

ИНОВАТИВНИ ИНФОРМАЦИОННИ ТЕХНОЛОГИИ ЗА ДИГИТАЛИЗАЦИЯ НА ИКОНОМИКАТА

Сборник с доклади



ИЗДАТЕЛСКИ КОМПЛЕКС – УНСС

УНИВЕРСИТЕТ ЗА НАЦИОНАЛНО И СВЕТОВНО СТОПАНСТВО
КАТЕДРА „ИНФОРМАЦИОННИ ТЕХНОЛОГИИ И КОМУНИКАЦИИ“

**ИНОВАТИВНИ
ИНФОРМАЦИОННИ
ТЕХНОЛОГИИ
ЗА
ДИГИТАЛИЗАЦИЯ
НА ИКОНОМИКАТА**

Сборник доклади от юбилейна научна конференция
по повод 55 годишнината от основаването
на катедра „Информационни технологии и комуникации“
20 октомври 2023 г.

Издателски комплекс – УНСС
София, 2024

Научна редакция и съставители:

доц. д-р Ваня Лазарова
гл. ас. Веска Михова

Програмен комитет:

проф. д.ик.н. Валентин Кисимов
проф. д-р Любен Боянов
проф. д-р Камелия Стефанова
доц. д-р Александрина Мурджева
доц. д-р Пламен Милев
доц. д-р Митко Радоев
доц. д-р Моника Цанева
гл. ас. д-р Мария Мързованова
ас. д-р Мариана Ковачева
ас. д-р Ивона Велкова

Организационен комитет:**Председател:**

гл. ас. д-р Веска Михова

Членове:

доц. д-р Ваня Лазарова
доц. д-р Дорина Кабакчиева
гл. ас. д-р Станимира Йорданова
гл. ас. д-р Иван Белев
гл. ас. д-р Гено Стефанов
ас. д-р Явор Табов

Всички права са запазени! Не се разрешават копиране, възпроизвеждане и разпространение на книги или на части от тях по какъвто и да е начин без писменото разрешение на ИК – УНСС. Издателският комплекс не носи отговорност за съдържанието и стила на изложение на авторите.

© Колектив

© ИЗДАТЕЛСКИ КОМПЛЕКС – УНСС

Изп. директор: Веселин Ангелов, тел. 02 8195 251
Зам. изп. директор: Стефан Власев, тел. 02 8195 551
Гл. редактор: Тодорина Недева, тел. 02 8195 564

УНИВЕРСИТЕТ ЗА НАЦИОНАЛНО И СВЕТОВНО СТОПАНСТВО
1700 София, район „Студентски“, ул. „8-ми декември“ № 19

UNIVERSITY OF NATIONAL AND WORLD ECONOMY
DEPARTMENT „INFORMATION TECHNOLOGIES AND COMMUNICATIONS“

**INNOVATIVE
INFORMATION
TECHNOLOGIES
FOR ECONOMY
DIGITALIZATION
(I I T E D)**

Conference Proceedings from the Jubilee Scientific Conference
on the 55th Anniversary of the Founding of the Department
„Information Technologies and Communications“
October 20, 2023

Publishing complex
Sofia, 2024

СЪДЪРЖАНИЕ

Съдържание	5
ОБЕКТНО-ОРИЕНТИРАН ПОДХОД ЗА СТРУКТУРИРАНЕ НА СЪДЪРЖАНИЕ.....	7
Иван Белев	
HOW OPEN-SOURCE LARGE LANGUAGE MODELS ARE DRIVING THE DIGITAL TRANSFORMATION WAVE.....	15
Bozhidar Bahov	
TOWARDS A UNIFIED CONCEPT FOR SHARING AND LINKING OF DATA RESOURCES IN COLLABORATIVE RESEARCH VIA WEB SERVICES.....	20
Plamen Milev	
ТЕХНОЛОГИЧНИ РЕШЕНИЯ И ДИГИТАЛИЗАЦИЯ В СВАТЪБЕНИЯ ТУРИЗЪМ....	30
Симона Маджарова	
ДИГИТАЛНА ТРАНСФОРМАЦИЯ НА СЧЕТОВОДНАТА ОТЧЕТНОСТ В БЮДЖЕТНИТЕ ОРГАНИЗАЦИИ.....	35
Митко Радоев	
РАЗВИТИЕ НА ТУРИСТИЧЕСКИЯ ПАЗАР В БЪЛГАРИЯ ЧРЕЗ ДИГИТАЛИЗАЦИЯ В ТУРИЗМА	42
Дарина Георгиева Цветанова	
Борис Георгиев Карагеоргиев	
ЕЛЕКТРОННИТЕ АДМИНИСТРАТИВНИ УСЛУГИ НА ОБЩИНИТЕ – РЕГИОНАЛНИ АСПЕКТИ И ИЗМЕРЕНИЯ.....	50
Асен Божиков	
ДИГИТАЛИЗАЦИЯ ВЪВ ФИНАНСОВИЯ СЕКТОР	59
Аглика Кънева	
ИЗКУСТВЕНИЯТ ИНТЕЛЕКТ – ОТ ПЕРЦЕПТРОН ДО CHATGPT	68
Ваня Лазарова	
ARTIFICIAL INTELLIGENCE AND TAX ADMINISTRATION IN BULGARIA	75
Georgi Emilov Hristov	
НАМАЛЯВАНЕ НА РЕКЛАМНИТЕ ИЗМАМИ В ДИГИТАЛНИЯ МАРКЕТИНГ С ПОМОЩТА НА ИЗКУСТВЕН ИНТЕЛЕКТ	79
Ивона Велкова	
ВЪЗМОЖНОСТИ И ПРЕДИЗВИКАТЕЛСТВА ПРЕД БАНКИТЕ ОТ ПРИЛОЖЕНИЕТО НА ИЗКУСТВЕН ИНТЕЛЕКТ ПРИ УПРАВЛЕНИЕТО НА КРЕДИТНИЯ РИСК	89
Кирил Аначков	
ГОЛЕМИ ДАННИ В ЗДРАВЕОПАЗВАНЕТО	97
Любен Зюмбилски	
КАТАЛОГИЗИРАНЕ НА ДАННИ В СРЕДА НА ГОЛЕМИ ДАННИ.....	102
Гено Стефанов	
ЕЗЕРА ОТ ДАННИ(DATA LAKES) КАТО ХРАНИЛИЩЕ ЗА ГОЛЕМИ ДАННИ. АНАЛИЗ НА AWS ПОДХОДА.....	108
Гено Стефанов	

SIGNIFICANCE AND COMPARABILITY OF UNSTRUCTURED AND SEMI-STRUCTURED DATA IN THE MODERN WEB	113
Yavor Tabov	
DESIGN AND IMPLEMENTATION OF PROJECT MANAGEMENT APPLICATION IN REACT.JS.....	120
Daniel Nalbantov	
Victor Del Aguila Angulo, Flavio Gorini, Alessio Mancini, Nicola Rispoli	
OPERATING AND ANALYSING OF BIG DATA OVER PENTAHO BUSINESS ANALYTICS, TABLEAU AND QLIK	125
Mariana Kovacheva	
ТЕХНИЧЕСКИ РЕСУРСИ ЗА ПОДПОМАГАНЕ НА ХОРАТА СЪС ЗРИТЕЛНИ УВРЕЖДАНИЯ.....	132
Елена Филипова	
ПРОДУКТОВИ КАТЕГОРИИ ЗА ХОРА СЪС ЗРИТЕЛНИ УВРЕЖДАНИЯ	139
Елена Филипова	
DATA PREPARATION TECHNIQUES AND PLATFORMS IN THE CONTEXT OF MACHINE LEARNING	144
Genka Miteva	
Aleksandrina Murdzheva	
РАЗШИРЯВАНЕ ВЪЗМОЖНОСТИТЕ НА ИНФОРМАЦИОННА СИСТЕМА ЗА РАБОТА С ГЛАС.....	152
Веска Михова	
АРХИТЕКТУРА НА ИНФОРМАЦИОННА СИСТЕМА ЗА РЕАЛНИ И СИМУЛИРАНИ СТОЙНОСТИ НА IoT УСТРОЙСТВА	158
Александър Танев	
УПРАВЛЕНИЕ НА ДИГИТАЛНАТА ТРАНСФОРМАЦИЯ – НОВИ ПОДХОДИ И РОЛЯ НА КОНСУЛТАНСКИТЕ КОМПАНИИ.....	165
Цветомир Цанов	

ОБЕКТНО-ОРИЕНТИРАН ПОДХОД ЗА СТРУКТУРИРАНЕ НА СЪДЪРЖАНИЕ

Иван Белев

Главен асистент, доктор, катедра „Информационни технологии и комуникации“, УНСС,
e-mail: i.belev@unwe.bg

Резюме

Научното изследване разглежда обектно-ориентиран подход при структуриране, съхранение и управление на съдържание в организацията. Представени са сценарии, чрез които се прави опит да се демонстрират ползите при използване на обектно-ориентирания подход в системи от тип Enterprise Content Management.

Ключови думи: системи за управление на съдържанието, enterprise content management, ECM, обектно-ориентиран подход, структуриране.

OBJECT-ORIENTED ECM APPROACH

Ivan Belev

Abstract

The paper explores the possibility to use an object-oriented approach for structuring, storing and managing content in the organization. The author presents use cases that try to demonstrate the advantages of using an object-oriented approach in Enterprise Content Management systems.

Keywords: enterprise content management, ECM, object-oriented approach, structuring.

Съдържание и управление на съдържание в организациите

Понятието съдържание заема централна част в контекста на текущата разработка. В дигиталната ера, в която живеем, съдържанието е основният ресурс на голяма част от организациите. То е основа за взимане на решения, както и фактор за иновации и взаимодействие в организациите. Обемът и разнообразието на съдържание, което се създава в организациите или е необходимо за дейността им, достига невиджани досега нива.

В контекста на организациите, съдържание обикновено означава информация във някаква форма. Традиционно съдържанието е под формата на хартиени документи. Но с еволюцията на технологиите и технологичният напредък на организациите, под съдържание вече може да се има предвид и следната информация:

- дигитални документи;
- мултимедийно съдържание;
- съдържание в социални мрежи;
- електронни съобщения;
- съдържание в web страници.

Разглеждайки класификацията по-горе става ясно, че под термина съдържание в общия случай може да се има предвид неструктурирани източници на информация в организацията. Независимо от формата на съдържанието, правилното управление на това богатство от

информация е ключово за успеха на всяка една организация. В това отношение основна роля имат системите за управление на съдържанието в организацията – Enterprise Content Management (ECM). По отношение на понятието Enterprise Content Management, може да приемем българския термин – Управление на съдържанието на предприятието. Авторите Alalwan и Weistroffer [2] правят обобщение на различни определения за ECM от няколко литературни източника ([4], [5] и [6]). общите точки в обобщените от тях дефиниции определят термина Управление на съдържанието на предприятието като съвкупност от стратегии, методи, инструменти, процеси, умения и технологии за управление на пълния жизнен цикъл на съдържанието в организацията – създаване/извличане, откриване, съхраняване, управление, опазване, търсене/намиране, публикуване и предоставяне. Gartner [7] добавят към определението за Enterprise Content Management и анализа на използването на съдържанието в организацията с цел по-добро осигуряване на правилното съдържание на крайните ползватели, където те имат нужда от него.

Традиционен подход за съхранение и структуриране на съдържание

Поглеждайки назад във времето, съхранението и структурирането на съдържание е много различно от това, което технологиите за управление на съдържанието предоставят. Преди използването на софтуер за целите на управление на съдържанието, организациите съхраняват своите документи, използвайки:

- физически папки.
- кутии.
- шкафове.
- библиотеки.
- складови помещения.

За откриването и управлението на документите се създават и поддържат индекси, каталози и други форми на обозначение, чрез които служител собственоръчно открива необходимия документ.

С въвеждането и разпространението на компютърните технологии, съхранението и структурирането на съдържание в организациите се премества от физически носители към дигитални форми на съдържание. Организациите създават дигитално пространство, в което съхраняват дигиталните си документи и друго съдържание. За да бъде лесно откриваемо се прилага дигитална аналогия на физическите папки – компютърни директории. Използвайки директории и възможността те да бъдат именувани по избран начин, както и тяхното свойство да могат да помещават други директории, образувайки йерархия, дава възможност на организациите да приложат подход за структуриране на съдържанието, с който да осигурят дигитален аналог на описаното в предходния абзац – форма на обозначение, чрез която да се осигури по-лесно откриване на файлове/документи.

Прилагайки описания по-горе подход на практика означава следното – организациите създават йерархична структура от компютърни директории, организирано по специфичен за организацията начин, в която разпределят дигиталните документи/файлове. При необходимост за търсене на даден документ/файл в последствие, той бива намерен като се търси в съответната директория в структурата. За целите на научното изследване ще бъде представен практически пример, чрез който да се илюстрират разглежданите подходи за структуриране и техните предимства и недостатъци. Представяйки различните подходи за структуриране по-долу и практическите примери, авторът ще използва термина „документи“ като обобщаващ термин за различните видове съдържание в дигитален формат.

