

Торайғыров университетінің
ҒЫЛЫМИ ЖУРНАЛ

НАУЧНЫЙ ЖУРНАЛ
Торайғыров университета

ТОРАЙҒЫРОВ УНИВЕРСИТЕТІНІҢ ХАБАРШЫСЫ

Физика, математика және компьютерлік
ғылымдар сериясы
1997 жылдан бастап шығады



ВЕСТНИК ТОРАЙҒЫРОВ УНИВЕРСИТЕТА

Серия: Физика, математика
и компьютерные науки
Издается с 1997 года

ISSN 2959-068X

№ 1 (2023)
Павлодар

**НАУЧНЫЙ ЖУРНАЛ
ТОРАЙГЫРОВ УНИВЕРСИТЕТА**

Серия: Физика, математика и компьютерные науки
выходит 4 раза в год

СВИДЕТЕЛЬСТВО

о постановке на переучет периодического печатного издания,
информационного агентства и сетевого издания
№ KZ91VPY00046988

выдано

Министерством информации и общественного развития
Республики Казахстан

Тематическая направленность

публикация материалов в области физики, математики,
механики и информатики

Подписной индекс – 76208

<https://doi.org/10.48081/SGQS7560>

Бас редакторы – главный редактор

Тлеукенов С. К., *ф.-м.г.д., профессор*

Заместитель главного редактора Испулов Н. А., *ф.-м.г.к., профессор*

Ответственный секретарь Жумабеков А. Ж., *PhD*

Редакция алкасы – Редакционная коллегия

Esref Adali,	<i>PhD докторы, профессор (Турция);</i>
Qadir Abdul,	<i>PhD докторы, профессор (Пакистан);</i>
Донбаев К. М.,	<i>ф.-м.г.д., профессор;</i>
Демкин В. П.,	<i>ф.-м.г.д., профессор (Российская Федерация);</i>
Жумадилаева А. К.,	<i>т.г.к., кауымд. профессор;</i>
Ибраев Н. Х.,	<i>ф.-м.г.д., профессор;</i>
Косов В. Н.,	<i>ф.-м.г.д., профессор;</i>
Сентова С. М.,	<i>пед.г.д., профессор;</i>
Шоканов А. К.,	<i>ф.-м.г.к., профессор</i>
Омарова А. Р.,	<i>технический редактор</i>

За достоверность материалов и рекламы ответственность несут авторы и рекламодатели

Редакция оставляет за собой право на отклонение материалов

При использовании материалов журнала ссылка на «Вестник Торайгыров
университета» обязательна

© Торайгыров университет

СЕКЦИЯ «КОМПЬЮТЕРНЫЕ НАУКИ»

FTAMP 20.53.19

<https://doi.org/10.48081/NOGC5873>

***Ж. С. Алимова¹, А. З. Даутова², А. О. Садыкова³,
Ж. Айсаұлы⁴**

^{1,2,3,4}Торайгыров университеті, Қазақстан Республикасы, Павлодар қ.

*e-mail: jarasovajanar@mail.ru

ДЕРЕКТЕРДІ ТАЛДАУДЫҢ БАҒДАРЛАМАЛЫҚ ҚҰРАЛДАРЫНА ШОЛУ

Бұл мақалада Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің ІРН АР14972847 «Деректердегі айқын емес байланыстарды анықтау мен терең талдауды жүзеге асыратын алгоритмі мен компьютерлік бағдарламасын әзірлеу» мемлекеттік бюджеттік жобасын орындау барысында, студенттерді осы тақырыптың төңірегінде дипломдық жұмыс жазуға тарту мақсатында орындалған, 6В06101-Информатика мамандығының студенті Ж. Айсаұлының бітіру жұмысында қарастырылған, деректерді талдаудың бағдарламалық құралдарына шолу жасалады. Қазіргі таңда, деректерді цифрлық түрде жинау мен сақтау ең сенімді, арзан және тиімді болып қалғаны белгілі. Сәйкесінше, осы жинақталған үлкен деректерді жинау ғана жеткіліксіз – оны қандай да бір жолмен пайдалану керек. Ал пайдалану үшін деректерді талдау қажет екені анық. Екінші жағынан, әлем цифрлық сипатқа ие бола салысымен, сандармен бейнеленген әлемді талдауға арналған озық компьютерлік аналитика қажет екендігі де рас. Аналитика өздігінен емес, нақты үлгілерді жасау және үлкен деректер жиынтығымен сипатталған жағдайларда дұрыс шешім қабылдау үшін қажет. Осындай кең ауқымды мәселені қарастырудың алғашқы қадамы ретінде, бұл мақалада, үлкен деректер ұғымына тоқтала отырып, оның анықтамасы, деректер қоры жинақталатын интернет ресурстар, және осы жинақталған деректерді талдаудың ақпараттық құралдары қарастырылады.

