilir. Hələlik aydın deyildir ki, retrospektiv və real zaman verilənlərinin eyni vaxtda emalı üçün analitik sistemlərin optimal arxitekturu necə olmalıdır.

Böyük verilənlərin emalı və saxlanması problemləri. *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 [9]; *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 da “hesablamalar deyil, verilənlər daha əsas sərvətdir” anlamına gəlir [3, 9, 46].

Big data verilənlərin emalı və saxlanması üsullarını, saxlanma qurğularını, saxlanması arxitekturasını, verilənlərə giriş mexanizmlərini köklü surətdə dəyişmişdir. Saxlanma qurğuları böyük həcmli verilənlərin əlyətərliliyini və operativ analizini təmin etmək imkanına malik olmalıdır [9]. Hazırda saxlama məsələsinin həllində informasiyanın qurğular arasında miqrasiyasını həyata keçirmək imkanına malik olan *DAS (Direct Attach Storage*, birbaşa qoşulmaqla saxlama), *NAS (Network Attached Storage*, şəbəkəyə qoşulmaqla saxlama), *SAN (Storage Area Networks*, verilənlərin saxlanması şəbəkəsi), *HSM (Hierarchical Storage Management*, verilənlərin iyerarxik idarə edilməsi), *ILM (Information Life-cycle Management*, informasiyanın həyat dövrünün idarə olunması) kimi texnologiyalar geniş istifadə olunur [3, 6]. Lakin irimiqyaslı paylanmış sistemlər səviyyəsinə gəldikdə isə verilənlərin bu saxlanma arxitekturlarının çatışmazlıqları və məhdudiyyətləri ortaya çıxır. Son zamanlar isə saxlama qurğularının yaddaş tutumunun artırılması, hesablama və yaddaş resurslarının klasterləşdirilməsi və virtuallaşdırılmasını həyata keçirən *grid* və *cloud computing* texnologiyalarının tətbiqi saxlama sahəsindəki problemləri, demək olar ki, aradan qaldıra bilmişdir [3, 9]. Bulud hesablamaları (*cloud computing*) saxlanma, böyük hesablamaların aparılmasında son dərəcə müvəffəqiyyətli yanaşmalardandır. Digər tərəfdən, bulud saxlaması verilənlərin tamlığına nəzarət baxımından onların təhlükəsizliyi problemini yaradır [9]. BV-nin emal və saxlama sistemlərinin enerji səmərəliliyinin qiymətləndirilməsi və optimallaşdırılması da çox ciddi elmi-tədqiqat problemlərindəndir [7, 8].

Böyük verilənlərin vizuallaşdırma problemləri. BV-nin analizində əsas məsələlərdən biri də nəticələrin təqdim olunması – vizuallaşdırılmasıdır. Ümumiyyətlə, verilənlərin vizuallaşdırılması onların emalı və analizində ən sadə və təbii üsullardandır. Bu üsul insanın təqdim olunmuş informasiya ilə daha tez tanış olmaq və nəticələri düzgün qiymətləndirməklə optimal qərar verməyinə kömək edir [36, 35].

Böyük ölçüyə və həcmə malik verilənlərin vizuallaşdırılması o qədər də sadə məsələ deyil. BV-nin xüsusiyyətləri nəzərə alınmaqla onların vizuallaşdırılması zamanı yaranan əsas problemlər aşağıdakılardır [35]:

