

BIG DATA (ГОЛЕМИ ДАННИ) В ОБРАЗОВАНИЕТО

Проф. д.п.н. Наталия Витанова

Шуменски университет „Епископ Константин Преславски“

Резюме. В статията е направен опит да се опишат възможностите на приложение на Big Data технологията в сферата на образованието. Представени са множество дефиниции, принципите за работа с Big Data, някои от вероятните приложения на технологията в процеса на обучение, както и вероятните бариери за нейното използване в образованието.

Ключови думи: Big Data; образование; учебен процес; студенти

Феноменът глобализация, икономическата интернационализация, интензивното развитие на информационните и комуникационните технологии (ИКТ) са основите на съвременното общество на знанието. Тези явления предполагат постоянна промяна в социалната сфера и непрекъснатата трансформация на начина на живот, на работа и на учене.

Дълго време съвременното общество се уповава/ослани на технологичните изменения. Сократ е на мнение, че писмеността създава забора в душата на ученика, и поради тази причина той не записва собствените си думи (известен факт е, че Сократ няма писмени съчинения). В някакъв смисъл може би е прав. Ако се сравнят нашите способности за запомняне с традициите в устното народно творчество (някои хора са рецитирали наизуст произведения като „Илиада“), съвременната култура може да се окаже абсолютно лишена от памет. Ако Сократ живееше в съвременния свят, вероятно щеше да е „потресен“ колко малко се помни и колко много се разчита на помощни средства извън собствените глави.

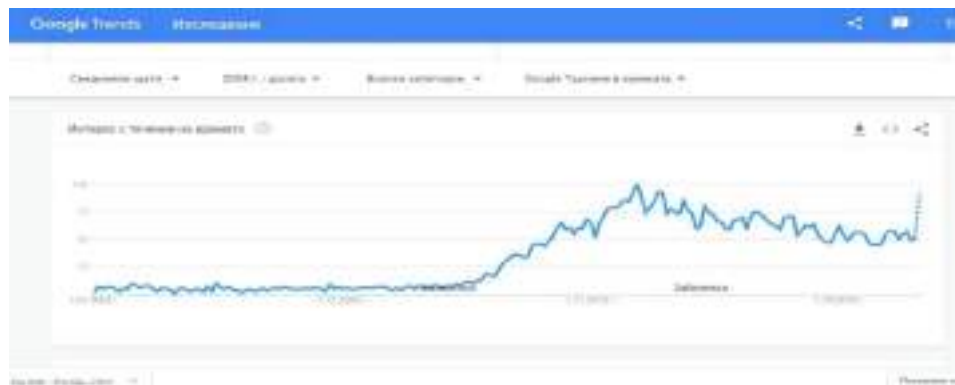
В постмодерния свят „все повече данни се генерират от световната интернет мрежа и се произвеждат от електронни сензори и устройства, които са навсякъде около нас. Обемът на данните, тяхното разнообразие и високата скорост, с която се създават, водят до създаване на концепцията „големи данни“, или Big Data. „Големите данни“ са съвършено нови източници за официалната статистика, с характеристики, различни от тези на традиционните източници. Промяната в характера на данните, тяхната наличност, в процеса по събиране и разпространение е фундаментална. Тази промяна, по същество, е промяна на парадигмата на конвенционалното статистическо изследване (Stateva 2017). „Големите данни“ са феномен, който се определя като „следващото голямо нещо в иновациите“

(Gobble 2013), „четвъртата парадигма на науката“ (Strawn 2012, p. 34), „следващата граница за иновации, конкуренция и производителност“ (Manyika 2011, p. 1), „следващата „управленска революция““ (McAfee & Brynjolfsson 2012, p. 3), „революция в науката и технологиите“ (Ann Keller, Koonin & Shipp 2012, p. 4) и др.

Бързото развитие на технологиите изисква непрекъснато адаптиране на образованието, за да бъде то в крак с технологичните изменения. В този контекст технологиите и образованието участват в своеобразна гонка (Goldin, Katz 2009). Когато образованието изостава от технологичния прогрес, то не подготвя достатъчно квалифицирани хора, които да са продуктивни и да работят ефективно. Как да се преодолее това изоставане и да се научат хората да живеят в епохата на безпрецедентни трансформации, когато обичайните професии изчезват след година и се появяват нови индустрии? Може би отговорът се крие в „четири-измерния модел“ на образование на XXI век, който предполага многостранност на човешкото развитие¹. „Четиримерното образование“ предлага ясен и ефективен, първи по рода си организационен модел на компетентностите, необходим за XXI век. Основната иновация не е в представянето на стандартен списък какво хората трябва да учат, а в създаването на определено пространство, в което учители, специалисти по образователни програми, методисти и ученици ще могат сами да вземат решение какво да учат, в собствен контекст за собственото бъдеще¹.