Кілтті сөздер: үлкен деректер, деректерді талдау, деректер қоры, бағдарламалық құралдар, компьютерлік аналитика.

Кіріспе

Тақырыптың өзектілігі бірнеше факторлармен айқындалады:

а) Қазіргі таңда, интернет желісінің дамуына орай цифрлық әлем өмірдің нормасына айналды. Ал, деректерді цифрлық түрде жинау мен сақтау ең сенімді, арзан және тиімді болып қалды. Сәйкесінше, осы жинақталған үлкен деректерді жинау ғана жеткіліксіз – оны қандай да бір жолмен пайдалану керек. Ал пайдалану үшін деректерді талдау қажет.

ә) Әлем цифрлық сипатқа ие бола салысымен, сандармен бейнеленген әлемді талдауға арналған озық компьютерлік аналитика қажет екендігі рас. Демек, үлкен деректер және оны талдау кез келген қызмет саласы үшін өзекті болды – бизнес, экономика, өнеркәсіп, мұнай және газ өнеркәсібі, банк ісі, маркетинг, мемлекеттік және муниципалды басқару, телекоммуникация және т.б. Аналитика өздігінен емес, нақты үлгілерді жасау және үлкен деректер жиынтығымен сипатталған жағдайларда дұрыс шешім қабылдау үшін қажет.

Аталған факторларды ескере отырып, осы мақалада үлкен деректерді талдаудың ақпараттық құралдарына тоқталу көзделіп отыр. Ендеше осы жұмыста алғашқы қарастырылатыны – үлкен деректер ұғымы және дереккөздері. Екінші ретте компьютерлік аналитика құралдарына шолу жасалады.

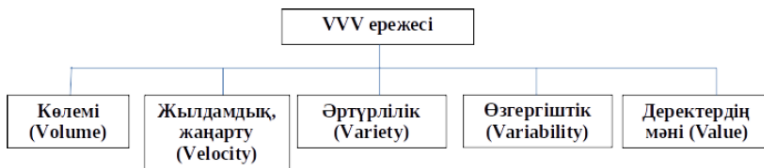
Материалдар және әдістер

Үлкен деректер (Big data) – бұл үлкен көлемді ақпараттардың жиынтығы дегенді білдіреді, және ол қандайда да бір цифрлық сақтау орындарында көбіне жүйесіз түрде сақталады. Дегенмен, үлкен деректерді әдеттегі құрылымдау және талдау құралдарымен өңдеу қиындықтар туғызады. Сондықтан, үлкен деректер ұғымын, әлсіз құрылымдалған үлкен көлемді деректерді іздеу, өңдеу және қолдану технологиялары тұрғысынан да қарастырады.

Тарихқа сүйенсек, үлкен деректер деген сөз тіркесін 2008 жылы Клиффорд Линч қолданған болатын. Ол Nature эксперт журналында жарық көрген мақаласында ақпараттық ағындардың өсуін – big data деп атаған еді. Мұнда ол, кез-келген біртекті емес деректердің күніне 150 Гб-тан артық массивін үлкен деректерге жатқызған еді.

Аналитикалық агенттіктердің статистикасына қарасақ, 2005 жылы әлем 4–5 эксабайт (4–5 миллиард гигабайт) ақпаратпен амалдар орындаса, ал одан бес жыл өткен соң оның көлемі 0,19 зеттабайтқа (1ЗБ = 1 024 ЭБ) дейін артқан. Ал, 2012 жылы – 1,8 ЗБ, 2015 жылы – 7 ЗБ, 2020 жылы – 42-45 зеттабайт ақпарат көлемімен амалдар орындауда [1].

Ақпараттар массиві үлкен деректер болуы үшін келесідей белгілері болуы керек [2] (1-сурет):



Сурет 1 – VVV ережесі

Көлемі – деректер физикалық шамасы және цифрлық тасымалдаушыда алатын кеңістігіне байланысты өлшенеді. Үлкен деректерге күніне 150 Гб-тан жоғары массивтер жатады.

Жылдамдық, жаңарту – ақпарат үнемі жаңарып отырады, және оларды өңдеу үшін үлкен деректердің интеллектуалды технологиялары қажетті.

Әртүрлілік – массивтегі ақпарат біртекті емес форматта, жүйсіз түрде және құрылымдалмаған күйде жинақталады. Мысалы, әлеуметтік желілер үлкен деректерді мәтіндер, видео, қаржылық транзакциялар, суреттер және басқалар түрінде қолданады.