- **vizual küy** (*visual noise*). Bu problem verilənlər dəstində obyektlərin bir-biri ilə həddindən artıq əlaqəli olmasından yaranır. Küy dedikdə, verilənlərin korlanması, təhrif olunması deyil, sadəcə, ekranda obyektlərin ayrılıqda görüntüsünün kiçilməsi və ya itməsidir. Bu da bütün görüntüdə faydalı informasiyanın alınmasını çətinləşdirir və əlavə emal lazım gəlir.
- **böyük təsvirin qavranılması** (*large image reception*). İnsanın beyni bəlli səviyyədə vizual görüntünü qəbul edə bilir. Verilənlərin qrafik vizuallaşdırılmasının dərk olunması səviyyəsi cədvəl vizuallaşdırılmasından yüksək olsa da, müəyyən məhdudiyyətlər mövcuddur. Belə ki, müəyyən səviyyəni keçdikdə insan həddindən artıq yüklənmiş vizual verilənlərdən əlavə informasiya əldə etmək qabiliyyətini itirir. Əlbəttə, vizuallaşdırma metodları verilənlərin çıxışda görüntüsünü təmin edən texniki qurğuların imkanları ilə məhduddur. Biz nə qədər daha müasir qurğu istifadə etsək də, yenə insanın dərk etmə məhdudiyyəti ilə qarşılaşırıq. Demək, verilənləri vizuallaşdırma metodları təkcə qurğuların imkanları ilə deyil, həm də insanın fiziki qavrayışları ilə məhdudlaşır. Qeyd olunan problemin həllində verilənlərin filtirlənməsi – kiçildilməsi yanaşmasından istifadə edilir.
- **informasiya itkisi** (*information loss*). Bu, vizual küy və böyük təsvirin qavranılması probleminin həllindən doğan problemdir. Qeyd olunan problemin həllinə tətbiq edilən yanaşmalar sonda istifadə olunan verilənləri azaltsa da, informasiya itkisi adlanan yeni problemin yaranmasına gətirib çıxarır. Çünki vizual informasiyanın azaldılması metodları obyektlərin yaxınlığı əsasında verilənlərin bir və ya bir neçə meyarə görə aqreqasiya və filtirlənməsini həyata keçirir. Bu yanaşmalar analitikləri yanılma bilər, daha vacib və maraq kəsb edən gizli obyektlər diqqətdən kənar qala bilər. Həmçinin dəqiq və lazımi informasiyanı almaq üçün verilənlərin aqreqasiyası prosesi böyük zaman və hesablama resursu tələb edə bilər.
- **yüksək məhsuldarlıq tələbləri** (*high performance requirements*). Qrafik analiz təsvirlərin yalnız statik vizuallaşdırılması ilə məhdudlaşmır, dinamik vizuallaşdırmadan da istifadə edilir ki, bu zaman da statik vizuallaşdırmada nəzərə çarpmayan bir problem yaranır. Vizuallaşdırmanın müəyyən sürətində prosesin məhsuldarlığına tələb yaranır. Çünki analiz prosesində böyük sayda verilənlərin filtirlənməsi üçün çox zaman və hesablama resursu tələb olunur.
- **təsvirlərin yüksək sürətlə dəyişməsi** (*high rate of image change*). Adından da göründüyü kimi, bu problem təsvirlərin yüksək sürətlə dəyişməsi ilə əlaqədardır. Yəni müşahidə zamanı insan sadəcə olaraq verilənlərin sürətli dəyişməsinə və ya onların ekrandakı intensivliyinə reaksiya göstərə bilmir. Dəyişən verilənlərin sürətinin azalması prosesin arzu olunan effektivliyini təmin edə bilmir. Ancaq insan reaksiyasının sürəti bu prosesdə müəyyən məhdudiyyətlər yaradır.

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

Böyük verilənlərin təhlükəsizlik problemləri. Ümumiyyətlə, verilənlərin təhlükəsizliyi hər zaman önəmli məsələlərdən hesab olunur. BV-nin meydana gəlməsiylə informasiya təhlükəsizliyi

baxımından yeni problemlər yaranmışdır. Problemlərə iki aspektdən yanaşmaq olar: 1) informasiya təhlükəsizliyi üçün *big data* analitikasının tətbiqi; 2) *big data* analitikasında informasiya təhlükəsizliyi [47, 48]. Hər iki məqam elmi tədqiqatçılar qarşısında duran ən aktual məsələlərdəndir. Belə ki, *big data* texnologiyalarının tətbiqi 15 il əvvəl istifadə olunan mövcud təhlükəsizlik modellərinin köhnəliyini, bu gün üçün adekvat olmadığını göstərmişdir. İnformasiya nə qədər rəqəmsallaşdırılırsa və əlavə informasiya toplanırsa, bir o qədər əlyətərli olur və onun istifadəçilərinin sayı da çoxalır. Nəticədə bədnəyyətçilər tərəfindən informasiyanın oğurlanması, təhrif olunması və şəbəkələrin sındırılması, fərdi məlumatların asanlıqla ələ keçməsi və s. kimi informasiya təhlükəsizliyi insidentləri yaranır.