Количеството данни в съвременния свят „избухва“. Увеличаващият се обем и детайлност на информацията, нарастването на социалните медии, мултимедиите и „интернет на нещата“ ще подхранват експоненциалния растеж на данните в обозримо бъдеще. Бързите темпове, с които се развива четвъртата индустриална революция, изискват преразглеждане на традиционния образователен императив, което предполага преосмисляне на понятия като „ефективно обучение“, „предметна учебна система“, както и образователния модел, основаващ се на трансфера на знанията от учителя към ученика. Непрекъснато нарастващите групи на децата със специални образователни потребности, на децата с развито мислене, на лидери, на децата със „златни ръце“, артистично надарените деца и децата със спортни заложби инициират развитието на образователна парадигма, чийто фундамент са технологии, които могат да правят информирани заключения за необходимостта от промяна на подходите и действията в образованието (Utěmov, Gorev 2018, p. 450). Една от възможностите е технологията за работа с големи данни (Big Data). Използването на Big Data в образованието е свързано с анализи на образователната система, измерване, събиране, анализ и представяне на структурирани и неструктурирани данни от огромни обеми за ученици и образователна среда, за да се разберат особеностите на функционирането и развитието на образователната система.

Терминът Big Data се появи сравнително скоро. Google Trends показва началото на активния ръст на използване на словосъчетанието, който датира от 2011 година.



Фигура 1. Ръст на използване на термина Big Data

Понятието добива особена популярност през последните няколко години. Big Data предизвиква огромно внимание в световен мащаб. Резултатите от търсене в Google по темата показват 8 070 000 000 резултата за 0,58 секунди (24.05.2022). Това явление е свързано главно с широкото разпространение и използване на мобилни устройства, социални медийни платформи, включително YouTube, Facebook и Twitter и концепциите, свързани с „интернет на нещата“.

В хода на своята популяризация терминът се определя по различен начин:

- Big Data са данни, по-големи от 100 GB (500GB, 1TB);
- Big Data са данни, които не могат да бъдат обработвани в Excel;
- Big Data са данни, които не могат да бъдат обработени на един компютър;
- Big Data са всякакви данни;
- Big Data не съществуват, те са измислени от маркетинголозите;
- Big Data са серия от подходи, инструменти и методи за обработка на

структурирани и неструктурирани данни с огромни обеми и значително разнообразие за получаване на възприеманите от човека резултати, които са ефективни в условията на непрекъснат растеж, разпределение върху многобройните възли на компютърна мрежа, формирани в края на 2000 г. като алтернатива на традиционните системи за управление на бази данни и решения от класа Business Intelligence;

– Big Data е съвкупност от големи обеми данни (структурирани и неструктурирани), които се извличат от различни източници, имат различни форми и с тях се правят анализи на момента;

– Big Data е масив от данни, чийто размер надхвърля възможностите на типичните бази данни за съхранение, управление и анализ на данните (Manyika et al. 2011, p. 1);

– Big Data обединяват техники и технологии, които извличат смисъл от данните на екстремалната граница на практичността;

– Big Data са много голям набор от данни, които могат да бъдат анализирани с помощта на компютър, за да се откриват шаблони, тенденции и асоциации, свързани най-вече с човешкото поведение и взаимодействие (Neamtu et al. 2018);

– Big Data са данни, идващи от различни канали, включително сензори, сателити, емисии в социалните медии, снимки, видео и мобилни телефони и GPS сигнали (Rich 2012);

– Big Data са данни, заснети от сензори, публикации в сайтове за социални медии, цифрови снимки и видеоклипове, записи за транзакции и GPS сигнали за мобилен телефон и т.н.²;

– Big Data са изключително големи набори от данни, свързани с поведението на потребителите, публикации в социалните мрежи, геолокации, сензорни изходи (Johnson 2012, p. 21);

– Big Data са данни от всичко, включително данни за потока кликания от мрежата до геномни и протеомични данни от биологични изследвания и медицина (Davenport et al. 2012);