Өзгергіштік – деректер ағындарының периодтары, маусымдылығы, өсу және түсулері болуы мүмкін. Әлсіз құрылымдалған ақпараттардағы өзгерістерді басқару өте қиын, және оларды өңдеуге қуатты технологиялар қажетті.

Деректердің мәні – ақпараттарды қабылдау және өңдеу үшін әртүрлі қиындықтар болуы мүмкін, ол өз кезегінде интеллектуалды жүйелердің жұмысына кедергі келтіреді. Машиналардың негізгі міндеті, осы келіп түскен ақпараттардың маңыздысын таңдап алып, оны тез арада құрылымдау.

Үлкен деректер технологиясының жұмыс принципі, қолданушыларды қандай-да бір зат немесе құбылыс туралы көбірек ақпараттандыруға негізделген. Мұндай ақпараттандыру немесе деректермен таныстырудың негізгі міндеті – барлық «қолдау» немесе «қарсы» жақтарын білуге көмектесу [2]. Жалпы алғанда, үлкен деректердің дереккөздеріне мыналарды жатқызуға болады:

- а) интернет – блогтар, әлеуметтік желілер, сайттар, ЖАҚ (СМИ) және әртүрлі форумдар;
- ә) корпоративтік ақпараттар – архивтер, транзакциялар, деректер қоры;
- б) есептеу құрылғыларының көрсеткіштері – метеорологиялық құралдар, ұялы байланыс датчиктері және басқалар.

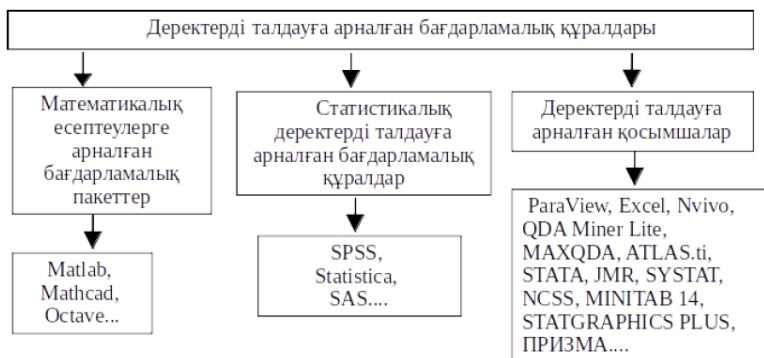
Аталған дереккөздердің ішінде, интернет желісінен алынатын, әлемдік деңгейдегі ақпараттар ағыны жинақталатын деректер қорларының мысалы ретінде мыналарды келтіруге болады (2-кесте):

Кесте 2 – Әлемдік деңгейдегі ақпараттар ағыны жинақталатын деректер қорлары

World Bank Open Data https://data.worldbank.org/	Бұл деректер қорында әлемнің дамуының 570-ке жуық көрсеткіштері бойынша статистикалық деректер жинақталған. Уақыт қатарлары 1960 жылдан бастап 208 ел бойынша ұсынылған. Экономикалық, әлеуметтік, қаржылық көрсеткіштер, қоршаған орта мен табиғи ресурстар қамтылған [3]
OECD.Stat https://stats.oecd.org/	Мұнда экономикалық ынтымақтастық және даму ұйымының мүшелері мен қоса, бұл ұйымға кірмейтін жекелеген елдердің де экономикалық, қаржылық, әлеуметтік, ғылыми-техникалық, салалық көрсеткіштері туралы деректер жинақталған [4].
Eurostat https://ec.europa.eu/eurostat	Бұл ашық деректер қорында Еуропалық одақ (ЕО) мүшелері туралы ақпарат жинақталған. Деректер қорында жалпы және аймақтық статистика, экономикалық және қаржы көрсеткіштері, демографиялық және әлеуметтік шарттар, өнеркәсіптер мен сауда туралы және басқа да көптеген деректер бар [5]
E u r o m o n i t o r International https://www.euromonitor.com/	Деректер қоры елдер, индустриялар, компаниялар және тұтынушылар бойынша деректерді қамтиды. Мұнда бизнес-ортаны, салалық көрсеткіштерді, нарықтың брендтер мен компаниялар бойынша үлесін, әлемнің ірі экономикалық салаларын және компаниялардың арасындағы өзара қатынастарды (B2B) талдау үшін қажетті ақпараттарды алуға болады [6]