Организация може да започне да съхранява своите документи/файлове в директории, като организира на първо ниво отделни директории за клиентските документи и за други вътрешни

за организацията документи, които не са свързани с нейните клиенти. Разглеждайки клиентските документи, на второ ниво може да се създаде отделна директория за всеки един клиент, в която да се съхраняват документите на дадения клиент. В случай, че клиентите имат дългосрочна история в организацията, на следващо ниво може да се създаде директория за всяка отделна година, в която организацията съхранява документи за дадения клиент. Всяка отделна директория на ниво година. В случай, че документите в рамките на един клиент за една година са много, на следващо ниво може да се създаде различна директория за основни типове документи, които са обект на търсене. Ако погледнем от гледна точка на търговски организации, това може да бъдат отделни директории някои често срещани типове документи в тази сфера:

- Оферти.
- Договори.
- Фактури.
- и др.

При така представения примерен сценарий, оферта на даден клиент (Client A) от 2024 година ще бъде съхранена в следната директория:

Клиентски документи -> Client A -> 2024 -> Оферти



Фигура 1: Пример за традиционен подход за съхранение и структуриране на съдържание в йерархия от директории

Този подход за структуриране и съхранение на съдържание би работил без особени проблеми при сравнително статични параметри, които се използват в организацията за организиране и откриване на съдържание, както и при сравнително малък брой уникални стойности на различните нива в йерархията.

В посочения пример става ясно, че ключова роля за структуриране на съдържанието в организацията и в последствие неговото лесно откриване имат именно параметрите или данните за всеки един съхраняван документ като единица съдържание. Тези параметри в терминологията на Enterprise Content Management се наричат свойства на документа (properties) или също така – метаданни (metadata). Милев и Табов [3] дефинират метаданните като структурирана информация, която описва, обяснява, открива или дава възможност лесно да бъде открит даден информационен ресурс.

Така реализирания подход за съхранение и структуриране на съдържание е приложим в организациите до достигане на определени нива на динамика в дейността, както и до достигане на определени обеми от съдържание. Бързо развиващата се среда на дейност на организациите от всякакъв тип в днешните условия водят до повишени нива на динамика и големи обеми, които извеждат на преден план недостатъците на представения традиционен подход:

- трудна навигация в структурата от директории при голям брой директории на всяко ниво от йерархията.
- трудно управление на структурата при постоянно променящи се изисквания по отношение на параметрите (метаданните) на отделните единици съдържание.
- трудно извличане на документи при необходимост да се търси по повече от една стойност на метаданните.

Може да бъде представен следните примери, за да се илюстрират недостатъците на този подход за структуриране и съхранение на съдържанието:

- При голям брой клиенти с дълга история в организацията (например в порядъка на милиони) би било много трудно да се открие директорията на конкретния клиент. Освен това всяко търсене трябва да премине през последователно намиране на съответната директория на всяко едно от нивата в йерархичната структура.
- При поява на евентуално ново изискване за отделяне на клиентите по техния тип например (физически лица и юридически лица, или по сегмент на клиента) ще се наложи структурата да бъде преобразувана като се добавят нови нива и се разпределят текущите директории в тях.
- Евентуално изискване да се намерят и извадят всички фактури за две избрани години без значение кой е клиентът би довело до множество търсения на различни нива, особено при голям брой клиенти. В този случай трябва да се търси вариант за допълнителна организация на документите само по техния тип например.

Описаното дотук дава основание да се потърси друг подход за структуриране на съдържанието. Такъв подход се предлага от системите за управление на съдържание от тип Enterprise Content Management.

Съхранение и структуриране на съдържание в ЕСМ

ЕСМ системите въвеждат допълнителен компонент за да съхраняват и използват в структуриран вид метаданните на съдържанието в организацията – обикновено чрез база от данни. Всички метаданни за даден документ (единица съдържание) се съхраняват като запис в база от данни заедно с описание на „пътя“ до физическото разположение на документа/файла. Документите продължават да бъдат съхранявани в структура от директории, но тя вече не се използва за йерархия, в която да се търси по името на определената директория.

Следвайки примера, който бе използван до момента, след въвеждане на описаната по-горе организация за съхраняване на документите, намирането на фактурите на клиент Client A за 2024-та година би било възможно след следната заявка в базата данни:

Всички документи за Име на клиент: Client A, Година: 2024, Тип на документ: Оферта

За разлика от традиционния подход, при този подход лесно може да бъдат правени сечения и комбинации между стойностите на различните метаданни, например:

Всички документи за Година: 2024, Тип на документ: Оферта или Договор

Използвайки такъв подход решава голяма част от проблемите описани до момента, но някои от тях остават, като се прехвърлят на друго ниво – в организацията на структурата на базата данни. Повишената динамика на изискванията по отношение на метаданните на документите и техните различни стойности води до трудности по отношение на:

- броя колони в базата данни.
- номенклатурата от стойности в базата данни.
- обособяването на различни таблици в базата данни за съхранение на документи с различен набор от метаданни.

Може да се представи следният опростен пример, за да се илюстрират част от описаните проблеми: Ако в организацията е необходимо да се съхраняват общи клиентски документи, за които има един набор от метаданни:

- тип на клиент (например физическо и юридическо лице).
- идентификатор на клиент.
- категория на документ (например Документи за идентификация).
- тип на документ (например лична карта, паспорт, пълномощно).
- дата на документ.
- потребител.

В същото време е необходимо да се съхраняват вътрешни документи с правно естество, които не са свързани с определен клиент и имат друг набор от метаданни:

- номер на дело.
- отдел/департамент.
- тип на документ.
- дата на документ.
- срок за резолюция.

Може да се добави и още един пример – съхраняване на документи, свързани с продуктите на клиента, при които има следния набор от метаданни:

- тип на клиент (например физическо и юридическо лице).
- идентификатор на клиент.
- категория на документ (например Договорни документи).
- тип на документ (например договор, анекс, фактура).
- Тип на продукт (например застраховка, кредит)
- дата на документ.
- потребител.

Разглеждайки изискванията за структуриране на съдържанието в посочените три примера се вижда, че в структурата на базата от данни за съхранение на метаданните на документите е необходимо да има различни таблици, в които да се разделят отделните видове документи. Същевременно между тях остават някои общи метаданни, които ще бъдат дублирани между отделните таблици. Това би довело до проблеми при извличането на данни, подобни на проблемите с традиционния подход. Може да се наложи да се обединяват резултати от повече от една заявка към базата от данни. Също така е възможно да има трудности при дефиниране и управление на колони за метаданни, които присъстват едновременно при повече от един вид документи – необходимост от индексирание, обединяване, сортиране, филтриране на резултати.

За решаване на част от тези проблеми може да бъде приложен изменен подход за съхранение и структуриране на съдържание в системите за управление на съдържанието, който е обектно-ориентиран.

Обектно-ориентиран подход за структуриране на съдържание в ЕСМ

Еволюцията на технологии води до дефинирането на обектно-ориентирания подход за реализация на различни задачи в света на ИТ. Може би най-разпознаваем е терминът за обектно-ориентирано програмиране, използван в редица програмни езици и технологии. Стъпвайки на предимствата на подход като дизайн модел [8], някои системи за управление на съдържанието успешно го прилагат за структуриране на метаданните на отделните единици съдържание (документи) в системата. Такава система например е IBM FileNet P8 Platform [9].

Използвайки обектно-ориентиран подход за структуриране на съдържанието в ЕСМ системите може да доведе до елиминиране на недостатъците, описани до момента в изследването. При този подход всички структурно-различни видове съдържание (документи) могат да се дефинират като отделни класове (и под-класове) в хоризонтална и вертикална (йерархична) структура. При създаването на структурата се следват няколко основни правила:

- Всички метаданни се дефинират като отделни *properties*, които могат да бъдат използвани в един или повече класове документи.
- Всеки клас документ може да приема определен набор от избрани метаданни (*properties*).
- Всеки под-клас наследява метаданните на своя „родител“ (*parent class*) и от своя страна може да приема и други метаданни, специфични за него, както и да има свои под-класове.

За представяне на предимствата на обектно-ориентирания подход за структуриране на съдържанието може отново да се даде пример, който е сходен на използваните до момента в изследването:

На най-високо ниво се дефинира клас Документ със следните метаданни:

- Заглавие.
- Дата на създаване.
- Потребител.

На следващо ниво като под-клас на клас Документ се дефинира клас Документ на организацията, който наследява метаданните на клас Документ и има в допълнение следните метаданни:

- Тип на документ.
- Статус.

На следващо ниво се създават два под-класа – Клиентски документ и Документ Човешки ресурси, които наследява метаданните на клас Документ и клас Документ на организацията, но в допълнение имат и собствени метаданни. Например за клас Клиентски документ:

- Идентификатор на клиент.
- Тип на клиент.

На последно ниво се дефинира клас Продуктов документ, който е под-клас на Клиентски документ, наследява всички метаданни нагоре в йерархията и има следните допълнителни метаданни:

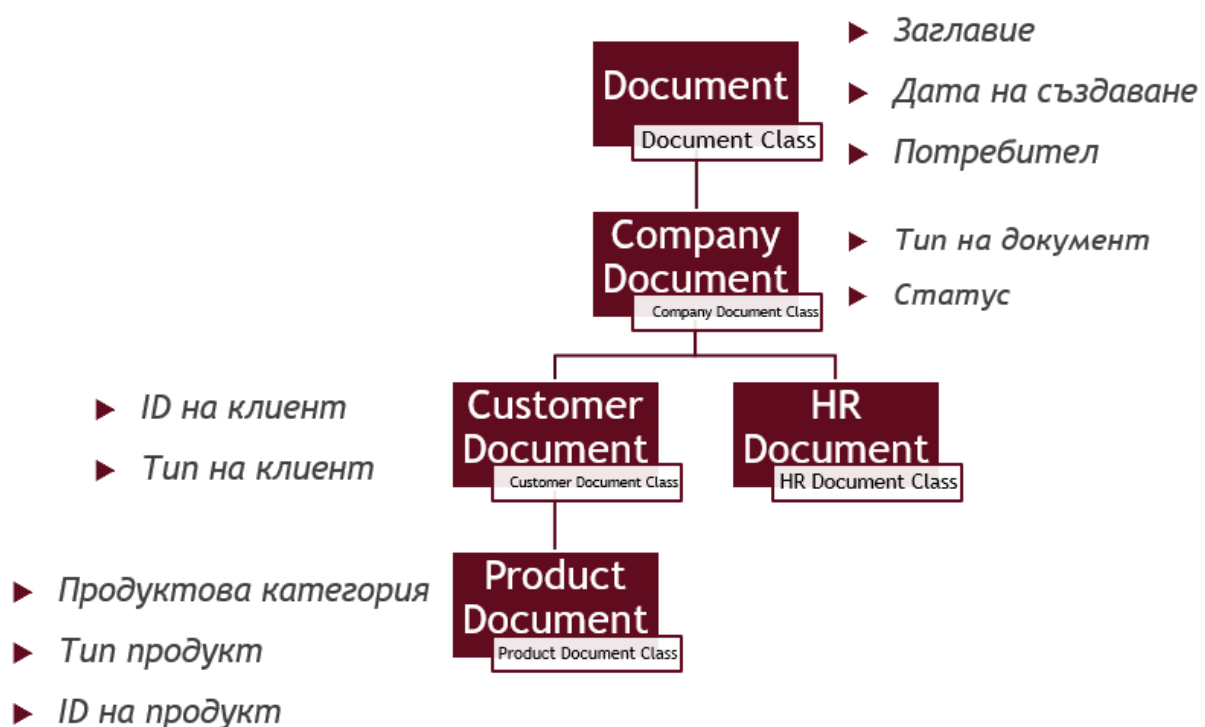
- Продуктова категория.
- Тип продукт.
- Идентификатор на продукт.

Така дефинираната структура дава следните предимства:

- На лице е ясно разграничение между структурно различните видове документи в организацията.

- Могат да се търсят различните видове документи, като се използва критерия за клас на документа в комбинация с метаданните на конкретния клас, независимо дали са негови собствени или наследени от „родителските“ класове.
- Същевременно е възможно да се търсят документи от даден клас, като търсенето да включва и документи от неговите под-класове, които отговарят на зададените критерии.
- Структурата се управлява сравнително лесно като за всеки нов клас или под-клас, може да се използват отново предварително дефинираните метаданни (properties) без да е необходимо да се създават нови. Това позволява лесно да се търси по един и същ критерий между различни класове и под-класове, които притежават общи метаданни.
- Сравнително лесно управление на структурата чрез добавяне на нови метаданни на високо ниво в даден клас, което води до автоматичното им създаване и в техните под-класове.

Дефинираният пример може да бъде представен визуално чрез следната диаграма:



Фигура 2: Пример за обектно-ориентиран подход за съхранение и структуриране на съдържание

Реализирайки подобен подход за структуриране на съдържанието в организацията дава възможност да се използват предимствата на системите от Enterprise Content Management по отношение на бързото намиране на дадена единица съдържание в комбинация с елиминиране на трудностите, свързани с дефиниране и управление на самата структура.

Могат да бъдат представени следните примери, за заявки за извличане на съдържание в дефинираната примерна обектно-ориентирана структура:

Всички документи от клас Продуктов документ за ID на клиент: Client A, Тип на продукт: Застраховка, Потребител: Иван Иванов, Дата на създаване: през 2023 година.