İnsanların razılığı olmadan onlara məxsus fərdi məlumatların analiz olunması etik və hüquqi cəhətdən yolverilməzdir, təhlükəsizlik və gizlilik baxımından çox ciddi problemdir [49].

BV-nin həcm, müxtəliflik və sürət kimi xüsusiyyətləri təhlükəsizlik və gizlilik problemini daha da kəskinləşdirmişdir. 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 artırma bilər [9]. Genişmiqyaslı “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 şəraitində ənənəvi təhlükəsizlik mexanizmləri yetərli deyildir. Eyni zamanda, verilənlərin axını çox çevik və sürətli təhlükəsizlik həlləri tələb edir [9, 47-49]. İnformasiya təhlükəsizliyi üzrə aparılan elmi-tədqiqatın əsas istiqamətlərini təşkil edən kibercümlərin aşkarlanması; informasiya sistemlərinin təhlükəsizliyi, təhlükəsizlik risklərinin idarə edilməsi, informasiya təhlükəsizliyinin qiymətləndirilməsi və s. kimi problemlər BV ilə bağlı oxşar problemin həllində tədqiqatçıları düşündürən əsas məsələlərdəndir [47].

Nəticə

Big data çox qiymətli bilik mənbəyi kimi bütün dünyada dövlət, biznes və elmi dairələrin diqqətini cəlb etməklə informasiya texnologiyaları sahəsinin ən səs-küylü mövzusunə və yeni multidissiplinar elmi tədqiqat istiqamətinə çevrilmişdir. Elmi tədqiqatların əsas istiqamətlərini isə böyük verilənlərin konsepsiyasının əsasını təşkil edən həcm, sürət və mürəkkəblilik kimi təbii xüsusiyyətlərindən doğan elmtutumlu məsələlər təşkil edir. Problemlər qrupuna verilənlərin toplanmasından tutmuş nəticənin istifadəçiyə təqdim olunmasına qədər olan bir çox məsələlər daxildir. 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. Məsələlərin həllində daha effektiv *data mining* sinfinin üsulları (assosiativ qaydalar, reqressiya, 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 və s. kimi metodların işlənməsi tədqiqatçıların qarşısında duran əsas məsələlərdəndir.

Ədəbiyyat

1. The digital universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Study report, IDC, November 2014, <http://www.emc.com/leadership/digital-universe/index.htm>
2. Diebold F. Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting / Discussion Read to the Eighth World Congress of the Econometric Society, Cambridge: Cambridge University Press, 2000, pp. 115-122.
3. Əliquliyev R.M., Hacırəhimova M. Ş. "Big Data" fenomeni: problemlər və imkanlar // İnformasiya texnologiyaları problemləri, 2014, №2, s. 3-16.