– Big Data са описание на обемно количество неструктурирани и полуструктурирани данни създавани от компания, или данни, които отнемат твърде много време и струват много пари за зареждане в релационна база данни за анализ (Rouse 2011);

– Big Data данни са данни, които не могат да се обработват пряко (Fisher, DeLine, Czerwinski, Drucker 2012, p. 53);

– Big Data са данни, които не могат да се заредят в работната памет на персоналния компютър (Havens, Bezdek, Leckie, Hall, Palaniswami 2012, p. 1130);

– Big Data са твърде големи данни, за да бъдат поставени в релационна база данни и анализирани с помощта на пакет за статистика/пакет за визуализация – данни, чийто анализ изисква масов успореден софтуер, работещ на десетки, стотици или дори хиляди сървъри (Jacobs 2009, p. 44);

– Big Data имат три основни характеристики: самите данни, анализа на данни и представяне на резултатите от анализите. След това са продуктите и услуги, които могат да бъдат обвити около един или всички тези елементи на големи данни;

– Big Data са културно, технологично и научно явление, което се опира на взаимодействието на: (1) технология: увеличаване на изчислителната мощност и алгоритмична точност за събиране, анализиране, свързване и сравняване на големи масиви от данни; (2) анализ: анализиране на големи масиви от данни за идентифициране на модели, чрез които да се отправят икономически, социални, технически и правни искове; (3) митология: широко разпространеното убеждение, че големите масиви от данни предлагат по-висока форма на интелигентност и знания, които могат да генерират прозрения, които преди са били невъзможни, с аурата на истина, обективност и точност (Boyd, Crawford 2012, p. 663).

Представените дефиниции поставят акцент върху различни аспекти на съдържанието на термина. Някои автори – B. D. Johnson, Thomas H. Davenport et al., се

фокусира върху разнообразието на източници на данни, други автори – Havens, Jacobs, Manyika et al., Rouse, наблягат на изискванията за съхранение и анализ на данните. IDC идентифицира три основни характеристики на „големите данни“: самите данни, анализирането на данните и представяне на резултатите от аналитика, които позволяват създаване на бизнес стойност по отношение на нови продукти или услуги. Boyd & Crawford предлагат по-цялостна дефиниция на „големи данни“, която включва технология (съхранение, изчислителна мощност), анализ (идентифициране на модели за икономически, социални, технически и правни претенции) и митология (убеждение, че „големите данни“ предлагат по-високо ниво на генериране на ценни данни). За „големите данни“ трябва да се мисли не само от гледна точка на анализа, но и от гледна точка на развитието на умения, които позволяват използването на ново поколение информационни технологии, инструменти и архитектури за събиране на данни от различни източници, за съхранение, за организиране, за извличане, за генериране на ценни прозрения и споделянето им със заинтересовани страни, за създаване и реализиране на конкурентни предимства. Следователно анализът на „големите данни“ може да се дефинира като холистичен подход за управление, обработка и анализ на 5V измеренията (т.е. обем, скорост, вариативност/разнообразие, достоверност и стойност) (1V (volume/обем) – количеството физически данни е значително; 2V (velocity/скорост) – скоростта на събиране на данни и скоростта на обработка на резултатите е сравнително висока; 3V (variety/вариативност) – вариативност/променливост на алгоритмите за обработка на различни видове събрани резултати; 4V (veracity/достоверност) – висока достоверност/надеждност на събраните данни, позволяваща формулирането на репрезентативни резултати; 5V (value/стойност) – стойността на натрупаните данни дава възможността да се формулират полезни разноаспектни зависимости на образователната система – б.м.), за да се създаде полезна представа за осигуряване на устойчива стойност, измерване на ефективността/производителността и установяване на конкурентни предимства (Wamba et al. 2015).

От представените дефиниции може да се обобщи, че с този термин се обозначават масиви от данни, които са толкова огромни и комплексни, че трудно могат да се обработват с традиционни приложения, като предизвикателствата са на всички нива – анализ, търсене, споделяне, съхраняване, трансфер, визуализация, поверителност и т.н.

Могат да се формулират следните основните принципи за работа с Big Data.

1. **Хоризонтална мащабируемост.** Тъй като може да има произволно много данни, всяка система, която включва обработка на Big Data, трябва да бъде разширяема.