Всички документи от клас Клиентски документ (включително документи от под-клас Продуктови документи) за Тип на клиент: Физическо лице, Тип на документ: Фактура, Статус на документ: Активен, Дата на документ: през месец Октомври 2023

Заклучение

Динамичният свят, в който всички организации развиват своята дейност поставя редица предизвикателства пред тях по отношение на съдържанието в организацията и управлението на неговия жизнен цикъл. За справяне с предизвикателствата организациите могат да използват системи за управление на съдържанието от тип Enterprise Content Management, избирайки такава система, която дава възможност да се приложи обектно-ориентиран подход при създаването на структурата на съдържанието. Използвайки такъв подход организациите могат максимално да се възползват от предимствата му и да успеят да се справят в голяма степен с трудностите, свързани със структурирането и управлението на съдържанието в организацията.

References

1. Cameron, S. A. (2011). *Enterprise content management: a business and technical guide*. BCS, The Chartered Institute.
2. Alalwan, J. A., & Weistroffer, H. R. (2012). *Enterprise content management research: a comprehensive review*. *Journal of Enterprise Information Management*, 25(5), 441-461.
3. Milev, P., & Tabov, Y. (2022). *Conceptual Approach For Presenting Text Data From Web-Based Information Systems In Structured Form*.
4. Smith, H.A. and McKeen, J.D. (2003), “*Developments in practice VIII: enterprise content management*”, *Communications of the Association for Information Systems*, Vol. 11, pp. 647-59
5. Woolley, R. and Fletcher, D. (2007), “*Research summary: enterprise content management*”, white paper, Department of Technology Services, available at: <http://utah.pdfs.com/awweb/awarchive?type=file&item=18380>
6. Fowler, D. (2008), “*Implementing enterprise content management using Microsoft SharePoint*”, Master’s thesis, University of Oregon, Eugene, OR
7. *Definition of Enterprise Content Management (ECM)* - Gartner Information Technology Glossary. (n.d.). Gartner. <https://www.gartner.com/en/information-technology/glossary/enterprise-content-management-ecm>
8. Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns: Elements of reusable object-oriented software*. Addison Wesley.
9. Zhu, W.-D., Buchanan, N., Oland, M., Poggensee, T., Romero, P. E., Snow, C., Worel, M., & Redbooks, I. B. M. (2011). *IBM FileNet P8 Platform and architecture*. IBM

HOW OPEN-SOURCE LARGE LANGUAGE MODELS ARE DRIVING THE DIGITAL TRANSFORMATION WAVE

Bozhidar Bahov

Ph.D. candidate, Department of Information Technologies and Communications, University of National and World Economy, e-mail: bozhidar.bahov@unwe.bg

Abstract

This paper examines the capabilities, associated challenges, and competitive advantages of open-source Large Language Models (LLMs). We analyze their role as a significant driver of digital transformation, offering advantages like cost-effectiveness and adaptability. Although they can contribute to risks around data privacy, bias, and security, the results of our research indicate that open-source LLMs can serve as viable alternatives for proprietary models.

Key words: LLMs, open-source, digital transformation

Introduction

Artificial intelligence has reached new highs in recent years, transforming our understanding and interaction with technology. One notable leap in this trajectory is the emergence and utilization of Large Language Models (LLMs). These deep learning-based models can generate human-like language, conduct complex calculations, and aid in problem-solving, transforming various professional sectors and disciplines. While proprietary LLMs have attracted considerable attention, it is the open-source ones that offer a fascinating landscape for diverse stakeholders, including academics, researchers, practitioners, and policymakers.

The significance of open-source LLMs is multifaceted. They promise to democratize AI by making advanced computational capabilities accessible to all. However, they also introduce various ethical and operational challenges that require careful examination. Data privacy, model hallucinations, inherent biases, and security vulnerabilities are at the forefront of discussions within the scientific and business communities.

This article systematically analyzes open-source LLMs in the contemporary digital ecosystem. It explores their practical applications, ethical and operational challenges, and competitive performance compared to proprietary counterparts. The objective is to provide a well-balanced perspective supported by empirical evidence, enabling a nuanced understanding of the capabilities and limitations of open-source LLMs and promoting responsible usage and informed decision-making.

Open-Source LLMs: A New Paradigm

As early as in the 2007 paper “The Need for Open-Source Software in Machine Learning,” Sonnenburg made the case that advancement is hindered by a lack of publicly accessible algorithmic implementations. To improve the accessibility, interoperability, and sharing of potent learning algorithms for various applications, machine learning requires open-source software. In a more recent academic article titled “How Open-Source Machine Learning Software Shapes AI,” Max Langenkamp suggests that open-source software plays a substantial role in generating economic value within the field of AI. According to Langenkamp's research, every dollar invested in open-source machine learning software generates a minimum of \$100 worth of global economic value (Langenkamp, 2022). This added value is generated by increasing experimentation, standardization, and community building made possible by open-source machine learning technologies.

On the other hand, LLMs have emerged as a specialized and highly impactful subset of machine learning technologies. These models are designed to manipulate and generate text based on extensive datasets. Beyond mere text generation, LLMs extend their functionalities to include tasks such as

language translation, content summarization, sentiment analysis, and even virtual tutoring and content moderation. These expansive functionalities can be accessed through different LLMs, most prominently classified into proprietary and open-source models. The former are commercial products governed by licenses. In contrast, the latter represents a transformative approach that melds the intricate capabilities of LLMs with the philosophical underpinnings of open-source development.

Building upon these classifications, it is essential to delve into the specific attributes that set open-source LLMs apart. In contrast to proprietary LLMs, which are constrained by ownership arrangements and accessibility limits, open-source LLMs offer more freedom. These models exemplify a paradigm characterized by inclusivity and adaptability, driven by collective participation and dispersed development principles. Not only is the source code accessible, but the fundamental structure, algorithms, and design principles are also openly available. This allows a diverse group of developers, researchers, and organizations to analyze, adapt, and improve these models to their specific requirements. Furthermore, utilizing open-source LLMs frequently leads to substantial cost reductions, matching the broader economic advantages of open-source machine learning.

The open-source architecture in LLMs offers more than just code visibility; it creates transparency that serves multiple stakeholders, including businesses without dedicated machine learning teams and academic researchers. It provides a clear and trustworthy environment that fosters compliance with ethical and legal standards, which is especially valuable for regulatory agencies and businesses prioritizing data ethics and compliance.

Another unique advantage of open-source LLM architecture is its impetus for innovation. Unlike their proprietary counterparts, where license terms often restrict customization and require additional time and costs, open-source models are inherently flexible. Enterprises and individual developers can fine-tune these models, add unique features, or train them on specialized datasets, leading to tailored solutions for various applications.

In addition, this architecture thrives on community contributions, marking a shift from centralized development to a more collaborative and dynamic ecosystem. Enterprises adopting open-source LLMs can benefit from many community-contributed enhancements, maintaining a competitive technological edge. These community contributions range from algorithmic optimizations and efficiency improvements to adding new functionalities like better translation algorithms or sentiment analysis modules. The decentralized nature of development ensures a responsive and continuously evolving model architecture, adapting to the rapidly changing landscape of AI and deep learning.

Companies and organizations like Hugging Face and EleutherAI have exemplified the strengths of the open-source approach in this realm. For instance, Hugging Face's Transformers library has become an indispensable resource for professionals and researchers in natural language processing. Its openness has created a dynamic community that continuously refines the models, adding features and troubleshooting issues. Similarly, EleutherAI's GPT-Neo stands as a testament to the democratizing potential of open-source AI. The model has afforded high visibility and ethical governance and catalyzed collective contributions to its continual improvement.

These initiatives demonstrate that the open-source approach can effectively address some of the drawbacks of proprietary models, particularly regarding transparency, ethical considerations, and collaborative advancements. By fostering an inclusive community of varied contributors, these cases suggest that open-source LLMs can be both innovative and equitable, thereby playing a vital role in steering the AI community towards a more balanced and responsible trajectory.

The Role of Open-Source LLMs in Digital Transformation

Open-source LLMs can disrupt and redefine established operational paradigms across many businesses and sectors. Central to this transformative power is the diversity of applications they enable, ranging from text generation to virtual tutoring and software development assistance. In the following section, we will examine how open-source LLMs contribute to digital transformation in three key areas: organizational language tasks, virtual education, and software development.

Open-source LLMs offer a compelling proposition for organizations keen on leveraging language generation capabilities. Take, for example, Falcon-40B, an LLM that operates under the liberal Apache 2.0 license. The ability to produce high-quality text in response to a given prompt offers an exceedingly vital tool for automating many activities, including writing emails, creating blog posts, and generating creative narratives. Doing so alleviates the manual labor associated with these tasks, rendering the workflow more efficient. Falcon-40B is just one instance among a growing repository of open-source LLMs that organizations can freely adopt and adapt according to their unique operational needs.

Online teaching is a growing domain with substantial influence from open-source LLMs. These models can be adjusted to provide customized educational experiences that follow the individuals' unique learning styles and requirements. By embracing customization, educational institutions and edtech businesses have the potential to enhance the efficacy of learning modules significantly. Additionally, the collaborative potential facilitated by the open-source nature of these tutoring models enables educators and developers to work together, potentially leading to improvements in the overall curriculum quality.

In the software development landscape, LLMs trained in codebases and programming languages have demonstrated proficiency in aiding developers. These models can auto-generate code snippets, identify programming errors, and even flag security-related issues. This can accelerate the development process and bring an additional layer of quality control, essential for creating robust applications. Unlike proprietary alternatives that might limit customization and inspection, open-source LLMs offer a democratized platform for enhancing coding efficiency and reliability.

Having discussed the practical applications of open-source LLMs, let's focus on how these models perform compared to their proprietary counterparts, using benchmarks and case studies as our lens. The open-source ecosystem has fostered a rich landscape of high-performing LLMs, such as Llama-2 by Meta AI and MPT-30B from MosaicML, which not only match but often surpass the capabilities of their proprietary counterparts. This performance is gauged through metrics and benchmarks, meticulously tracked by platforms like the Open LLM Leaderboard. These models are part of a larger, dynamic ecosystem fed by community contributions, ensuring their adaptability and responsiveness to the rapidly evolving landscape of AI and deep learning.

A notable experiment conducted by Waleed Kadous from Anyscale is an informative case study comparing the efficacy of proprietary and open-source Large Language Models (LLMs). The experiment focused on measuring the factual accuracy of these models in summarizing news articles, utilizing a dataset of 373 hand-labeled news report statements for evaluation. Proprietary models like GPT-4 and GPT-3.5-turbo were compared against Llama-2. The study revealed that Llama-2's 70b version achieved a factual accuracy rate of 81.7%, closely following GPT-4's 85.5% but significantly better than GPT-3.5-turbo with its 67%. This suggests that the gap in quality between proprietary and open-source models is gradually narrowing.

Additionally, Llama-2 showcased remarkable cost-effectiveness, being 30 times cheaper than GPT-4 for the nearly equivalent levels of factual accuracy. Despite this, GPT-4 is still better at following instructions, showcasing that there's still room for improvement. Nevertheless, the open-source nature of Llama-2 offers unparalleled flexibility and visibility, allowing for model fine-tuning and avoiding vendor lock-in.

Open-source LLMs are at the forefront of digital transformation, serving as flexible, transparent, and powerful tools for various applications across sectors. Their utility spans from enhancing productivity in mundane tasks to facilitating breakthroughs in complex problem-solving. By aligning their capabilities with the essence of open-source development, they democratize access to advanced AI solutions and catalyze innovation, fostering an environment conducive to rapid, ethical, and efficient digital transformation.

Challenges and Limitations of Open-Source LLMs

This chapter aims to dissect the intricate ethical and operational challenges of open-source LLMs, like the risks of privacy infringement, "hallucinations," bias, issues surrounding consent, and security vulnerabilities, while highlighting the community-driven mechanisms for mitigation.

When a project is open to contributions from multiple entities, ensuring the privacy and integrity of the data used in training the models becomes increasingly complex. For example, when training data comes from various sources like public forums, government databases, and private contributions, tracking and securing all data points is increasingly difficult. This decentralized structure also introduces variability in performance and quality standards, making uniformity a pressing issue.

Nevertheless, these challenges are manageable through joint community effort. By setting up stringent data-handling protocols, explicit contribution guidelines, and rigorous review processes, open-source projects can strive for ethical development while sustaining high-quality benchmarks.

The presence of "hallucinations" in LLM outputs is another cause for concern, mainly because of their ability to generate coherent and fluent writing. The hallucinations produced by the LLM are simply deceptive manifestations that arise from inadequate, conflicting, or deceptive information on which the model was trained. Still, the intrinsic openness of the model serves as a protective measure by enabling a wider community to thoroughly examine, verify, and enhance its validity. This capability has the potential to identify and rectify flaws compared to a closed, proprietary system more effectively.

Bias represents another glaring limitation. For example, an LLM could be tailored to produce text that inadvertently perpetuates gender stereotypes or racial prejudices, thus raising both ethical and social challenges (Zhao et al., 2017). On the other hand, the collaborative nature of open-source environments often promotes a diversity of perspectives, which can serve as an effective countermeasure against inherent biases. By pooling the collective intelligence of a varied user base, open-source LLMs can be refined to be more equitable and less skewed.

In addition, the issue of consent in the training data for open-source LLMs is closely connected to broader debates about data governance and ethics in AI. Because open-source models can be adapted for diverse applications, ensuring that the original training data complies with existing laws and regulations becomes paramount. The notion of "accountability" in data gathering raises essential questions about data ownership, especially when dealing with sensitive sectors like healthcare or finance. Moreover, incorporating feedback mechanisms within the open-source framework can be a practical measure to ensure continuous alignment with ethical standards (Mittelstadt et al., 2016).