Statista https://www.statista.com/	600 салалар мен 150 ел бойынша ерекше статистика мен есептер жиналған деректер қоры. Мысалы, World of Warcraft ойнайтын ойыншылардың санын, немесе индияның қанша тұрғыны смартфон қолданады деген сияқты сұрақтарға жауап табуға болады [7]
International Monetary Fund Data https://www.imf.org/en/Data	Жалпы әлем бойынша және жекелеген елдер бойынша экономикалық, қаржы көрсеткіштері және валюта бағамы туралы уақыт қатарлары түрінде деректер сақталған [8]
OPEC Data/Graphs https://www.opec.org/opec_web/en/20.htm	ОПЕК мұнай саласының деректері (бағасы, салықтар, мұнай қоры, өндіру және сату туралы ақпараттар) жинақталған [9]
WTO Statistics https://www.wto.org/index.htm	<u>Бұл деректер қорында әлемнің елдері бойынша сауда ағыны, тарифтар, тарифтік емес шаралар және үстеме бағамен сауда үлесі туралы деректер бар [10]</u>
egov «Ашық деректер» порталы https://data.egov.kz/	Ашық үкімет компоненттерінің бірі, мүдделі азаматтарды Қазақстан Республикасының мемлекеттік органдарынан шығатын әртүрлі деректер жиынтығына толық қол жеткізумен қамтамасыз ету мақсатында құрылған [11]
Қазақстан ұлттық банкі https://www.nationalbank.kz/ru	Бұл деректер қорында ел экономикасының ақша-кредит саласын, ел резиденттерінің бейрезиденттермен қарым-қатынасын, экономиканың қаржы секторының, бағалы қағаздар нарығының және ұлттық төлем жүйесінің ахуалын сипаттайтын Қазақстан Ұлттық Банкінің статистикалық деректері ұсынылған [12]
Дүние жүзілік денсаулық сақтау ұйымы еуропалық аймақтық кеңсесінің деректер қорында https://www.euro.who.int/en/home	Денсаулық сақтау саласындағы, мониторинг құралдары, және осы саладағы іс-әрекеттердің негізгі бағыттарын бағалау бойынша статистикалық деректер жинақталады. Бұл қорда халықаралық жиынтық көрсеткіштерді салыстырудан бастап, осы денсаулық сақтау саласындағы арнайы іс-әрекеттермен айналысатын бағыттарына дейін және ауруды мұқият бақылау нәтижелеріне дейін деректерді табуға болады [13].
EM-DAT https://www.emdat.be/	Табиғи және техногендік апаттардың алдын алу. EM-DAT 1900 жылдан бүгінгі күнге дейін әлемде 22 000-нан астам жаппай апаттардың пайда болуы мен салдары туралы маңызды негізгі деректерді қамтиды. Мәліметтер қоры әртүрлі көздерден, соның ішінде БҰҰ агенттіктерінен, үкіметтік емес ұйымдардан, сақтандыру компанияларынан, ғылыми-зерттеу институттарынан және баспасөз агенттіктерінен жинақталған [14].

Бұл аталғандар әлемдік деңгейдегі ақпараттар ағыны жинақталатын деректер қорларының бірнеше мысалы ғана. Демек, деректерді цифрлық түрде жинау мен сақтау ең сенімді, арзан және тиімді болып отыр. Сәйкесінше, осы жинақталған үлкен деректерді жинау ғана жеткіліксіз – оны қандай да бір жолмен пайдалану керек. Ал пайдалану үшін деректерді талдау қажет екені анық. Екінші жағынан, әлем цифрлық сипатқа ие бола салысымен, сандармен бейнеленген әлемді талдауға арналған озық компьютерлік аналитика қажет екендігі де рас. Компьютерлік аналитика құралдары ретінде математикалық есептеулерге арналған, сондай-ақ, статистикалық деректерді талдауға арналған бағдарламалық құралдар мен қосымшаларды айтуға болады (2-сурет).



Сурет 2 – Деректерді талдауға арналған бағдарламалық құралдары

Математикалық есептеулерге арналған арнайы бағдарламалар Matlab, Mathcad және Octave бағдарламалық пакеттері болып табылады:

1) Matlab – автоматтандырылған математикалық есептеулер жүйесі. Математикалық есептеулер саласындағы барлық әдістерді қамтиды (<https://ww2.mathworks.cn/en/products/matlab-online.html>).

2) Mathcad – бұл ғылым мен техниканың, білім берудің әр түрлі облыстарында массалы математикалық есептерді шешуді автоматизациялауға бағытталған компьютерлік математиканың белгілі жүйесі.

3) Octave (<https://octave-online.net/>) – Matlab-пен үйлесімді жоғары деңгейлі тілді пайдалана отырып, математикалық есептеулерге арналған тегін бағдарламалық қамтамасыз ету жүйесі.