2. **Толерантност на повредите.** Принципът на хоризонтална мащабируемост предполага, че в клъстера може да има много машини. Например клъстерът Надоор на Yahoo има повече от 42 000 машини (тази връзка позволява да се видят размерите на клъстери в различни организации). Това означава, че част от тези

машини могат да се повредят. Методите за работа с големи данни трябва да отчитат възможността за подобни грешки и да се справят без значителни последици.

3. Локалност на данните. В големите системи данните се разпределят в голям брой машини. Ако данните са физически разположени на един сървър и се обработват на друг, цената на трансфера на данни може да надвиши цената на самата обработка. Следователно един от най-важните принципи за проектиране на Big Data решения е принципът на локализация на данните – ако е възможно, обработването на данните да се осъществява на същата машина, на която те се съхраняват.

Всички съвременни средства за работа с Big Data, по един или друг начин, следват тези три принципа.



Фигура 2. Хранилище за данни на Google

Прогнозата на IDC, публикувана в доклада „Изследване на дигиталната вселена“ (Digital Universe Study: Big Data, Bigger Digital Shadows and Biggest Growth in the Far East) от 2011 година, е, че данните ще нараснат с 50 пъти през 2020 г., а неструктурираните данни, като файлове, имейли и видео, ще представляват 90% от тези данните. Развлеченията, а не корпоративните данни, ще заемат по-голямата част от облака (46,7%). Ключов драйвър са машинно генерирани данни, които ще нараснат от 11% от цифрова вселена през 2005 година до над 40% през 2020 година. Инвестициите и разходите за IT хардуер, софтуер, услуги, telco (телекомуникационна компания) и персонал, които се приемат за „инфраструктура“ на дигиталната вселена, ще нараснат с 40% между 2012 и 2020 година. В доклада е посочено, че човечеството е навлязло в „ерата на Zettabyte“ – зетабайт (ZB) е трилион гигабайта, или милиард терабайта). Според IDC през 2020 г. цифровата вселена ще достигне 40³.

Каква е причината за увеличения обем на данни? Отговорът е: в света, в който живеем, източниците на данни са много – постъпващи данни от измерващи устройства, събития от радиочестотни идентификатори, от камерите за сигурност, от мобилните комуникации, електронната поща, записите на транзакции в

банковата система и търговската мрежа, от измервателни уреди и устройства в метеорологията, астрономията, транспорта, медицината, потоци данни от социалните мрежи, потоци данни за местоположението на абонатите на клетъчните мрежи и много други източници, които събират информация. Освен това се наблюдава и „тенденция съществуващите по-стари данни – документи, книги и фотографии, да се сканират и също да се представят в цифрова форма, която освен удобства при съхраняването създава отлични условия за обработката и извличането на ценна информация от тези данни“ (Barneva 2017). Изброените източници на данни (а и други) са отправна точка за възникването на Big Data. По този начин те проникват във всички сфери на човешката дейност – във фирмите, в научните изследвания, в държавния сектор и др.

Организациите създават огромни обеми от данни като голяма част от тях е във формат, който не съответства на традиционния структуриран формат на базите данни – видеозаписи, текстови документи, машинен код, геопространствени данни и т.н. Цялата тази информация се съхранява в разнообразни хранилища, понякога извън пределите на организацията. В резултат, независимо че имат достъп до огромен обем от данни, организациите нямат необходимите инструменти, за да намират зависимости между тези данни и на тази основа да правят значими изводи. Освен това данните се обновяват все по-често. Стига се до ситуация, в която традиционните методи за анализ на информацията не могат да се справят с огромните обеми постоянно обновявани данни, което води до необходимостта от нови Big Data технологии.

Възникват много въпроси. Какви данни са необходими? Какви качества, компетентности да се развиват в учениците? Как да се проследи напредъкът им, използвайки числа? Образованието е много сложна система, тя се занима-



Фигура 3. Източници на данни

ва с хора, поради тази причина данните трябва да се прилагат по отношение на всички компоненти на учебната среда: стандарти, програми, изпитни системи, технологии, както и контекста, в който учениците получават знанията.

Във филогенезиса си образователната система е натрупала значително количество данни. Въпросът за това как достъпно да започне обработката на този голям обем от данни, отпада благодарение на широкото използване на информационните и комуникационните технологии.

Според В. В. Утьомов и П. М. Горев в областта на образованието се разграничават пет основни типа данни:

- лични данни;
- данни за взаимодействието на учениците с електронни системи за обучение (електронни учебници, онлайн курсове);
- данни за ефективността на учебните материали;
- административни (общосистемни) данни;
- прогнозни данни (Utemov, Gorev 2018, p. 450).