Finally, the security aspect brings in a unique set of challenges. The openness that defines open-source LLMs can also make them vulnerable to exploitation. Cybercriminals can employ these models for malicious tasks. Additionally, there is a risk of leaking personally identifiable information (PII), especially when the LLM is fine-tuned on databases that might contain such data (Carlini et al., 2018). Conversely, the community can often detect security vulnerabilities that are patched more quickly, thanks to the significant number of contributors analyzing the code. This "many eyes" approach can result in more robust and secure LLMs in the long term.

In summary, while open-source LLMs share certain challenges with proprietary models, they also offer unique opportunities for mitigation through community analysis, rapid iteration, and ethical transparency. The open-source paradigm thus presents a complex yet promising landscape for the development and governance of LLMs.

Conclusion

In exploring open-source Large Language Models (LLMs), we have covered a diverse and evolving landscape that stretches from opportunities and practical applications to ethical and operational challenges. These models, underpinned by a rich ecosystem and community contributions, not only compete with their proprietary counterparts but, in many instances, exceed them in performance and cost-effectiveness. However, this innovation comes with its own hurdles. The ethical considerations of bias, privacy, data consent, and operational challenges like hallucinations and security vulnerabilities

are ever more present. Yet, the open-source model's inherent transparency and collaborative nature provide unique opportunities for addressing such issues.

As the role of LLMs continues to expand across sectors and applications, the open-source community is uniquely positioned to shape this evolution to prioritize ethical governance, community collaboration, and technological advancement. Researchers should continue to study and benchmark open-source LLMs to better understand their long-term impact. Developers and businesses should consider incorporating these models thoughtfully, aligning their capabilities with moral and practical requirements. Policymakers should craft regulations that balance innovation with ethical considerations. Finally, the general public should remain critical and optimistic, actively engaging with these technologies to shape their development. As we stand on the cusp of this technological frontier, it is crucial to leverage the strengths of the open-source model to guide the responsible and innovative use of LLMs in our digital future.

References

1. Sören, Sonnenburg, et al. (2007). The Need for Open Source Software in Machine Learning. *Journal of Machine Learning Research*, doi: 10.5555/1314498.1314577
2. Max, Langenkamp., Daniel, N., Yue. (2022). How Open Source Machine Learning Software Shapes AI. doi: 10.1145/3514094.3534167
3. Open source large language models: Benefits, risks and types. (2023, September 27). <https://www.ibm.com/blog/open-source-large-language-models-benefits-risks-and-types/>
4. W. Kadous (2023), Llama 2 is about as factually accurate as GPT-4 for summaries and is 30X cheaper. Retrieved from <https://www.anyscale.com/blog/llama-2-is-about-as-factually-accurate-as-gpt-4-for-summaries-and-is-30x-cheaper>
5. H. Jama, The Rise of Open Source LLMs. May 6, 2023, from <https://medium.com/@hamsa.a.j/the-rise-of-open-source-llms-bdf566393107>
6. Zhao, J., et al. (2017). Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
7. Mittelstadt, B. D., et al. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 2053951716679679.
8. Carlini, N., et al. (2018). The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks. In *28th USENIX Security Symposium (USENIX Security 19)*.

TOWARDS A UNIFIED CONCEPT FOR SHARING AND LINKING OF DATA RESOURCES IN COLLABORATIVE RESEARCH VIA WEB SERVICES

Plamen Milev

Associate Professor, PhD. Department of Information Technologies and Communications, University of National and World Economy, e-mail: pmilev@unwe.bg

Abstract

In an era where collaborative research is increasingly becoming the norm, the necessity for efficient and effective data sharing and linking mechanisms is paramount. This paper delves into the journey towards a unified concept for sharing and linking of data resources, emphasizing the pivotal role of web services in realizing this vision. From the challenges posed by diverse data formats and the need for interoperability to the potential for real-time collaboration, the research underscores the transformative potential of web services. Drawing on protocols such as OAI-PMH and the GO-FAIR initiative's principles, the study outlines the integral function of web services in adhering to standards of findability, accessibility, interoperability, and reusability (FAIR). Furthermore, platforms such as Zenodo, OpenAIRE, EOSC, and BPOS are discussed in light of their contributions to the open science movement. The research bridges the divide between the theoretical need for a unified data sharing concept and the practical technological solutions offered by web services, paving the way for a more integrated and collaborative future in scientific research.

Key words: Data sharing, data linking, web services, open science, collaborative research.

JEL: C88, L86.

Introduction

The modern scientific landscape thrives on collaboration. With researchers across the globe working together, sharing data efficiently and effectively has become crucial. This collaborative approach produces vast amounts of data from various sources, each with its unique format and structure. Navigating this complex environment without a consistent method for sharing and linking data can slow down progress, as researchers spend more time managing data than actually analyzing it. Web services offer a promising solution to this challenge. They are tools designed to operate over the internet, allowing different software applications to communicate seamlessly. For researchers, this means easier access to data, consistent communication between different research tools, and the potential for real-time collaboration. More than just a technical fix, web services could be the key to unlocking the full potential of collaborative research by streamlining data sharing and integration. It's in this context that the importance of a unified approach to data sharing emerges. Without standardized mechanisms to share and interlink datasets, researchers may find themselves mired in administrative tasks, struggling with data compatibility issues, or confronting barriers to accessing crucial datasets. These logistical hurdles can hinder the rapid exchange of insights, dampening the innovative spirit that collaborative research seeks to foster. Web services are designed to foster communication between different software applications, regardless of their underlying architecture or platform. In the research ecosystem, web services can act as vital bridges, connecting disparate data sources and ensuring that they speak a common language. As we delve deeper into this topic, we will explore the transformative potential of web services in the context of collaborative research. We will examine how they can address current challenges, highlight the platforms that exemplify their successful integration, and envision a future where collaborative research is truly synergized through the capabilities of web services.

This paper ventures into the realm of web services, aiming to highlight their instrumental role in forging a unified concept for sharing and linking data resources in collaborative research. We shall embark on a journey that traverses the challenges of the current research landscape, the transformative potential of web services, and the practical implications of integrating these services into the broader framework of scientific collaboration. Drawing from established protocols, open science platforms, and forward-thinking initiatives, this research seeks to pave the way for a future where collaborative research is not only facilitated but truly optimized through the power of web services.

Landscape of sharing and linking of data resources

In the contemporary digital era, the ability to share and link data resources effectively is pivotal in facilitating collaborative research efforts. This paper seeks to delineate the prevailing methodologies and frameworks that underscore data sharing and linking in collaborative research, exploring their evolution, advantages, and limitations. The importance of collaborative research cannot be understated in the modern academic and scientific landscape. Collaborative research, underpinned by the efficient sharing and linking of data resources, paves the way for multidisciplinary discoveries and innovations [1]. As research projects become increasingly complex and data-driven, the necessity for robust methodologies and frameworks for data sharing and linking becomes evident. This paper aims to scrutinize the current landscape of these methodologies and frameworks, providing a comprehensive overview of their characteristics and functionalities. In the nascent stages of collaborative research, data sharing was predominantly centralized, with data repositories being managed by a single entity. This approach, however, gradually gave way to decentralized and hybrid models as the volume of data increased and the necessity for multi-institutional collaboration became apparent [2]. Over the years, various frameworks have been developed to facilitate data sharing and linking, often characterized by evolving technological and policy considerations:

- Centralized approaches to data sharing and linking involve the consolidation of data resources into a single platform or repository. This approach ensures uniform data formats and standards, which can potentially simplify data management and analysis processes. However, it can also pose challenges in terms of data security and privacy [1].
- On the other hand, decentralized approaches allow for data resources to be stored and managed across various platforms and entities. While this approach fosters data ownership and autonomy, it can create hurdles in achieving data interoperability and integration [3].
- Hybrid approaches attempt to amalgamate the advantages of both centralized and decentralized approaches, fostering data sharing and linking while mitigating respective disadvantages. These approaches are often characterized by sophisticated technological infrastructures that facilitate data integration and interoperability [1].

In terms of frameworks, a considerable focus has been placed on both technical and policy aspects:

- From a technical standpoint, considerations such as data formats, integration, and interoperability are pivotal. Emerging frameworks are increasingly leveraging technologies such as blockchain to facilitate secure and efficient data sharing and linking [4].
- In parallel, policy and governance aspects have garnered attention, with emphasis on data privacy, security, and ownership. Various frameworks have emerged to navigate the complex landscape of data governance, focusing on fostering collaboration while safeguarding data rights and privacy [5].

The landscape of data sharing and linking in collaborative research is characterized by a rich tapestry of methodologies and frameworks. As research endeavors continue to evolve, the development of robust, secure, and efficient frameworks for data sharing and linking will remain a focal point. This analysis underscores the importance of ongoing efforts to refine and optimize these frameworks, fostering a collaborative research environment that is both innovative and secure. Data sharing and linking are central processes in collaborative research, aiding in the development of a cohesive and comprehensive research landscape. Despite the advancements, several challenges persist, hindering the

full realization of a collaborative research ecosystem. Concurrently, the evolution of a unified concept for data sharing and linking presents promising opportunities for overcoming these challenges and fostering enhanced collaboration in research. We delineate the prominent challenges and explore the potential opportunities that a unified concept could usher. Data sharing and linking in collaborative research face a myriad of challenges that span across technical, policy, and ethical dimensions. These challenges are elaborated as follows:

- Technical challenges comprise issues related to data interoperability, integration, and standardization. Disparate data formats and inconsistencies in data structures often impede seamless data sharing and linking [3].
- Policy challenges encompass complexities in data governance, including issues surrounding data ownership, privacy, and security. The diverse regulatory landscapes across different regions further complicate policy adherence [5].
- Ethical challenges involve safeguarding the privacy of individuals and ensuring the responsible use of data. Ethical considerations necessitate frameworks that promote transparency and consent in data usage [1].

The conceptualization and implementation of a unified concept for data sharing and linking portend significant opportunities in fostering collaborative research:

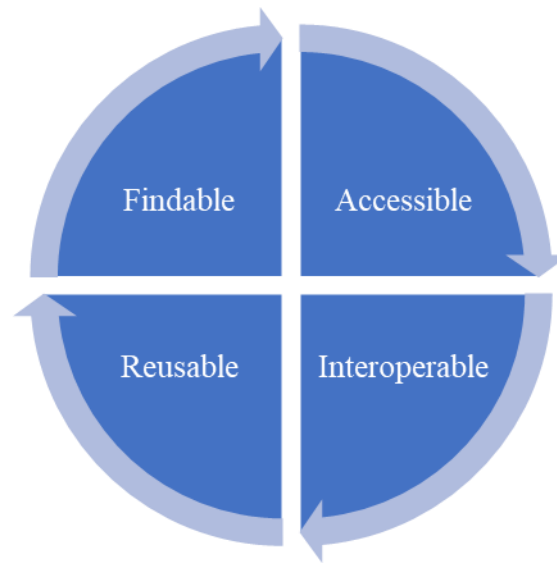
- A unified concept can facilitate greater collaboration by streamlining data sharing processes and fostering inter-disciplinary research endeavors [2].
- Technological advancements such as the integration of blockchain technology can offer secure and efficient mechanisms for data sharing and linking, paving the way for innovative research approaches [4].
- A unified concept can aid in the harmonization of policy frameworks, fostering a cohesive approach to data governance and facilitating compliance with diverse regulatory mandates [5].

As the landscape of collaborative research evolves, addressing existing challenges and leveraging emergent opportunities becomes critical. A unified concept for data sharing and linking stands as an important stage, promising to revolutionize the collaborative research arena by fostering enhanced collaboration, leveraging technological advancements, and harmonizing policy frameworks. The journey towards realizing this unified concept necessitates concerted efforts from all stakeholders, paving the path for a collaborative and innovative research ecosystem.

Sharing data can lead to new insights and discoveries, often beyond the initial scope of the original research. Open datasets can be used by multiple researchers to validate findings, explore new hypotheses, or combine with other data sources for richer analyses. Reducing duplication of efforts by making use of already collected data can save time and resources. This is particularly important in expensive domains, like large-scale experiments or medical research, where reproducing data collection can be cost-prohibitive. Open sharing of data can increase trust in scientific results, as independent parties can validate and reproduce findings. Shared data facilitates collaborations across disciplines, institutions, and even countries. These collaborations can lead to interdisciplinary projects and breakthroughs.

Technological aspects of sharing and linking of data resources

The advent of the digital era has brought forth an unprecedented surge in data generation and consumption. Ensuring the effective management, sharing, and utilization of this vast ocean of data, particularly in research, is paramount. Central to this pursuit are the GO-FAIR principles, which lay out a comprehensive roadmap for outlining fundamental principles that aim to bolster the discoverability and utility of data [3]. Specifically, the principles highlight the importance of making data findable, accessible, interoperable, and reusable – often abbreviated as FAIR (Fig. 1) [6].