Сондай-ақ, статистикалық деректерді талдауға арналған кең таралған бағдарламалық құралдар ретінде SPSS, Statistica, SAS сияқты бағдарламаларды айтуға болады [15]:

4) SPSS – әлеуметтік ғылымдар саласындағы статистикалық талдауларда кеңінен қолданылатын бағдарлама. Бұл сондай-ақ нарықты зерттеушілер, денсаулық сақтау саласындағы зерттеушілер, зерттеу компаниялары, үкімет, білім саласындағы зерттеушілер, маркетингтік ұйымдар, деректер өндірушілер [16] және басқаларға да қолданылады.

5) Statistica (<http://statsoft.ru/>) – статистикалық және графикалық талдау, болжау, data mining, жеке қосымшаларды жасау, оларды кіріктіру, және басқалар үшін қуатты әрі қолдануға ыңғайлы құралдарды ұсынады. Ол туралы барлық ақпараттар, оқыту құралдары мен бағдарламаны жүктеу көздерін Statsoft компаниясының сайтынан табуға болады.

6) SAS – қаржы менеджменті, тауекелдерді басқару, маркетинг, жеткізу тізбектерін басқару үшін негізгі Business Intelligence өнімін ұсынады. Есептеулерде салаларға байланысты ерекшеліктерді ескере отырып, әртүрлі салалар үшін сол салаға сәйкес шешу жолын таңдауға болады

Бұлардан басқа да деректерді талдауға арналған қосымшалар көптеп кездеседі:

7) ParaView – деректерді визуализациялауға және талдауға арналған қосымша – <https://www.paraview.org/>. Бұл қосымшада электрондық кесте және мәтінді өңдеу қолданбалары оқшау бағдарламалар ретінде де және пакеттің бөлігі ретінде де қол жетімді. Компьютерде әріптерді, құжаттарды құру, сандарды қысқарту және қаржылық болжамдар жасау сияқты бірқатар тапсырмаларды жылдам орындауға мүмкіндік береді. Электрондық кесте және мәтінді өңдеу бағдарламалары ұқсас, бірақ электрондық кесте қолданбасы сандық деректерге арналған, ал мәтінді өңдеу қолданбасы құжаттарды құруға бағытталған.

Бұл бағдарлама – таратылған жадты есептеу ресурстарын қолдана отырып, өте үлкен мәліметтер жиынтығын талдау үшін жасалған. Қосымшаны суперкомпьютерлерде мәліметтер жиынтығын талдауға, сондай-ақ кішігірім мәліметтер жиынтығына арналған ноутбуктарда пайдалануға болады.

8) Excel арқылы деректерді талдау. Excel электрондық кестелері есептеу және деректерді басқару процесін автоматтандыруға мүмкіндік береді, ақпаратты талдау нәтижелерін графикалық түрде ұсынуға болады.

9) NVivo – көптеген көздерден алынған деректерді сапалы ұйымдастыруға және оңай басқаруға мүмкіндік беретін қосымша (<https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/try-nvivo>). Деректерді тікелей Word және Excel бағдарламаларынан, сондай-ақ Twitter, Facebook, YouTube және SurveyMonkey сияқты құралдардан әлеуметтік медиа мен сауалнама деректерін импорттау үшін кіріктірілген мүмкіндіктері бар. Кез-келген дереккөзден зерттеу деректерін импорттауға және олармен жұмыс істеуге мүмкіндік береді, соның ішінде: сауалнамалар, сұхбаттар, мақалалар, бейне,

электрондық пошта, әлеуметтік медиа және веб-мазмұн, толық немесе қарапайым мәтін, PDF, аудио, цифрлық фотосуреттер, электрондық кестелер.

10) QDA Miner Lite – жоғары сапалы зерттеулерді жүргізуге арналған тегін бағдарламалық қамтамасыз ету пакеті (толық ақпаратты <https://guides.library.wheaton.edu/QualResearch/QDAMiner> сайтынан табуға болады). Оны сұхбаттар, ашық жауаптар, транскрипттер, жазбалар және т.б. сияқты мәтіндік деректерді талдау үшін пайдалануға болады. Сонымен қатар,

а) қимылсыз бейнелерді талдау;

б) Word, HTML, PDF, кәдімгі мәтін, CSV және Excel сияқты әртүрлі форматтағы құжаттарды импорттау;

в) Ағаш құрылымында ұйымдастырылған кодтарды пайдалана отырып, интуитивті кодтау;

г) Кодталған сегменттерге түсініктемелер немесе ескертпелер қосу;

д) Мәтін сегменттерін шығарып алуға және кодтауға арналған мәтінді іздеу құралы;

е) Кодтау жиілігін талдау;

ж) Кестелерді Excel, Word немесе CSV файлдарына экспорттау.