На основа на анализ на много подходи и модели в техните презентации И.Д. Фрумин идентифицира три основни области на големите данни:

- свързана с мисленето (предимно критично и креативно мислене);
- свързана с взаимодействие с други хора (комуникация и сътрудничество);
- свързана с взаимодействие със себе си (саморегулация, рефлексивност и самоорганизация).

В съвременния свят управлението на мегаполисите е почти невъзможно без големи данни (Big Data). Тук става въпрос за потоци от трафик, здравеопазване, въвеждане на система за изкуствен интелект, градоустройствени решения, превенция на престъпността и редица други аспекти. Образованието, от тази гледна точка, е най-трудната област. Факт е, че няма да се появят големи данни, ако не бъде създадена модерна инфраструктура. Първият етап се свързва с появата на нови компютри, високоскоростен интернет и мрежа от електронни услуги. Днес милиарди хора използват цифрови услуги ежедневно. Един от елементите на основната инфраструктура в образованието е електронният дневник, който е важен инструмент. За да се разбере мащабът на големите данни, трябва да се отбележи, че в него се записват милиони училищни оценки на месец – списък на използваните образователни материали, видове тестове, всичко се записва в дневника. Чрез електронния дневник всеки ученик остава дигитален отпечатък, който отразява практически всички аспекти на обучението му – как се справя със задачите, какво е научил, в какви кръжоци участва, какви музеи, театри, технологични паркове посещава, какво е проектирал и изобретил, какви книги чете, в какви олимпиади участва, какви компетентности притежава и т.н.

Вторият етап вероятно ще е свързан със създаването и развитието на електронни училища, с библиотека от разнообразни дигитални образователни материали. Създаването на такава инфраструктура не е достатъчно, необходимо

е да се изгради и система за мотивация на потребителите и да се наситят ресурсите със съдържание.

Задача на следващия етап е да се създадат инструменти за анализ, с които учителите да оценяват резултатите от обучението, да прогнозира бъдещи постижения, да установят липса на знания и умения и да предприемат навременни мерки за отстраняването им в съответствие с образователните потребности на всеки ученик. По този начин училището получава възможност да изгради индивидуална образователна траектория, като отчита максимално способностите и потребностите на всеки ученик. Големите данни ще съдействат и на учениците да изградят/формират персонално портфолио, което да включва резултати от олимпиади, получени сертификати, завършени проекти и други постижения за целия период на обучение в училище.

Резултатите от тестове са лесни за интерпретиране в областта на математиката, естествените науки, четенето, но е много по-трудно да се прецени дали един ученик може да мисли. Когато става въпрос за познавателни умения – например критично мислене или за социални умения, т.е. за способността за взаимодействие с други хора, това са компетентности, които са трудни за измерване с помощта на инструментите на XX век. Инструментите на математическата статистика, машинното обучение, компютърната лингвистика, технологиите за изкуствен интелект отварят пътя за синтез на различни източници на данни и дават възможност за подобряване на образователната система.

Масовото навлизане и използване на технологиите в образованието, електронното обучение, стандартизираните тестове, електронното изпитване (в някои училища) оставят „дигитални отпечатъци“ и са една от причините за натрупването на голям обем от данни всеки ден, които могат да се използват с цел подобряване процеса на обучение. „Тези данни съдържат информация за скоростта на четене на материала, времето за отговор, най-често посещаваните страници, отговорите на въпроси от стандартни тестове и др. както на индивидуалния студент/ученик, така и сумарни данни за всеки урок и всеки тест“⁴. След анализ на данните може да се получи информация за успеваемостта, за грешките и пропуските на всеки ученик и на тази основа учителят да коригира (ако е необходимо) собствената си дейност, така че учениците да възприемат и усвояват по-добре учебното съдържание. Анализът на данните в реално време дава възможност за адаптиране и индивидуализиране на учебното съдържание, така че всеки ученик да получава помощ, обяснения, насочващи въпроси върху учебното съдържание, което го затруднява. Позитивното е, че е възможно и обратното – за учениците, за които представеното учебно съдържание е лесно и те се справят успешно с поставените задачи, трудността може да бъде повишена, така че да се удовлетворят техните по-високи потребности в интелектуален план. Така всеки ученик следва собствената си образователна траектория, усвоявайки учебното съдържание със собствен темп. Във всеки момент може да

види дейността си и да се сравни със съучениците си. Ако ученикът не проявява признаци на активност в рамките на две седмици, платформата изпраща известия до учителите, за да могат те бързо да се свържат с него и да го подкрепят. По този начин се преодолява един от недостатъците на съвременното обучение, свързан с това, че учителят преподава учебното съдържание със степен на трудност и с темп, ориентиран към средното познавателно равнище на учениците (които са по-голямата част от състава на класа), а децата с по-ниско, по-високо познавателно равнище и децата със специални образователни потребности практически остават изолирани от процеса на обучение.