Source: GO-FAIR Initiative (2023)

Figure 1: FAIR data principles

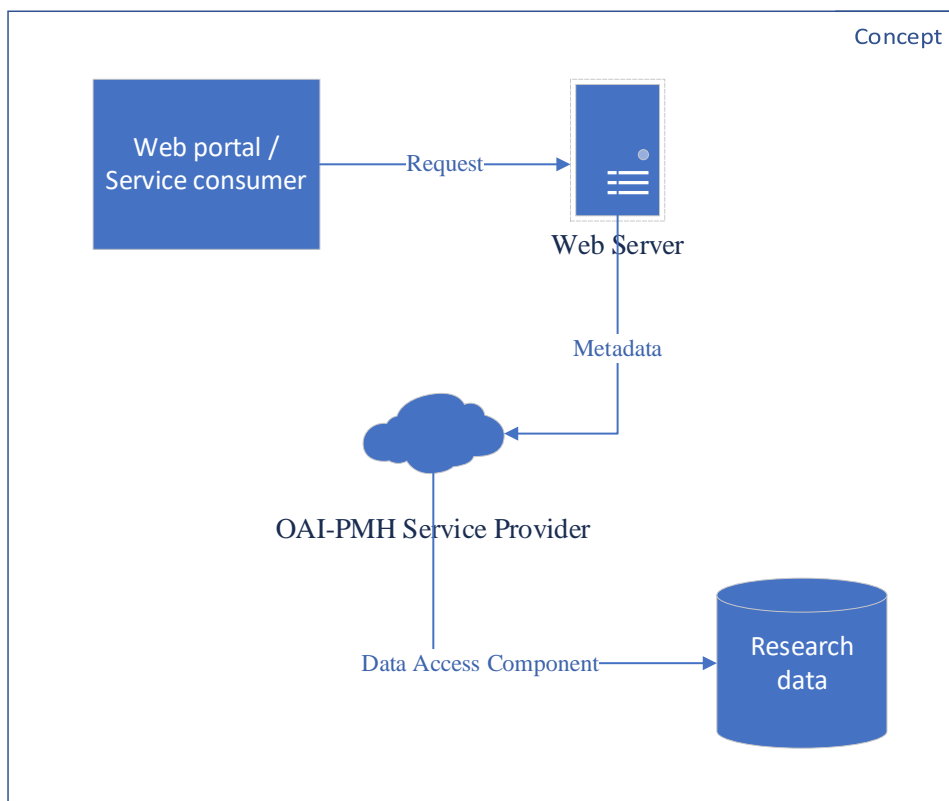
Platforms such as Zenodo, OpenAIRE, EOSC, and BPOS have been at the forefront in actualizing these principles, albeit with tailored approaches suited to their specific user demographics and regional mandates. Positioned as a universal repository, Zenodo is more than just a storage platform [7]. Its significance is accentuated when viewed through the prism of the FAIR principles. Each piece of data or research output uploaded to Zenodo is endowed with a unique Digital Object Identifier (DOI). This not only heightens the findability of the data but also assures its longevity and traceability. Open access, a defining feature of Zenodo, ensures that data remains accessible to researchers globally. Furthermore, its capability to support an array of data formats ensures that data from disparate sources can be integrated, promoting interoperability. By clearly delineating licensing terms, Zenodo also champions the reusability of data, ensuring researchers can build upon previous works with clarity on rights and attributions. Serving as a beacon in the open science movement in Europe, OpenAIRE is a testament to the power of collaboration [8]. Its vast repository, teeming with publications, datasets, and other research products, is a treasure trove for scholars. Advanced search algorithms ensure data findability, while its commitment to open access aligns with accessibility mandates. What sets OpenAIRE apart is its proactive approach to promoting interoperability. By providing tools and guidelines for data harmonization, it ensures that datasets from varied domains can be seamlessly merged. Moreover, by setting standards and offering data management recommendations, OpenAIRE fosters data reusability, ensuring the longevity of research outputs. The European Open Science Cloud (EOSC), an ambitious venture by the European Commission, is a paradigm shift in how research data is viewed and managed [9]. Its mission transcends mere storage. It aims to provide a cohesive environment where data from various research infrastructures across Europe can be harmonized and utilized. By creating a network of interconnected data sources, EOSC amplifies data findability. A robust governance structure, complemented by clear policies, ensures data accessibility while minimizing misuse. Efforts to integrate resources from diverse research domains underscore its commitment to interoperability. Finally, by championing open science and emphasizing proper data attribution, EOSC ensures that data remains reusable. As Bulgaria's foray into the open science arena, BPOS is a reflection of the country's commitment to democratizing research [10]. Its centralized structure enhances data findability, acting as a one-stop portal for Bulgarian research outputs. The commitment to standardizing data formats and aligning with broader European directives ensures interoperability. By providing clear guidelines on data usage and rights, BPOS ensures that research outputs are not only accessible but also primed for reuse. The essence of the GO-FAIR principles is to facilitate a seamless research ecosystem where data is not just a byproduct but a valuable asset. Platforms like Zenodo, OpenAIRE, EOSC, and BPOS, each in its unique way, are emblematic of these principles in action. As research continues to break

boundaries, both geographically and disciplinarily, such platforms will be instrumental in ensuring that data remains a unifying force, driving collaboration, innovation, and discovery.

The technological landscape has seen rapid developments that hold the potential to revolutionize the domain of data sharing and linking in collaborative research. The advent of sophisticated technologies could pave the way for a unified concept that fosters efficiency, security, and interoperability in data sharing and linking. The integration of Artificial Intelligence (AI) and Machine Learning (ML) in data sharing and linking has the potential to revolutionize data analysis and interpretation. AI and ML can facilitate the automation of data processing, helping in the identification of patterns and trends, which can significantly enhance collaborative research efforts [11]. Cloud computing offers a scalable and flexible platform for data storage and sharing. It enables real-time access to data from anywhere, fostering collaboration and facilitating seamless data integration across various research groups. Furthermore, cloud computing can aid in efficient data management through services like Data as a Service (DaaS), which can streamline data access and usage [12]. The Internet of Things (IoT) can be leveraged to enhance data collection and sharing in research. Big Data and IoT are one of the most promising technologies in our time [13]. IoT is a modern paradigm that describes the connection of almost everything to the Internet [14]. IoT devices can generate a vast amount of real-time data, which can be integrated into research databases for comprehensive analyses. This technology fosters connectivity and data exchange, opening new avenues for collaborative research [15]. The amalgamation of the IoT revolution with platforms upholding the GO-FAIR principles is a testament to the evolving nature of research.

Concept of sharing and linking of data resources via web services

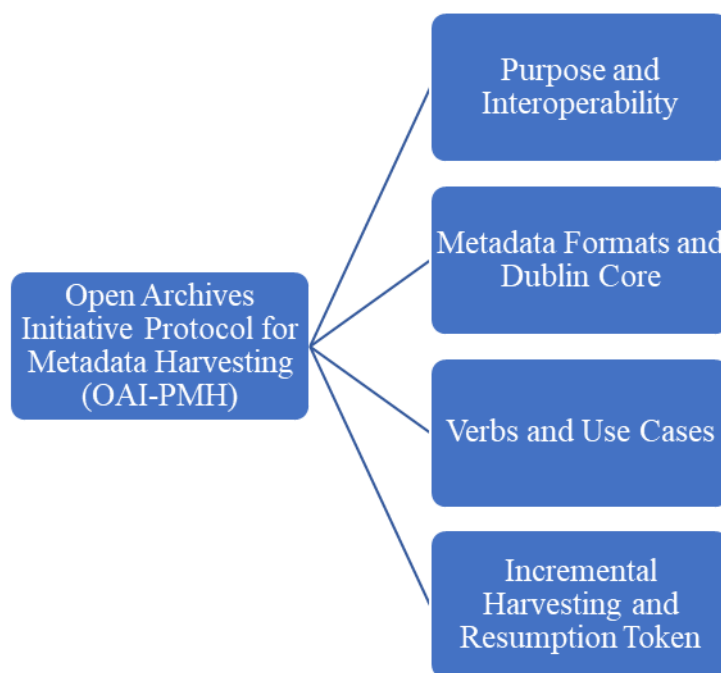
Technological advancements stand as a cornerstone in the evolution of a unified concept for data sharing and linking. Concept development is a prior process to the technological implementation of a corresponding prototype solution [16]. Sharing and linking of data resources in a unified way also requires the corresponding reorganization of business process management in organizations in the context of their digital transformation [17]. In the context of the present study, we will propose a concept for sharing and linking of data resources within the alliance of leading European universities in business, economics, and social sciences with rich experience in analyzing societal change, known as ENGAGE.EU (Fig. 2).



Source: Author

Figure 2: Conceptual approach for sharing and linking of data resources via web services

The objective of the concept is to show and share data resources on topics of societal change. We collect data of the repositories and existing data centers of different partner universities. Sharing and linking data resources will foster open science strategies in research [18]. For the purposes of sharing and linking of data resources, the OAI-PMH protocol is used. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a protocol developed for harvesting (or collecting) metadata descriptions of records in archives. It allows automated agents to access metadata from repositories that support the protocol (Fig. 3) [19].

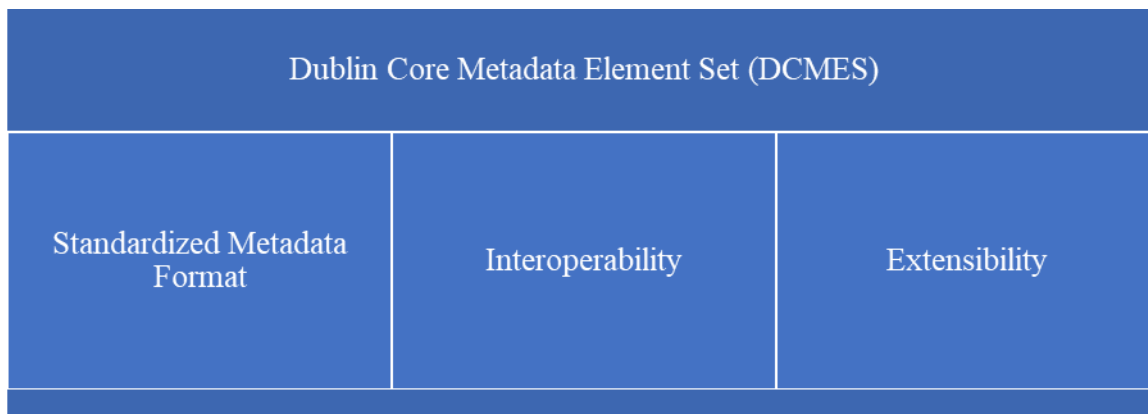


Source: Open Archives Initiative Protocol for Metadata Harvesting (2023)

Figure 3: Key points about OAI-PMH

Its primary goal is to facilitate the efficient dissemination of content. OAI-PMH allows for the exposure and sharing of metadata, making it easier for search engines or other services to find and access content from various repositories. One of the significant advantages of OAI-PMH is that it promotes interoperability. Repositories can expose their metadata in a standard format, allowing service providers to aggregate this metadata from various sources easily. While OAI-PMH can support multiple metadata formats, the only format that repositories are required to support is Dublin Core [20], a simple and standardized metadata schema. However, repositories can also offer their metadata in other formats. OAI-PMH operations are based on the use of verbs. These verbs allow users to get information about the repository. The protocol is used in various applications, especially in the realm of digital libraries, institutional repositories, and scholarly communication. OAI-PMH supports incremental harvesting, allowing service providers to request only those records that have been added or changed since a previous harvest. For situations where the response to a request is too large to be returned in one go, OAI-PMH uses the concept of a resumption token. This token can be used to continue a request from where it left off. The connection between OAI-PMH and Dublin Core can be understood through their intertwined roles in enhancing the accessibility and interoperability of digital resources (Fig. 4).

For repositories just starting with metadata exposure via OAI-PMH, the Dublin Core's simplicity offers a straightforward path. They can begin with basic metadata descriptions and later opt for more detailed or specialized schemas as needed. Given the widespread adoption of both Dublin Core and OAI-PMH in the digital library and repository communities, their combined use has become something of a de facto standard. This widespread acceptance ensures that tools, best practices, and community support are readily available. When repositories expose their metadata via OAI-PMH using Dublin Core, they ensure that their digital resources can be more easily discovered and accessed by a global audience. Aggregators, search engines, and other digital services can harvest this metadata, making the content of these repositories more widely known and accessible.



Source: Dublin Core (2023)

Figure 4: Connection of DCMES to OAI-PMH

At its core, OAI-PMH is a protocol developed for collecting (or "harvesting") metadata descriptions from repositories. It is based on the widespread use of web standards, particularly HTTP and XML. OAI-PMH itself is a web service, allowing different software on different servers to interact with each other over the internet. It uses HTTP for the transport of requests and responses, making it accessible and straightforward to integrate with various platforms. One of the foundational principles of web services is the standardization of requests and responses to ensure smooth communication between different software applications. OAI-PMH adopts this principle by defining specific "verbs" (like GetRecord, ListIdentifiers, etc.) that dictate the type of requests and expected responses. This ensures that any OAI-PMH compliant harvester knows exactly how to request data and what kind of response to expect, regardless of the specific repository it's communicating with. Web services need a standardized way to represent data that both the requester and the responder can understand. OAI-PMH uses XML for this purpose, a common choice for web services. XML allows for structured representation of metadata, which can be easily parsed and processed by different software applications. One of the main reasons for using web services is to achieve interoperability between different software systems, often running on different platforms and written in different programming languages. OAI-PMH is designed with this in mind. By adhering to the OAI-PMH protocol, repositories ensure that their metadata can be harvested and utilized by a diverse range of systems. OAI-PMH is flexible regarding the metadata formats it supports. While it mandates the support for a simple Dublin Core format, repositories can offer their metadata in multiple formats. This extensibility is a key feature of many web services, ensuring they can adapt to evolving requirements without overhauling the entire system. Web services are often used to automate tasks that would be tedious or impractical for humans to perform manually. In the context of OAI-PMH, automation is evident in how harvesters can periodically query repositories for new or updated records, ensuring metadata collections are up-to-date. This automated process can scale to accommodate large repositories or frequent updates. The proposed concept assumes the presence of an intermediate layer between the web server and data repositories to implement the OAI-PMH protocol. The use of an additional layer is common for concepts that include a metadata model [21]. Within concepts that involve data, it is a common principle to treat data as a very valuable asset and to occupy a central place in the respective concept [22]. In the context of sharing and linking of data resources, data should be available in real time from any point in the world, which is also considered one of the main reasons for the emergence of new directions in databases [23].