4. Big data: The next frontier for innovation, competition, and productivity. Analyst report, McKinsey Global Institute, May 2011, <http://www.mckinsey.com>
5. Fan J., Han F. & Liu H. Challenges of Big Data analysis // National Science Review, 2014, vol. 1, no. 2, pp. 293–314.
6. Chen P.L., Zhang C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data // Information Sciences, 2014, vol. 275, pp.314–347.
7. Chen M, Mao S, Liu Y. Big data survey // Mobile Networks and Applications, 2014, vol.19, no.2, pp.171–209.
8. Gandomi A., Haider M. Beyond the hype: Big data concepts, methods, and analytics // International Journal of Information Management, 2015, vol. 35, pp. 137–144.
9. Jina X., Benjamin W. Waha, Chenga X., Wang Y. Significance and Challenges of Big Data Research, 2015, vol.2, no.2, pp. 59–64.
10. Halevi G. The Evolution of Big Data as a Research and Scientific Topic // Research Trends, 2012, no.30, pp.3-6.
11. Raghavendra K. et.all. The anatomy of big data computing // Software: practice and experience, 2016, no. 46, pp.79–105.
12. Codd E. F. A Relational Model of Data for Large Shared Databanks // Communication ACM, 1970, vol.13, no.6, pp.377-387.
13. DeWitt D., Gray J. Parallel database systems: the future of high performance database systems // Communication ACM, 1992, vol.35, no.6, pp. 85–98.
14. Ghemawat S., Gobiuff H. and Leung S.T. The Google File System / Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles, New York, USA, October 2003, pp. 29–43.
15. Dean J., Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters / Proceedings of the Sixth Symposium on Operating System Design and Implementation, volume 6 of OSDI '04, Berkeley, CA, USA, 2004, pp.137–150.
16. Hadoop MapReduce, http://www.hadoop.apache.org/docs/stable/mapred_tutorial.html
17. Hadoop Distributed File System, <http://www.hadoop.apache.org/docs>
18. Diebold F. On the Origin(s) and Development of the Term "Big Data". Pier working paper archive, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania, 2012, http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf
19. Maier M. Towards a Big Data Reference Architecture, 2013, http://www.win.tue.nl/~gfletche/Maier_MSc_thesis.pdf
20. İmamverdiyev Y.N. Big data texnologiyalarının böyük perspektivləri və problemləri // İnformasiya cəmiyyəti problemləri, 2016, №1, s. 23-34.
21. Clifford L. Big data: How do your data grow? // Nature, 2008, vol.455, pp. 28–29.
22. Kaisler S. et.all. Money W. Big Data: issues and challenges moving forward / Proceedings of the 46th Hawaii International Conference on System Sciences, 2013, pp. 995–1004.
23. Tole A.A, et.all. Big Data challenges // Database Systems Journal, 2013, vol. 4, no. 3, pp. 31–40.
24. Laney D. 3D Data Management: Controlling Data Volume, Velocity and Variety. Technical report, META Group, Inc (now Gartner, Inc.), February 2001, <http://www.blogs.gartner.com/doug-laney/files/2012/01>
25. Alıquliyev R.M., İsmayılova N.T. Bibliometric Analysis of Big Data Research / “Big data: imkanları, multidissiplinar problemləri və perspektivləri” I respublika elmi-praktiki konfransının əsərləri, Bakı şəhəri, 25 fevral 2016-cı il, səh. 58-60.
26. Hey T., Tansley S., Tolle K. (Eds.), The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Corporation, 2009, 287 p.
27. Zhang Y., Zhao Y. Astronomy in the Big Data Era // Data Science Journal, 2015, vol. 14, no.11, pp.1-9.