Анализът на големите данни позволява на учениците да работят по индивидуални програми, т.е. да се персонализира обучението, тъй като тези данни дават информация какъв тип е ученикът, с каква част от съдържанието си взаимодейства, как се осъществява това взаимодействие, с кого и как си взаимодейства в процеса на обучение, къде е той, какви са интересите му, кога му е скучно, как преминаването на учебното съдържание се отразява на образователните му резултати, на кой етап от обучението се нуждае от помощ. Така ученето става адаптивно и личностно ориентирано. Анализът на големите данни може да ориентира учениците да изберат такова образование и кариера, които отговарят на личните им качества и интереси.

Използването на Big Data в образованието дава възможност за промяна на подходите за мониторинг и оценяване на образователните резултати. Мониторингът става постоянен. Интересът на учениците към непрекъснат мониторинг е свързан с факта, че анализът на данните позволява да се направи неговият учебен план индивидуален. Интересът на учителите е свързан с възможността за получаване на информация и обратна връзка от учениците за съдържанието (учебни ресурси), което те създават (интересно /не интересно, трудно/лесно и т.н.), както и за ефективното разпределение на ресурсите. Динамиката на образователните резултати се записва постоянно, като въз основа на тези данни могат да се формират модели (повтарящи се), чрез които да се преценява развитието на ученика.

Big Data дават възможност да се предвиди и контролира рискът от отпадане на учениците от училище. На основата на анализ на данни за академичната история на учениците, тяхната активност в дигиталната образователна среда и демографските данни се изчислява нивото на риска за отпадане за всеки ученик. Стойността на този показател периодично се изпраща на класния ръководител и на самия ученик. За по-голяма яснота данните могат да се визуализират: учениците, за които е най-вероятно да завършат срока/учебната година, се отбелязват в зелено, тези в риск – в жълто, а тези, които се нуждаят от незабавна помощ от класния ръководител/учител, са маркирани с червено. Подобна интерактивна система за големи данни дава възможност да се подобрят резултатите от обучението и да се намали процентът на отпадане.

Съществуват някои бариери, които затрудняват внедряването и използването на Big Data в образователния процес:

- разпръснатите данни и непостоянният достъп до тях (данните могат да бъдат защитени с парола, съхранявани на лични компютри и следователно недостъпни за агрегиране);
- страхът от загуба на данни, споровете за продължителността на тяхното съхранение, проблеми с обучението на персонала;
- конфиденциалността (проблеми със сигурността на данните и злоупотребата при споделени устройства) (Harper, Parkerson 2015).

Проблемите, свързани с поверителност и сигурност, в комбинация с неправилно тълкуване и злоупотреба с данни, трябва да се вземат предвид при прилагането на тази технология. Така ще се сведат до минимум моралните проблеми, правните и етичните рискове за учениците и учителите.

При използването на Big Data (в т.ч. и в образованието) е възможно е да възникнат проблеми, свързани с някои технологични съображения:

- **събирането на данни** – как организациите получават началните/оригиналните данни за съхранение, обработка и отчитане в инструментите за анализ;
- **съхранението на данни** – средата, използвана за поддръжка на аналитичните данни на организацията;
- **обработката на данни** – действията, извършени върху съхранената информация, за да я превърнат в бизнес интелект за анализиране;
- **отчитането и визуализирането на данни** – официалното представяне на резултатите от обработената заявка⁵.

Тези данни трябва да бъдат обработвани в реално време, като потребителите очакват да получат резултат в рамките на 60 секунди. Проблем е, че не всички данни са структурирани така, че да се класифицират бързо и лесно, голяма част от информацията е неструктурирана, а част е „полуструктурирана“.

Могат да възникнат и някои етични проблеми, свързани с правото на собственост върху данните и кой трябва да има достъп до тях, в кои случаи е необходимо да се забрани събирането и/или анализът на данни и как това да се регламентира практически, как да се избегне фалшивата информация разпространявана по интернет, подмяната на информацията и много други.