The principles that underpin OAI-PMH – standardization, interoperability, and efficient data dissemination – align closely with the broader goals of creating a unified concept for data sharing and linking. By leveraging such protocols, the research community can move towards a more collaborative and integrated environment where data resources from various sources can be effortlessly shared, linked, and utilized. The connection between OAI-PMH and Dublin Core represents a synergy between a protocol designed for efficient metadata harvesting and a standardized, simple metadata schema. Their combined use has greatly facilitated the growth and success of digital repositories, ensuring that digital

resources are both accessible and discoverable on a global scale. OAI-PMH is a shining example of web services applied to the domain of metadata harvesting. Its design principles, rooted in web standards, facilitate easy and standardized access to metadata across diverse repositories. This ensures that metadata from different sources can be aggregated, searched, and utilized efficiently, promoting the broader goals of open access and interoperability in the digital world.

Conclusion

In an increasingly digitalized research ecosystem, the imperative for effective data sharing and linking has never been more pronounced. The evolution of the research domain, with advancements like IoT, reinforces the necessity for protocols and platforms that facilitate seamless, efficient, and interoperable data management. Drawing from established concepts like OAI-PMH and the principles of web services, we glean insights into how such mechanisms can be developed, standardized, and universally adopted. Central platforms like Zenodo, OpenAIRE, EOSC, and BPOS, while varying in their scope and functionality, share a common goal – promoting the open sharing of data and research outputs. Their alignment with the GO-FAIR principles ensures not just the availability of data but its accessibility, interoperability, and reusability. The application of these platforms, bolstered by national initiatives like BPOS in Bulgaria, underscores the global push towards open science and research transparency. The integration of web services, especially in the context of OAI-PMH, exemplifies how technology can be harnessed to address the challenges of data sharing. The future of research collaboration will inevitably rest on the confluence of technology and policy, ensuring that data isn't just shared, but shared with purpose, clarity, and a commitment to advancing the collective knowledge of the global research community. As we look forward, it's evident that a unified concept for data sharing and linking, underpinned by robust technological infrastructures and harmonized policies, will be pivotal.

Acknowledgement

This work was supported by the Erasmus+ project “The European University engaged in societal change – ENGACE.EU” (Research Grant No. 101004071).

References

1. Borgman, C. L. (2017). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.
2. Kahn, R., & Wilensky, R. (2006). A framework for distributed digital object services. *International Journal on Digital Libraries*, 6(2), 115-123.
3. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.
4. Zyskind, G., Nathan, O., & Pentland, A. (2015). Decentralizing privacy: Using blockchain to protect personal data. In *2015 IEEE Security and Privacy Workshops* (pp. 180-184). IEEE.
5. European Commission. (2018). *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data*.
6. GO-FAIR Initiative. (2023). Retrieved from GO-FAIR Website: <https://www.go-fair.org>
7. Zenodo. (2023). Retrieved from Zenodo Website: <https://zenodo.org/>
8. OpenAIRE. (2023). Retrieved from OpenAIRE Website: <https://www.openaire.eu/>
9. European Open Science Cloud. (2023). Retrieved from EOSC Website: <https://eosc-portal.eu/>
10. Bulgarian Portal for Open Science. (2023). Retrieved from BPOS Website: <https://bpos.bg/>
11. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
12. Mell, P., & Grance, T. (2011). *The NIST definition of cloud computing*.

13. Stefanov, G. (2019). Analysis of Cloud based ETL in the Era of IoT and Big Data. In *Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education*. 198-202. ICAICTSEE.
14. Boyanov, L. (2021). A Conceptual Approach for Industrial Internet of Things Assessment. *Godishnik na UNSS*, (2), 97-107.
15. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645-1660.
16. Tabov, Y. (2019). Application of Concept for Structuring of Web Data. In *Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education*. 266-270. ICAICTSEE.
17. Belev, I. (2018). Software Business Process Management Approaches for Digital Transformation. *Godishnik na UNSS*, (1), 109-119.
18. ENGAGE.EU. (2023). The European University Engaged in Societal Change. Retrieved from ENGAGE.EU Website: <https://www.engageuniversity.eu/>
19. Open Archives Initiative Protocol for Metadata Harvesting. (2023). Retrieved from OAI-PMH Website: <https://www.openarchives.org/pmh/>
20. Dublin Core. (2023). Retrieved from Dublin Core Metadata Element Set (DCMES) Website: <https://www.dublincore.org/>
21. Mihova, V. (2015). Methods of Using Business Intelligence Technologies for Dynamic Database Performance Administration. *Economic Alternatives*, (3), 105-116.
22. Delchev, D., & Lazarova, V. (2021). Big Data Analysis Architecture. *Economic Alternatives*, (2), 315-328.
23. Radoev, M. (2021). Contemporary Trends in the Development of Databases. *Ikonomiceski i Sotsialni Alternativi*, (1), 5-15.

ТЕХНОЛОГИЧНИ РЕШЕНИЯ И ДИГИТАЛИЗАЦИЯ В СВАТБЕНИЯ ТУРИЗЪМ

Симона Маджарова

Докторант, катедра „Икономика на туризма“, УНСС, e-mail:simona.madzhurova@abv.bg

Резюме

Научният доклад има за цел да анализира как технологиите променят сватбения туризъм и кои са най-важните тенденции в отрасъла. Вниманието е фокусирано върху технологичните решения в процеса на организиране на сватбата и медения месец. Изследователският интерес е насочен към дигитализацията и предизвикателствата, които тя поставя пред сватбения сектор. Изтъкнато е значението на нуждите на сватбената индустрия, които подтикват въвеждането на нови подходи. Анализирани са бариерите за сватбения туризъм след пандемията от Covid-19. Специално място е отделено на генерирането на нови технологични продукти за удовлетворяване на клиентите. Авторът защитава тезата, че технологичните решения и дигитализацията в сватбения туризъм са основа за успешното развитие на конкурентен пазар за сватби и медени месеци.

Ключови думи: сватбен туризъм, технологии, дигитализация, иновации

JEL: O31, O39

TECHNOLOGICAL SOLUTIONS AND DIGITALIZATION IN WEDDING TOURISM

Simona Madzharova

Abstract

The research report aims to analyze how technology is changing wedding tourism and what are the most important trends in the industry. Attention is focused on technological solutions in the process of organizing the wedding and honeymoon. The research interest is focused on digitalization and the challenges it poses to the wedding sector. The importance of the needs of the wedding industry, which prompts the introduction of new approaches, is highlighted. The trends for wedding tourism after the Covid-19 pandemic have been analysed. A special place is devoted to the generation of new technological products to satisfy customers. The author defends the thesis that technological solutions and digitization in wedding tourism are the basis for the successful development of a competitive market for weddings and honeymoons.

Key words: wedding tourism, technology, digitization, innovation

Туризмът, сватбите и технологиите

Туризмът е един от най-важните сектори на световната икономика, генериращ заетост и средства във всеки край на планетата. През последните години той претърпя голяма трансформация благодарение на технологиите, които промениха начина, по който туристите планират, наслаждават се и споделят своите пътувания.

Сватбата е един от най-важните и незабравими дни в живота на всеки човек. Това е ден, за който мечтаем от малки и си представяме в най-дребния детайл. В тази връзка, отнасяйки туризма към сватбите и говорейки за сватбен туризъм, обикновено клиентите инвестират много средства и винаги очакват резултата да бъде перфектен. Технологиите са ключов елемент в тази връзка,

защото минимизират човешките грешки и улесняват осъществяването на цялостния процес по организация и изпълнение на сватбено тържество. Развитието на туризма разчита на набор от доказани техники и стратегии за неговото популяризиране (Парушева, 2017, с. 16).

Технологичните решения променят начина, по който туристите планират своето сватбено пътуване или организация на тържество. Засилва се тенденцията клиентите да оценяват доброто обслужване според наличието на технологии в конкретния туристически обект. (Гоцева, 2021, с. 601). Днес е възможно да бъдат резервирани украса, изпълнение, настаняване, полети, екскурзии и всичко нужно – изцяло онлайн. Това може да се случи от всяка точка на света и по всяко време на денонощието. В допълнение има множество приложения, сайтове, социални платформи и форуми, които помагат на младоженците да намерят най-добрата сделка, въз основа на техните предпочитания и бюджет. Благодарение на дигитализацията и технологиите, туристите могат да сравняват цени, да четат коментари и оценки от други потребители. По този начин съставят плана, който най-добре отговаря на техните нужди. Този тип технология позволява на потребителите да виждат в реално време всички удобства, които се предлагат. Освен това, дава възможност на клиентите да се доближат възможно най-близо до предстоящото изживяване (Гоцева, 2020, с. 155).

Веднъж достигнали до момента на сватбата или мястото за меден месец, технологиите продължават да бъдат основен инструмент за клиентите на сватбения сектор. Туристическата индустрия е силно повлияна от различни световни тенденции (Parusheva, 2017, p. 268). Социалните платформи за комуникация непрестанно обновяват и променят своето съдържание. Социалните платформи като Facebook, Instagram, Tik-Tok, позволяват споделяне в реално време на сватбеното тържество или пътуване. Мобилните приложения за навигация, позволяват на туристите да се ориентират в непознато място и да намерят най-интересните туристически атракции. Освен това има приложения, които предлагат информация за историята, културата и гастрономията на всяко място, което им позволява да научат повече за местната култура. Има и такива, които позволяват резервации в последния момент, в случаите когато просто не е останало време покрай цялата суматоха на празника. Благодарение на това ново измерение, се появяват напълно нови сектори на икономиката, свързани с поддръжката и развитието на различни платформи (Рангелов, 2022, с.146).

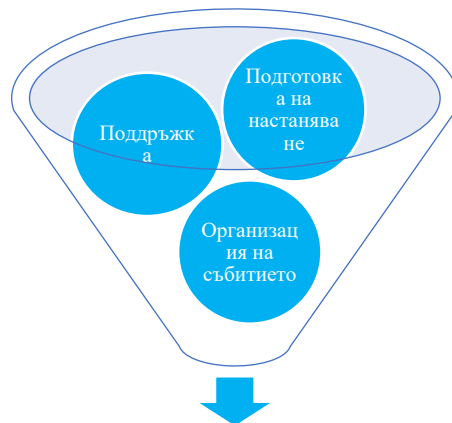
Друга от най-важните нововъведения в туризма, е виртуалната и добавената реалност. Тази технология позволява на бъдещи младоженци да преживеят дестинация, преди да пътуват до там, което може да им помогне да решат дали това е мястото, което искат да посетят. Виртуалната реалност може да бъде и маркетингов инструмент за туристически дестинации. Тя позволява да се покаже мястото в най-добрия му вид и да се даде представа за изживяването, което туриста очаква. Добавената реалност, от своя страна, позволява взаимодействието с околната среда по по-богат и по-пълноценен начин. Всичко споменато е с добавена стойност за сватбения сектор, тъй като допринася за висококачествено изпълнение на продуктите и услугите, които се генерират.

Технологиите променят благоприятно сватбения туризъм, предлагайки на клиентите инструменти и ресурси, за да планират по-добре и да се наслаждават на своите пътувания. В допълнение, превръщат сватбените услуги в по-достъпни, по-евтини и по-персонализирани, което води до увеличаване на броя на туристите и генериране на нови бизнес възможности.

Пример за технологични решения в сватбения туризъм

Съзнавайки важността на технологиите, испански хотел, насочен основно към организация на сватбени тържества са решили да заложат на иновациите, така че нито един детайл да не бъде пренебрегнат при организирането на толкова сложно събитие. Благодарение на опита от повече от 20 години и изключителните си съоръжения, имение в Толедо превръща всяко събитие в незабравимо преживяване. То е разположено на 5 хектара, като има различна природна среда, бутиков хотел и 4 вили за провеждане на събития. Екипът се отличава със страстта, която имат в генерирането на среда, базирана на изтънчени декорации и творения. Те създават проект за всяко тържество, влагайки душата си в него (<https://tecnohotelnews.com>).

Екип от повече от 60 души участва в сватбено тържество и трябва да действа с абсолютна координация. Групата от градинари е отговорна за това градините да изглеждат великолепно, монтажният екип трябва да работи усилено, за да гарантира, че мебелите, декорациите и домакинските предмети са разпределени според изискванията на клиента. Освен това, повече от десет готвачи работят няколко дни предварително, така че гастрономическото изживяване да отговаря на останалите усещания, които присъстващите на събитието ще усетят. Фризьори, хотелски персонал, цветари, сладкари, персонал по поддръжката и контролен екип, също имат необходимост от координация за безпроблемно протичане на всяко тържество. За да рационализира управлението на тези проекти с максимална ефективност, фирмата избра да внедри система, която дигитализира всички ежедневни операции: поддръжка, подготовка на настаняване и организация на събитието (вж. Фигура 1).