11) MaxQDA – сапалы және аралас әдістермен – ұйымдастыру, талдау, визуализация, презентация әдістерімен – деректерді талдауға арналған әлемдік жетекші бағдарламалық құрал (<https://www.maxqda.com/> – арнайы сайтында). Деректердің барлық түрлерін талдауға болады: мәтіндерді, суреттерді, аудио/бейне файлдарды, веб-сайттарды, твиттер, фокус-топтық талқылауларды, сауалнама жауаптарын және т.б. талдауға мүмкіндік береді.

12) Atlas.ti – интуитивті зерттеу құралдары мен озық технологияның көмегімен іске асырылатын түсініктерді ашуға көмектеседі – <https://atlasti.com/>. Кез келген қажеттілік үшін жасалған: негізгі талдау тапсырмаларынан бастап ең терең зерттеу жобаларына дейін.

13) Stata – деректер туралы ғылымның барлық қажеттіліктерін – деректерді өңдеу, визуализация, статистика және автоматтандырылған есеп беруді қамтамасыз ететін толық, біріктірілген бағдарламалық құрал пакеті (<https://www.stata.com/>).

14) JMR – деректерді талдау қосымшасы.

15) Systat – дербес компьютерлерге арналған статистикалық жүйе.

16) NCSS – статистикалық өңдеу саласындағы кәсіби емес мамандарға арналған қосымшасы. Жүйе интерфейсі көп терезелі.

17) MiniTab 14 – бұл пайдаланушы интерфейсін жақсы ұйымдастырылған және жұмыс нәтижелерін визуализациялау үшін жақсы мүмкіндіктері бар, қолдануда өте оңай бағдарламалық пакет.

18) StatGraphicsPlus – өте күшті статистикалық бағдарлама. 250-ден астам статистикалық функцияларды қамтиды, анық, теңшелетін есептерді жасайды.

19) Призма – интуитивті интерфейс бірнеше минут ішінде деректерді талдауға және жоғары сапалы графиктер құруға мүмкіндік береді. Бағдарламада көптеген зерттеулер үшін жеткілікті болатын негізгі жиі қолданылатын статистикалық функциялар бар.

Қарастырылған барлық бағдарламалық құралдар деректердің көптеген көздерінен әртүрлі форматтағы ақпараттарды жүктеуге, және осы деректер массивіне математикалық өңдеу әдістерінің көптеген түрлерін қолдана отырып, талдаудың – ұйымдастыру, талдау, визуализация, презентация әдістерімен нақты зерттеу деректерін талдауға мүмкіндік береді.

Нәтижелер және талқылау

Деректерді талдау бойынша зерттеу жұмыстарын бастамас бұрын, кез-келген тұтынушы қолдана алатын ашық арнайы дереккөздерін және сол деректер қорындағы ұсынылатын қосымша талдау құралдарын қарастырып алған жөн. Егер зерттеу нәтижелерін қорыту үшін аталмыш деректер қорында ұсынылған мүмкіндіктер жеткіліксіз болған жағдайда, компьютерлік аналитика құралдарына жүгіну қажет. Осы мақалада, деректерді жинақтауға арналған дереккөздер мен осы деректерді талдауға арналған бағдарламалық құралдарға шолу жасалды.

Бұл нәтижелер Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің ИРН АР14972847 «Деректердегі айқын емес байланыстарды анықтау мен терең талдауды жүзеге асыратын алгоритмі мен компьютерлік бағдарламасын әзірлеу» мемлекеттік бюджеттік жобасын орындау барысында, студенттерді осы тақырыптың төңірегінде дипломдық жұмыс жазуға тарту мақсатында, В06101-Информатика мамандығының студенті Ж.Айсаұлының бітіру жұмысында қарастырылды.

Қорытындылар

Сонымен, жоғарыда айтылғандарды қорыта келе, деректерге талдау жасауға арналған бағдарламалық құралдар деректердің көптеген көздерінен әртүрлі форматтағы ақпараттарды жүктеуге, және осы деректер массивіне математикалық өңдеу әдістерінің көптеген түрлерін қолдана отырып, талдаудың – ұйымдастыру, талдау, визуализация, презентация әдістерімен нақты зерттеу деректерін талдауға мүмкіндік беретіні анықталды. Яғни, математикалық және статистикалық талдау кезінде сипаттамалық нәтижелерді алуға толық мүмкіндік береді. Дегенмен, деректерді терең талдауы сұрақтарына келетін болсақ, бұл бағдарламалық құралдардың қауқарсыз екенін байқауға болады. Демек, тақырыпты осы бағытта зерттеуге негіз бар деген сөз.