Според J. Avella et al. анализът на данни, свързани с обучението, е обещаващо изследователско поле, но за целта е необходимо да се познават добре методите (визуален анализ на данните, социален анализ, семантичен анализ и извличане на знания от данни), предимствата и предизвикателствата (проследяване, събиране, оценяване и анализиране на данните, нуждата от оптимизиране на учебната среда, въпроси относно етиката и уединението и други), свързани с използването му (Avella et al. 2016).

За да могат Big Data да се използват в сферата на образованието, е необходим значителен финансов ресурс. Аргументите за инвестиции в Big Data вари-

рат – от намаляване на рисковете до повишаване качеството на продуктите и услугите. Два от аргументите са безусловни лидери – по-бързото и обосновано приемане на решения и задълбочено изучаване на потребителите. Но истинското предимство е скоростта на получаване на отговор на възникналите въпроси. Използвайки средства за анализ на Big Data, организацията получава отговор за секунди и за минути, а не за часове и дни, а това радикално влияе върху процеси, които протичат в тях. Проблемът е да се задават правилните въпроси.

Данните се цифровизират, следователно е необходимо тяхното обработване и отчитане, което води до взривен ръст в изчислителната мощност и скоростта на предаване на информацията. В близко бъдеще уменията за работа с големи данни вероятно ще се окажат ключови, тъй като количеството информация расте стремително, а с това – и възможностите на цифровите технологии, които улесняват научния прогрес, виртуализацията и автоматизация на много процеси. Съвременните технологични възможности и големите масиви данните променят съществуващите начини за събиране, съхранение и обработка на информация. Тоталният преход на обществото към дигитализация увеличава количеството данни, с което днес се оперира. Специалистите в сферата на образованието, включително от училищната институция, са изправени пред необходимостта от висококачествено използване на разнообразни, понякога неструктурирани данни, чийто обем постоянно нараства. Най-новите технологии могат значително да опростят процесите на извличане, управление, анализ и интерпретация на огромен масив данни, както и да подобрят ефективността на работата с тях. Технологията Big Data има голям потенциал за революционна промяна в сферата на образованието. Възможностите за приложението ѝ са практически безкрайни и до голяма степен се детерминират от смелостта на заинтересованите страни и всички участници в образователния процес.

БЕЛЕЖКИ

1. ФЕЙДЛ, Ч., М. БЯЛИК, Б. ТРИЛЛИНГ. Четырехмерное образование. http://nios.ru/sites/nios.ru/files/poleznoe/4D_Education_0.pdf, последно влизане 30.04.2020, 10:53 ч.]
2. IBM. (2012). What is big data? Retrieved February 25, 2013, <http://www-01.ibm.com/software/data/bigdata/>, последно влизане 07.05.2021, 07:54 ч.
3. IDC Digital Universe Study: с. 4 – 12.
4. БЪРНЕВА, Р. (2017). Големите данни и образованието в информационно общество. <http://sci-gems.math.bas.bg:8080/jspui/bitstream/10525/2820/1/ERIS2017-book-p01.pdf>, последно влизане 05.05.2020, 12:14 ч.].
5. DAVID, V. (2005). Technological Considerations in Learning Analytics. <http://www.clomedia.com/2005/06/29/technological-considerations-in-learning-analytics/>, последно влизане 06.05.2020, 09:27 ч.

6. STATEVA, G. (2017). Problemi pri statisticheskia analiz na „golemi danni“ (big data), avtoreferat. https://www.uni-svishtov.bg/portal/getFile/37/Avtoreferat_G.Stateva.pdf, posledno vliane 06.05.2021, 09:56 ch. [Статева, Г. (2017). Проблеми при статистическия анализ на „големи данни“ (big data), автореферат. https://www.uni-svishtov.bg/portal/getFile/37/Avtoreferat_G.Stateva.pdf, последно влизане 06.05.2021, 09:56 ч.]
7. BIG DATA ОТ А ДО Я. Часть 1: Принципы работы с большими данными, парадигма MapReduce, <https://habr.com/ru/company/dca/blog/267361/>, последно влизане 05.05.2021, 10:04 ч.
8. RICH, S. (2012). Big Data Is a „New Natural Resource“. <http://www.govtech.com/policy-management/BigData-Is-a-New-Natural-Resource-IBM-Says.html>, последно влизане 06.05.2021, 10:52 ч.
9. ROUSE, M. (2011). big data (Big Data) <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>, последно влизане 06.05.2021, 11:14 ч.
10. <http://www.unkniga.ru/vishee/9614-bolshie-dannye-v-obrazovanii.html>, последно влизане 05.05.2021, 09:14 ч.
11. <https://trends.google.com/trends/explore?date=all&geo=US&q=big%20data>, последно влизане 05.05.2021, 09:38 ч.
12. <https://calcme.biz/big-data>, последно влизане 05.05.2021, 10:21 ч.
13. www.whizpr.be, последно влизане 05.05.2021, 11:16 ч.