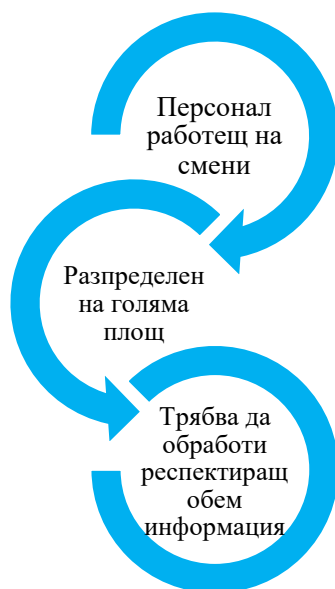
Source: Систематизация на автора по информация от <https://tecnohotelnews.com>

Фигура 1: Операции в управлението на проект

Управлението на такъв голям екип, който се движи през толкова обширно пространство, налага да има комуникационен инструмент, който е гъвкав, интуитивен и лесен за използване от служителите, и който позволява на хора от много различни среди да боравят с него. Тези предизвикателства са оптимално решени чрез внедряването на „hotelkit“. Тази компания е създадена от австрийски хотелиер през 2012 г., за да преодолее вътрешните комуникационни трудности, които съществуват в собствения му хотел. Те са общи за повечето хотели днес: персонал, който обикновено работи на няколко смени, разпределен е на голяма площ и трябва да се справя с респектиращ обем информация (вж. Фигура 2). Благодарение на тази технология, служителите лесно докладват подробности за вършената работа, проверяват списъци с повтарящи се задачи, изпълняват конкретни работни поръчки, консултират се с ръководствата на различните отдели, получават корпоративни новини и известия (<https://tecnohotelnews.com>).

Технологичното решение в този хотел, успява да елиминира хартиения носител за предаване на инструкции и известия между своя екип – независимо къде се намира всеки човек. Персонала използва своя смартфон за всякакви запитвания и комуникация. Освен това се намаля времето за реакция на всяко клиентско търсене и времето за разрешаването на всеки възникнал проблем или въпрос. Подобно на повече от 800 хотела в различни страни по света, които също използват тази система, този в Испания успява да дигитализира своите операции и да предложи отлично обслужване на своите клиенти. Този пример би бил изключително полезен за български хотели, които също са насочени основно в обслужване и осъществяване на сватбени празненства.

Source: Систематизация на автора по информация от <https://tecnohotelnews.com>



Фигура 2: Вътрешни комуникационни трудности с персонала, срещани във всеки хотел

Дигитализацията и предизвикателства, които поставя пред сватбения сектор

Технологиите се превръщат в съюзник на представителите на сватбения сектор, още повече в периода след пандемията. Те имат двойна цел – да възвърнат доверието на потребителите и да постигнат оперативна ефективност. Очертават се различни предизвикателства, пред които са изправени компаниите в сектора в процеса на дигитализация (<https://www.hosteltur.com>).

Това възстановяване на доверието трябва да се извърши от нулата след въздействието, което коронавируса оказва върху туристическия бранш. С дефинирането на нови, по-прозрачни модели за сигурност, позволяващи проследяване по цялата верига със стойност и дистанционни решения във всички процеси. В този смисъл възприемаме възхода на виртуалните асистенти.

Една от промените, които ще се запазят след пандемията, е да се направят публично достойни информационните потоци, които досега бяха за вътрешна употреба. Те имат положителен ефект, именно за да се генерира това доверие в клиента. Освен това сега клиентите са склонни да предоставят по-подробна информация, защото знаят за какво ще бъде използвана. В тази връзка, трябва да се работи в насока да не бъде загубено това доверие. Целта е да се установят трайни взаимоотношения чрез насърчаване на персонализацията, благодарение на автоматизацията и изкуствения интелект. Както и комбиниране на различни източници на данни и извличане на съответната информация от тях.

Дигитализацията е наложена в един от най-важните отрасли на туризма: туристическите агенции и организирането на събития. И с ковид и постковид успяхме да възприемем и открием, че те са съществени и необходими. От една страна, с туристическите агенции – по време на пандемията – те станаха основни при управлението на случаи на пътуване, полети, круизи, хотели и много други. Благодарение на усилията, работата и отдадеността на агентите, тези потребители, които са резервирали чрез агенции, подчертават бързината, ефективността и най-вече добрата работа

на агентите. Има няколко проучвания и статистики, че благодарение на този безупречен мениджмънт търсенето на почивки чрез агенции ще расте. От една страна, очаква се следващите години агенциите да увеличат броя на работниците. От друга страна, частта за управление на събития, благодарение на новите технологии, имаше голяма трансформация офлайн и онлайн и въпреки че се очаква, че офлайн няма да изчезне, онлайн ще остане и ще бъде силно наложен. Тези мрежи служат като платформа за комуникация с туристите и оценка на споделените преживявания между тях (Парушева, 2018, с. 180). Много годишни събития, от хилядите, които се провеждат, ще станат онлайн събития, така че тези които са посветени на бранша, ще трябва да имат повече дигитално обучение. Но други офлайн събития ще продължат и със сигурност с по-голяма сила, пример за тях са именно сватбите. Всички сватби, които не можаха да се състоят, несъмнено ще бъдат отложени напред във времето, така че ще има нужда от повече работници в организацията и сектора на събитията. Затова, проспериращите туристически дестинации в света, разбират необходимостта от ефективно и интелигентно развитие чрез нови технологии (Димитрова, 2021, с. 30).

References:

1. Гоцева, К. (2021). Взаимодействие между новите технологии и опита на гостите в хотелиерската индустрия, *Сборник доклади от юбилейна научна конференция с международно участие „Туризм и свързаност“, по случай 55 години от създаването на специалност „Туризм“*, 30 – 31 октомври, Варна, стр. 596-603.
2. Гоцева, К. (2020). Прогнози за развитие на технологиите в хотелиерството. *Сборник доклади от научна конференция „Туризм, образование, бизнес“*, Колеж по туризм – Бургас, 24 – 25 септември, с. 152-158.
3. Димитрова, Р. (2021). Управление на интелигентни туристически дестинации, *Списание „Пари и култура“*, Vol. 3/2021, Висше училище по застраховане и финанси, стр. 25-33.
4. Парушева, Т. (2018). Устойчиво развитие на културния туризм в дигитална среда. *Сборник доклади от Международна научна конференция „Туризмът и иновациите“*, посветена на 55 години от създаването на Колеж по туризм – Варна, 14 – 15 септември, Издателство „Наука и икономика“, Икономически университет – Варна, стр. 175-183.
5. Parusheva, T. (2017). *Cultural Tourism: Markets and Strategies*, Publishing House “Avangard Prima”, Sofia.
6. Parusheva, T. (2017). Global Trends in Leisure and Cultural Tourism, *Scientific Journal – Economics and Finance, Collection of Scientific Articles – Agenda Publishing House, Coventry, United Kingdom*, pp. 265-270.

ДИГИТАЛНА ТРАНСФОРМАЦИЯ НА СЧЕТОВОДНАТА ОТЧЕТНОСТ В БЮДЖЕТНИТЕ ОРГАНИЗАЦИИ

Митко Радоев

Доцент, доктор, катедра "Информационни технологии и комуникации", УНСС
e-mail: mradoev@unwe.bg

Резюме

Дигитализацията и дигиталната трансформация в почти всички сфери на дейност на субектите от икономиката и публичния сектор са на дневен ред. Има редица изследвания в сферата на дигиталната трансформация на счетоводната отчетност, но те са насочени изключително към бизнеса. Същевременно, организациите в бюджетната сфера имат своята специфика, която трябва да бъде отчетена, за да се осъществи успешно процесът на дигитална трансформация. В настоящия доклад се анализират специфичните особености на счетоводството в бюджетните организации в Република България, както и на самите организации. На тази база се формулират основните направления за дигитална трансформация на бюджетното счетоводство. Представят са някои от проблемите и предизвикателствата пред този процес, както и възможните подходи за решаването им.

Ключови думи: дигитализация, дигитална трансформация, счетоводство, бюджетно счетоводство

JEL: M400, M410, O310

DIGITAL TRANSFORMATION OF ACCOUNTING IN THE BUDGETED ORGANIZATIONS

Mitko Radoev

Abstract

Digitalization and digital transformation in almost all spheres of activity of subjects from the economy and the public sector are on the agenda. There are a number of studies in the field of digital transformation of accounting, but they are focused exclusively on business. At the same time, organizations in the budget sector have their specifics, which must be taken into account in order to successfully implement the digital transformation process. This report analyzes the specific features of accounting in budget organizations in the Republic of Bulgaria, as well as the organizations themselves. On this basis, the main directions for digital transformation of budget accounting are formulated. They present some of the problems and challenges facing this process, as well as possible approaches to solving them.

Keywords: digitalization, digital transformation, accounting, budget accounting

Въведение

Темата за дигитализацията и дигиталната трансформация е изключително актуална в последните години. Почти няма човешка дейност, която да не е засегната от тези промени, задвижвани най-вече от развитието на информационните технологии. Естествено, и счетоводната отчетност няма как да остане встрани от тази всеобща тенденция. Настоящия доклад си поставя следните цели:

- Да обсъди и прецизира използваните понятия, както и основните направления на дигиталната трансформация на счетоводната отчетност;
- Да анализира специфичните особености на счетоводната отчетност в бюджетните организации и да формулира основните насоки за дигитална трансформация на счетоводството в бюджетните организации;
- Да представи проблемите и някои възможни решения за дигиталната трансформация на счетоводната отчетност в бюджетните организации.

Дигитализация и дигитална трансформация на счетоводната отчетност

Дигитализация и дигитална трансформация

Терминът **дигитализация** в повечето случаи се използва в тесен смисъл - като цифровизация (digitization) на данните, съхранявани на аналогови носители, и преобразуването им в цифров вид. Типичен пример за това е сканирането на хартиен документ и записването му във вид на файл в някакъв формат.

Ясно е, че само цифровизацията на данните има ограничена полезност, ако не бъде допълнена от подходяща обработка на тези данни. Понякога терминът дигитализация се използва и в по-широк смисъл, като обхваща не само цифровизацията на данните, но и използването на нови средства и технологии за обработка на цифровите данни.

Пълният ефект от цифровизацията и новите технологии може да се постигне при оптимизиране на бизнес процесите в организацията. Цялостното реструктуриране и оптимизиране на дейностите, базирано на цифровизацията и на използването на подходящи информационни средства и технологии, наричаме **дигитална трансформация**. Успешната дигитална трансформация идва не от въвеждането на нови технологии, а от трансформирането на организацията, за да се възползва от възможностите, които новите технологии предоставят [1].

Насоки за дигитализация и дигитална трансформация на счетоводната отчетност

Различните автори дефинират различни специфични компоненти на дигитализацията и дигиталната трансформация в счетоводството, като например [2]:

1. еднородност на системите, използвани във фирмата;
2. разширяване на интерфейсите към външни системи;
3. безхартиено счетоводство;
4. подобряване на съгласуваността на данните;
5. оценка на качеството и точността на данните;
6. отчитане в реално време;
7. сложни анализи на големи данни;
8. автоматизиране на рутинните процеси;
9. високи очаквания към одиторите.

В други изследвания се акцентира на [3]:

1. разширяване на обхвата от бизнес дейности, които намират своето отражение в счетоводството;
2. разработване на подходи за интегриране на различни видове счетоводни дейности;
3. ускоряване на процеса по осчетоводяване;
4. подобряване на начина, по който се натрупва и съхранява счетоводна информация.

След анализ на спецификата на бюджетните организации, ще бъдат формулирани основните направления за дигитална трансформация на бюджетното счетоводство.

Специфика на счетоводната отчетност в бюджетните организации в Република България

Счетоводството в бюджетните организации има основните характеристики на всяко счетоводство. От друга страна, то има редица особености, които заедно с особеностите на самите

бюджетни организации трябва да се вземат предвид при формулиране на насоките за дигитална трансформация.

Специфика на бюджетното счетоводство

- **Отчитане на бюджетни средства.** Основната разлика между бюджетното и фирменото счетоводство се състои в отчитането на бюджетните средства. Отделно от счетоводното отчитане на приходите и разходите (на начислена основа) се съставя и отчет за приходите и разходите на бюджетни средства (на касова основа).
- **Отчитане на ангажименти.** През последните години беше въведено детайлното отчитане на поетите ангажименти от бюджетните организации. Това в значително степен утежнява и усложнява счетоводната отчетност без съществена практическа полза. Ресурсите, изразходвани за отчитане на незначителни по размер ангажименти в много случаи надхвърлят евентуалните ползи.
- **Отчитане на задължения към доставчиците.** Наложил се като правило да се отчитат задължително всички задължения към доставчиците, независимо от това, че в много случаи възникването на задължението съвпада с неговото погасяване или е поне в същия отчетен период. Това също увеличава в значителна степен обема на работата, особено когато се съчетае с отчитането на ангажиментите.
- **Операции с пари в брой.** Въпреки непрекъснатото намаляване на дела на тези операции, все още съществуват плащания в брой, например заплащане на държавни и местни такси и цени на услуги от граждани и юридически лица, изплащане на стипендии на учащите се и т.н.