Пайдаланған деректер тізімі

- 1 **Smolan, R., Erwit, J.** «The Human Face of Big Data» [Текст] / Publisher : Against All Odds Productions, 2012. – p. 224
- 2 **Pang Ning Tan V., Steinbach M.** «Introduction to Data Mining»-2nd-Edition. [Текст] / Publisher : AK.R, 2015.– p.152
- 3 Дүниежүзілік ашық деректер қоры / <https://data.worldbank.org>
- 4 OECD.Stat ашық деректер қоры [Электрондық ресурс]. – / [Электрондық ресурс]. – <https://stats.oecd.org/>
- 5 Eurostat ашық деректер қоры / [Электрондық ресурс]. – <https://ec.europa.eu/eurostat>.
- 6 Euromonitor International / [Электрондық ресурс]. – <https://www.euromonitor.com/>.
- 7 Statista ашық деректер қоры / [Электрондық ресурс]. – <https://www.statista.com/>
- 8 International Monetary Fund Data / [Электрондық ресурс]. – <https://www.imf.org/en/Data>
- 9 OPEC Data / Graphs / [Электрондық ресурс]. – https://www.opec.org/opec_web/en/20.htm
- 10 WTO Statistics / [Электрондық ресурс]. – <https://www.wto.org/index.htm>
- 11 Egov «Ашық деректер» порталы / [Электрондық ресурс]. – <https://data.egov.kz/>
- 12 Қазақстан ұлттық банкі ашық деректер қоры / [Электрондық ресурс]. – <https://www.nationalbank.kz/ru>.
- 13 Дүниежүзілік денсаулық сақтау ұйымы еуропалық аймақтық кеңсесінің деректер қоры / [Электрондық ресурс]. – <https://www.euro.who.int/en/home>
- 14 EM-DAT / [Электрондық ресурс]. – <https://www.emdat.be/>
- 15 **Кричевец, А. Н., Корнеев, А. А., Рассказова, Е. И.** Основы статистики для психологов. [Текст] / М. : Акрополь, 2019. – 286 с.
- 16 **Гунарто, Х.** Әлеуметтік зерттеулерге арналған параметрлік және параметрлік емес деректерді талдау: IBM SPSS. [Текст] LAP академиялық баспа. ISBN 978-6200118721. – 2019.

REFERENCES

- 1 **Smolan, R., Erwit, J.** «The Human Face of Big Data» [Text] / Publisher: Against All Odds Productions, 2012. – p. 224.
- 2 **Pang Ning Tan V., Steinbach M.** «Introduction to Data Mining»-2nd-Edition. [Text] / Publisher: AK.R, 2015.– p. 152.

3 Dыniezhыzilik ashy`k derekter qory` [World Open Database] / [Electronic resource]. – <https://data.worldbank.org>

4 OECD.Stat / [Electronic resource]. – <https://stats.oecd.org/>

5 Eurostat ashy`k derekter qory` [Eurostat open database] / [Electronic resource]. – <https://ec.europa.eu/eurostat>.

6 Euromonitor International / [Electronic resource]. – <https://www.euromonitor.com/>.

7 Statista / [Electronic resource]. – <https://www.statista.com/>

8 International Monetary Fund Data / [Electronic resource]. – <https://www.imf.org/en/Data>

9 OPEC Data / Graphs / [Electronic resource]. – https://www.opec.org/opec_web/en/20.htm

10 WTO Statistics / [Electronic resource]. – <https://www.wto.org/index.htm>

11 Egov «Ashy`k derekter» portaly` [Egov «Open Data» portal] / [Electronic resource]. – <https://data.egov.kz/>

12 Qazaqstan ылty`k banki ashy`k derekter qory` [Open database of the National Bank of Kazakhstan] / [Electronic resource]. – <https://www.nationalbank.kz/ru>.

13 Dыniezhыzilik densauly`k saқтаu ұйы`my` europaly`k аймақты`k кеңsesiniң derekter qory` [World Health Organization Regional Office for Europe database] / [Electronic resource]. – <https://www.euro.who.int/en/home>

14 EM-DAT / [Electronic resource]. – <https://www.emdat.be/>

15 **Krichevecz, A. N., Korneev, A. A., Rasskazova, E. I.** Osnovy` statistiki dlya psixologov. [Fundamentals of Statistics for Psychologists] [Text] / M. : Akropol`, 2019. – 286 p.