REFERENCES

- AVELLA, J. ET AL., 2016. Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. *Online Learning*, vol. 20, pp. 13 – 29.
- BOYD, D. & CRAWFORD, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, vol. 15, no. 5, pp. 662 – 679.
- DAVENPORT, T. H.; BARTH, P. & BEAN, R., 2012. How Big Data Is Different. *MIT Sloan Management Review*, vol. 54, no. 1, pp. 43 – 46.
- FOSSO WAMBA, S.; AKTER, S.; EDWARDS, A.; CHOPIN, G. AND GNANZOU, D., 2015. How ‘big data’ can make big impact: findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, vol. 165, pp. 234 – 246.
- FISHER, D.; DELINE, R.; CZERWINSKI, M. & DRUCKER, S., 2012. Interactions with Big Data Analytics. *Interactions*, vol. 19, no. 3, p. 50.
- GOLDIN, C. D.; KATZ, L. F., 2009. *The Race between Education and Technology*. Cambridge, MA: Harvard University Press.
- GOBBLE, M. M., 2013. Big Data: The Next Big Thing in Innovation. *Research Technology Management*, vol. 56, no. 1, pp. 64 – 66.

- HARPER, E. M.; PARKERSON, S., 2015. Powering big data for nursing through partnership. *Nurs. Adm. Q.*, vol. 39, no. 4, pp. 319 – 324.
- HAVENS, T. C.; BEZDEK, J. C.; LECKIE, C.; HALL, L. O. & Palaniswami, M., 2012. Fuzzy c-Means Algorithms for Very Large Data. *Fuzzy Systems. IEEE Transactions on*, vol. 20, no. 6, pp. 1130 – 1146. Available from: doi: 10.1109/tfuzz.2012.2201485.
- JACOBS, A., 2009. The Pathologies of Big Data. Association for Computing Machinery. *Communications of the ACM*, vol. 52, no. 8, p. 36.
- JOHNSON, B. D., 2012. The Secret Life of Data. *The Futurist*, vol. 46, no. 4, pp. 20 – 23.
- KELLER, A.; KOONIN, S. & SHIPP, S., 2012. Big data and city living - what can it do for us? *Significance*, vol. 9, no. 4, pp. 4-7. Available from: doi: 10.1111/j.1740-9713.2012.00583.x
- MANYIKA, J.; CHUI, M.; BROWN, B.; BUGHIN, J.; DOBBS, R.; ROXBURGH, C. & BYERS, A. H., 2011. *Big data: the next frontier for innovation, competition and productivity*. McKinsey Global Institute.
- MCAFEE, A. & BRYNJOLFSSON, E., 2012. Big data: the management revolution. *Harvard business review*, no. 1.
- NEAMTU, R. ET AL., 2018. *The impact of Big Data on making evidence-based decisions*, in *Frontiers in Data Science*.
- STRAWN, G. O., 2012. Scientific Research: How Many Paradigms? *EDUCAUSE Review*, vol. 47, no. 3, p. 26.
- UTEMOV, V. V.; GOREV, P. M., 2018. Razvitie obrazovatelynyh sistem na osnove tehnologii Big Data. *Nauchno-metodicheskiy elektronniy zhurnal Kontsept*, no. 6, pp. 449 – 461.

BIG DATA IN THE EDUCATION

Abstract. In the article an attempt is made to describe the possibilities of application of Big data technology in the field of education. It presents many definitions, the principles of working with Big Data, some of the possible applications of technology in the learning process, as well as possible barriers to its use in education.

Keywords: big data; education; learning process; students

✉ **Prof. Natalia Vitanova, DSc.**

ORCID iD: 0000-0002-6093-9672

Faculty of Education

University of Shumen “Bishop Konstantin Preslavski”

Shumen, Bulgaria

E-mail: n.vitanova@shu.bg