28. Hays R. R., Daker-W. G. The care.data consensus? A qualitative analysis of opinions expressed on Twitter // *BMC Public Health*, 2015, vol. 15, no. 838, pp. 2-13.
29. Greene C.S., Tan J., Ung M., Moore J.H., and Cheng C. Big data bioinformatics // *Journal of Cellular Physiology*, 2014, vol.229, no.12, pp.1896–1900.
30. Wu Z. From Big Data to Data Science: A Multi-disciplinary Perspective // *Big Data Research*, 2014, vol. 1, p.1.
31. Jagadish H.V. Big Data and Science: Myths and Reality // *Big Data Research*, 2015, vol. 2, no 2, pp. 49–52.
32. Wu X., Zhu X., Wu G.Q., Ding W. Data mining with bigdata // *IEEE Transactionson Knowledge and Data Engineering*, 2014, vol.26, no. 1, pp. 97–107.
33. Alguliev 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.
34. Omar Y.Al-J. et.all. Efficient Machine Learning for Big Data: A Review // *Big Data Research*, 2015, vol. 2, no. 3, pp. 87–93.
35. Gorodov E., Gubarev V. Analytical Rewiew of Data Visalization Methods in Application to Big Data // *Journal of Electrical and Computer Engineering*, 2013, 1-7 p.
36. Olshannikova E. et.all. Visualizing Big Data with augmented and virtual reality: challenges and research agenda//*Journal of Big Data*, 2015, vol. 2, pp.2-22.
37. 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.
38. Akerkar R. *Big Data computing*. Boca Raton, FL: CRC Press, Taylor&Francis Group, 2013, 562 p.
39. Sun Z., Pambel F., Wang F. Incorporating big data analytics into enterprise information systems, *Lecture Notes in Computer Science*, 2015, vol. 9357, pp. 300-309.
40. Kambatla K., Kollias G., Kumar V., Grama A. Trends big data analytics// *Parallel and Distributed Computing*, 2014, vol.74, no.7, pp. 2561-2573.
41. Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao and Athanasios V. Vasilakos, Big data analytics: a survey // *Journal of Big Data*, 2015, 2(21), 1-32.
42. Jiang J. Information extraction from text. In C. C. Aggarwal, & C. Zhai (Eds.), *Mining text data* (pp. 11–41). United States: Springer, 2012.
43. Kalampokis E., Tambouris E. and Tarabanis, K. Understanding the Predictive Power of Social Media // *Internet Research*, 2013, vol. 23, no. 5, pp. 544–559.
44. Barbier, G., & Liu, H. Data mining in social media. In C. C. Aggarwal (Ed.), *Social network data analytics*, United States: Springer, 2011. pp. 327–352.
45. Pekka P., Pakkala D. Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems // *Big Data Research*, 2015, vol. 2, no 4, pp.166-186.
46. Клеменков П.А., КузнецовС.Д. Большие данные: современные подходы к хранению и обработке // *Труды Института системного программирования РАН*, том 23, 2012, с.143-156.
47. Alguliyev R., Imamverdiyev Y. Big Data: Big promises for information security / *Proceedings of the IEEE 8th International Conference on Application of Information and Communication Technologies (AICT)*, Astana, Kazakhstan 15-17 October, 2014, pp. 216–219.
48. Hacırahimova M., “Big data” texnologiyaları və informasiya təhlükəsizliyi problemləri // *İnformasiya texnologiyaları problemləri*, 2016, №1, s. 49-56.
49. Lei X., Chunxiao J., Jian W., Jian Y., Yong R., Information Security in Big Data: Privacy and Data Mining // *IEEE Access*, 2014, vol.2, pp.1149–1176.

УДК 004.9:351

Алигулиев Расим М.¹, Гаджирагимова Макруфа Ш.², Алиева Айбенлиз С.³

^{1,2,3}Институт Информационных Технологий НАНА, Баку, Азербайджан

¹secretary@iit.ab.az, ²makrufa@science.az, ³aybeniz63@rambler.ru

Актуальные научно-теоретические проблемы больших данных

В связи с техническим и технологическим развитием за последние 10 лет в мире были сгенерированы цифровые данные очень больших размеров. На сегодняшний день это превратилось в самую обсуждаемую тему везде, где только возможно: в газетах, журнальных статьях, блогах и т.д. Современному обществу Big Data предоставила новые возможности, а для научных исследователей – проблемы. В отличие от массовых средств информации и сектора бизнеса, в статье термин «Big Data» рассматривается как научно-исследовательский объект. Представляется краткое изложение концепции Big Data. Берутся во внимание основные факторы его развития в роли исследуемого направления. Кроме того, анализируются актуальные научно-теоретические проблемы, которые находятся в центре внимания исследователей.

Ключевые слова: большие данные, аналитика больших данных, аудиоаналитика, видеоаналитика, социальные медиа-аналитика, визуализация, безопасность.

Rasim M. Aliguliyev¹, Makrufa Sh. Hajirahimova², Aybaniz S. Aliyeva³

^{1,2,3}Institute of Information Technology of ANAS, Baku, Azerbaijan

¹secretary@iit.ab.az, ²makrufa@science.az, ³aybeniz63@rambler.ru

Current scientific and theoretical problems of big data

Abstract-A very large size of digital data is generated in the world on the basis of the relation of technical and technological development in last ten years ,as a result “big data” term has occurred. Today,it has become the main subject of extensive discussion in everywhere- magazine's, newspaper's articles, blogs and so on. Big Data brings new opportunities to modern society and challenges to scientists. Big data term is viewed as a mass media, an object of scientific research opposed to the business sector in the article. A brief statement of big data concept is given, key factors that stipltuaing its development as a research field are noted. Current scientific and theoretical challenges are also analyzed.

Key words: big data, big data analytics, audio analytics, video analytics, social media analytics, vizualization, security.