16 Gunarto, X. Әлеуметтік зерттеулерге арналған параметрлік және параметрлік емес деректерді талдау : IBM SPSS. [Parametric and Nonparametric Data Analysis for Social Research : IBM SPSS.] [Text] LAP akademiyaly`k baspa. ISBN 978-6200118721. – 2019.

Материал 06.03.23 баспаға түсті.

***Ж. С. Алимова¹, А. З. Даутова², А. О. Садыкова³, Ж. Айсаулы⁴**

^{1,2,3,4}Торайғыров университет, Республика казахстан, г. Павлодар

Материал поступил в редакцию 06.03.23.

ОБЗОР ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ АНАЛИЗА ДАННЫХ

В данной статье проведена обзор программных средств анализа данных, рассмотренных в дипломной работе студента специальности 6В06101-Информатика Ж.Айсаулы, выполненная в ходе выполнения государственного бюджетного проекта Ж. С. Алимовой ИРН AP14972847 Комитета науки Министерства науки и высшего образования Республики Казахстан «Разработка алгоритма и компьютерной программы для обнаружения и углубленного анализа неявных взаимосвязей данных», с целью привлечения студентов к написанию дипломной работы по данной теме. В настоящее время известно, что сбор и хранение данных в цифровом виде остается самым надежным, дешевым и эффективным. Соответственно, мало собрать эти накопленные данные – их нужно как-то использовать. И понятно, что анализ данных необходим для использования. С другой стороны, также верно и то, что как только мир станет цифровым, возникнет потребность в передовой компьютерной аналитике для анализа мира, представленного числами. Аналитика нужна не сама по себе, а для создания точных моделей и принятия правильных решений в ситуациях, описываемых большими наборами данных. В качестве первого шага к рассмотрению столь масштабной проблемы в данной статье основное внимание уделяется понятию больших данных, их определению, интернет-ресурсам, на которых собирается база данных, и информационным инструментам для анализа этих собранных данных.

Ключевые слова: большие данные, анализ данных, базы данных, программные средства, компьютерная аналитика.

***Zh. S. Alimova¹, A. Z. Dautova², A. O. Sadykova³, Zh. Aisauly⁴**

^{1,2,3,4}Toraighyrov University, Republic of Kazakhstan, Pavlodar

Material received on 06.03.23.

OVERVIEW OF DATA ANALYSIS SOFTWARE

In this article, IRN AP14972847 of the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan, during the implementation of the state budget project «Development of an algorithm and a computer program for the identification and in-depth analysis of ambiguous relationships in data», was carried out in order to attract students to write a thesis on this topic, 6B06101-Informatics A review of software tools for data analysis, considered in the graduation work of Zh. Aisauly, a student of the specialty. Nowadays, it is known that the collection and storage of data digitally remains the most reliable, cheap and efficient. Accordingly, it is not enough to collect this accumulated big data - it must be used in some way. And it is clear that data analysis is necessary for use. On the other hand, it is also true that as soon as the world becomes digital, there will be a need for advanced computer analytics to analyze the world represented by numbers. Analytics is needed not by itself, but to create accurate models and make the right decisions in situations described by large data sets. As a first step to consider such a wide-ranging problem, this article focuses on the concept of big data, its definition, Internet resources where the database is collected, and informational tools for analyzing this collected data.

Keywords: big data, data analysis, databases, software tools, computer analytics.

Теруге 06.03.2023 ж. жіберілді. Басуға 30.03.2023 ж. қол қойылды.

Электрондық баспа

7,50 Mb RAM

Шартты баспа табағы 9,1. Таралымы 300 дана. Бағасы келісім бойынша.

Компьютерде беттеген: Е. Е. Калихан

Корректор: А. Р. Омарова, Д. А. Кожас

Тапсырыс № 4020

Сдано в набор 06.03.2023 г. Подписано в печать 30.03.2023 г.

Электронное издание

7,50 Mb RAM

Усл.печ.л. 8,4. Тираж 300 экз. Цена договорная.

Компьютерная верстка Е. Е. Калихан

Корректор: А. Р. Омарова, Д. А. Кожас

Заказ № 4020

«Toraighyrov University» баспасынан басылып шығарылған

«Торайғыров университеті» КЕ АҚ

140008, Павлодар қ., Ломов к., 64, 137 каб.

«Toraighyrov University» баспасы

«Торайғыров университеті» КЕ АҚ

140008, Павлодар қ., Ломов к., 64, 137 каб.

+7(718)267-36-69

e-mail: kereku@tou.edu.kz

www.vestnik.tou.edu.kz

<https://vestnik-pm.tou.edu.kz/>