Специфика на бюджетните организации

- **Голям брой разпоредители с бюджетни средства.** Докато първостепенните и второстепенните разпоредители са сравнително ограничен брой, то броят на третостепенните разпоредители, особено в общините и някои министерства е изключително голям. Според Единната бюджетна класификация за 2023 г. броят на първостепенните разпоредители от централното управление е 120, а от общините 265 или общо 385 [4]. Същевременно, само в Столична община има 60 второстепенни и 428 третостепенни разпоредители с бюджетни средства [5]. Това е предпоставка за трудна координация и налагане на еднакви стандарти на работа.
- **Съществени разлики в дейността.** Въпреки че предмет на изследването са бюджетните организации като цяло, те не са хомогенна група по отношение на извършваните в тях дейности. Тук са включени министерства и други институции от централното управление, висши училища, общини, общински предприятия, училища, детски градини, детски ясли, кухни, домове за стари хора и т.н. Всички те имат специфична дейност, която трябва да се има предвид при анализа.
- **Недостатъчно финансиране.** На практика почти всички бюджетни организации са недостатъчно финансирани в по-малка или по-голяма степен. Това означава недостатъчно средства за технически средства, софтуер, както и за заплащане и квалификация на персонала.
- **Недостатъчна подготвеност на кадрите.** Като следствие от предходните фактори (необходимост от голям брой специалисти, за които няма достатъчно ресурси), съществена част от специалистите нямат достатъчна степен на подготовка, в това число и компютърна грамотност.
- **Недостатъчна ефективност.** В малките бюджетни организации от рода на общински училища, детски градини, ясли и др. един специалист трябва да се занимава както със счетоводната отчетност и отчитането на бюджетни средства, така и с много други съпътстващи дейности. Това възпрепятства специализацията и води до по-ниска ефективност на труда на тези специалисти.

- **Различен начин на управление на бюджетните средства.** Все още не всички общини са включени в системата за електронни бюджетни разплащания (СЕБРА). Това обуславя различен начин на работа в тях.

Основни направления за дигитална трансформация на счетоводството в бюджетните организации

Анализът на спецификата на бюджетните организации и воденото в тях счетоводство води до формулирането на следните направления за дигитална трансформация на бюджетното счетоводство:

1. **Еднородност на системите, използвани в организацията.** Това е задължително условие за извършване на дигитална трансформация на процесите в организацията. Изолираните и несъвместими помежду си софтуерни продукти и системи възпрепятстват истинската дигитална трансформация. Най-добрият вариант би бил създаването на единна информационна система, обхващаща всички сфери на дейност, каквито са ERP системите.
2. **Съвместимост на системите, използвани в системата на първостепенния разпоредител с бюджетни средства.** Използваните софтуерни системи във второстепенните и третостепенните разпоредители трябва да бъдат ако не напълно унифицирани, то поне съвместими, за да се гарантира достатъчно високо ниво на интеграция в системата на първостепенния разпоредител.
3. **Разширяване и усъвършенстване на интерфейсите към външни системи.** Докато данните, генерирани в рамките на организацията могат да бъдат обхванати веднага от информационната система и да се използват в счетоводството, то постъпващите от външни източници данни също трябва да бъдат дигитализирани и обработвани автоматизирано. Става дума за банковите документи, както и документите от доставчиците, клиентите и др. В идеалния случай те постъпват директно в счетоводната система през подходящ интерфейс [6].
4. **Безхартиено счетоводство.** Както вътрешните, така и външните документи следва да бъдат електронни. Трябва да се предвиди надеждното съхраняване в рамките на информационната система както на информацията от тези документи, така и на самите документи. Постепенно използването на хартиени документи трябва да се намалява, докато изчезне напълно.
5. **Автоматизиране на рутинните процеси.** Голяма част от процесите в счетоводството могат да бъдат автоматизирани. Автоматизираното постъпване на електронни документи в информационните системи може да бъде последвано от автоматизирана обработка. Могат да бъдат прилагани успешно схеми за автоматично осчетоводяване на по-голямата част от документите. Това ще доведе до ускоряване на процеса по осчетоводяване, рационализиране и повишаване на ефективността от работата на счетоводителите [7].

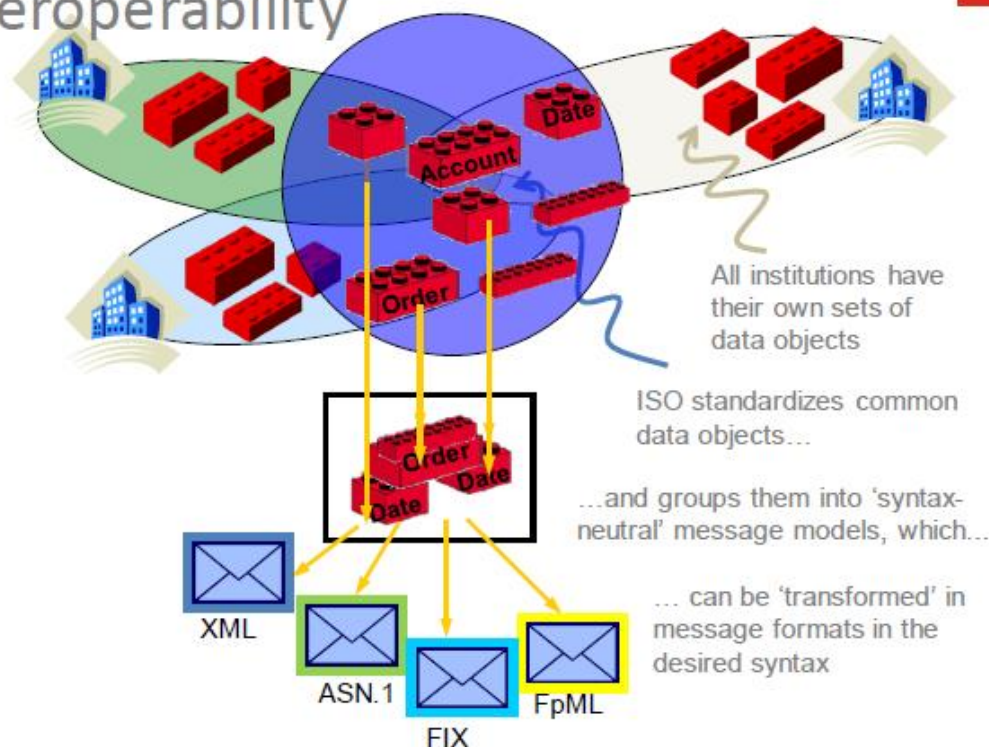
Някои направления за дигитализация на фирменото счетоводство поне засега не са особено съществени за бюджетното счетоводство. Такива са най-вече отчитането в реално време и сложните анализи на големи данни. В бъдеще може би ще има необходимост и от такива дейности, но в момента бюджетното счетоводство е далече от подобни проблеми.

Проблеми и възможни подходи при дигитализацията на счетоводната отчетност в бюджетните организации

Ще анализираме всяко от основните направления за дигитална трансформация на бюджетното счетоводство, за да открием потенциални проблеми и да предложим варианти за решение.

1. **Еднородност на системите, използвани в организацията.** Състоянието в момента е доста далеч от желаното. В повечето случаи се използват отделни приложения за счетоводство, работни заплати, дълготрайни активи, материални запаси и др. **Основен проблем** е закупуването и внедряването на ERP система, която да интегрира всички дейности. Това не е по силите на болшинството бюджетни организации и най-вече на второстепенните и третостепенните разпоредители с бюджетни средства. Те не разполагат нито с финансови ресурси, нито с квалифициран персонал, който да извърши внедряването и впоследствие да осъществява поддръжка на системата. **Възможно решение** би било централизирано финансиране и създаване на облачно базирана ERP система за бюджетни организации, която да бъде достъпна за всички. Някои бюджетни организации, които извършват по-специфични дейности, може да финансират разработването на допълнителни модули за тези дейности.
2. **Съвместимост на системите, използвани в системата на първостепенния разпоредител с бюджетни средства.** Обичайно счетоводните системи във второстепенните и третостепенните разпоредители са твърде разнообразни и не позволяват директен обмен на информация. Така **основен проблем** се явява събирането и обобщаването на отчетна информация. Много първостепенни разпоредители са създали собствени системи за обобщаване на отчетните данни, но това налага повторното им въвеждане от счетоводителите. **Възможно решение**, както и в по-горе дискутирания случай, е използването на еднакви счетоводни системи, които да имат и функционалност за обобщаване на информацията на по-горно ниво. Алтернативен подход би бил да се дефинират стандартизирани интерфейси за предаване на данни към висшестоящите организации, а евентуално и към останалите държавни органи. Използваните в момента файлове на Excel за обмен на счетоводни данни и данни за приходите и разходите по бюджета не са най-доброто решение. Има и други практически решения за интегриране на съществуващите системи с облачно базирани системи [8]
3. **Разширяване и усъвършенстване на интерфейсите към външни системи.** До известна степен решение може да се търси в използването на универсален софтуер, предлагащ подобна функционалност. **Основният проблем**, обаче, е в околната среда. Няма как едностранно да се дигитализира обмена на данни само от гледна точка на бюджетните организации, без това да засегне всички контрагенти (банки, доставчици, клиенти и т.н.). **Решението** е в нормативната регламентация и налагане на стандарти. Комуникацията с електронни документи трябва да стане задължителна, като същевременно се наложат единни стандарти за обмен на данни. Това постепенно се реализира в практиката - от 01.10.2023 са в сила изисквания за структурата, формата и съдържанието на платежните документи за целите на изпълнение на преводи от/към сметки на бюджетни организации в XML формат съгласно стандарт ISO 20022 [9]. Този стандарт определя основните обекти и съобщенията в банковата и финансовата сфера (Фиг. 1). Подобни стандарти трябва да се наложат и при обмен на фактури и други документи. В тази посока вече има съществуващи платформи за обмен на електронни фактури, но би било добре това да се регламентира нормативно.

Using ISO 20022 modelling to reach interoperability



Източник: Intro to ISO 20022 [10]

Фигура 1: Използване на стандартизирани обекти и съобщения в ISO 20022

4. **Безхартиено счетоводство.** Има две основни предизвикателства пред реализирането на изцяло безхартиено счетоводство - създаването на цялостна система за документооборота в организацията, която да регистрира и надеждно да съхранява в електронен вид както вътрешните, така и външните документи и, от друга страна, постепенното елиминиране или предефиниране на процесите, предполагащи използването на хартиени документи. Такива процеси, например, са тези, свързани със заплащане в брой на държавни или местни данъци и такси. Без трети независим орган (например банка), който да удостовери извършването на плащането, заплатилото сума физическо лице няма как да докаже впоследствие извършването на плащането освен с класически хартиен документ. **Решение** за документооборота отново може да се намери в цялостната информационна система на организацията. Колкото до проблемните бизнес процеси - може постепенно да се ограничи в максимална степен плащането в брой и същевременно да се мисли за електронно удостоверяване на всички извършени сделки, обвързано с електронната самоличност на физическите лица.
5. **Автоматизиране на рутинните процеси.** Това е традиционно направление, което и сега се развива успешно в счетоводния софтуер. Това е начинът за решаване на проблема с недостатъчната ефективност на счетоводителите и ангажирането им с твърде много рутинни задачи по следене на аналитично ниво на разчетите с доставчиците и клиентите, както и ненужно усложнената отчетност, свързана с поетите ангажименти по договори. **Предизвикателствата** в тази област са свързани най-вече с точното регламентиране на начина на комуникация с външните системи, така че да се елиминира рискът от грешки и несъответствия в партидите. Освен софтуерните решения, не бива да се изключват и някои чисто нормативни и административни **решения**, като например ограничаване на отчитането на ангажиментите само до най-съществените (над 1 милион лева),

окупняване на звената, извършващи счетоводното отчитане, така че те да обслужват множество третостепенни разпоредители и др.

Заклучение

Дигитализацията и дигиталната трансформация на счетоводната отчетност в бюджетните организации не бива да изостава от общото развитие. Независимо от проблемите и предизвикателствата, породени най-вече от недостатъчната ресурсна обезпеченост на тази дейност, могат да се търсят решения, които да ускорят процеса на дигитална трансформация. В настоящия доклад, освен че бяха дефинирани някои от основните проблем, бяха предложени и възможни конкретни подходи за решаването им.

References:

1. Лазарова, В., (2020). Дигитализация в счетоводството. София, Авангард Прима, 2020. ISBN 978-619-239-465-3.
2. Лазарова, В., (2019). Дигитализация и дигитална трансформация в счетоводството, *Икономически и социални алтернативи*, 2019/2, стр. 97-106.
3. Kovalevska, N., Nesterenko, I., Lutsenko, O., Nesterenko, O., Hlushach, Y., (2022). Problems of accounting digitalization in conditions of business processes digitalization., *Amazonia Investiga*, Vol. 56, стр. 132-141, <https://doi.org/10.34069/AI/2022.56.08.14>
4. Министерство на финансите (2023). Единна бюджетна класификация за 2023 г. Официална интернет страница. [Online] 2023. <https://www.minfin.bg/upload/55401/ЕБК-2023.zip>.
5. Столична община. (2021) Приложение № 9 към Решение № 107 по Протокол № 29/25.02.2021г. Столичен общински съвет. [Online] 2021. <https://council.sofia.bg/documents/20182/937233/R.107-PR.9.pdf>.
6. Butunov, S. B., & Azamatova, G. I. (2020). The digitalization of accounting: from the past to the future. *ISJ Theoretical & Applied Science*, 04 (84), стр. 514-519.
7. Gherman, M., Molociniuc, M., Grosu, V. (2021). Digitalization of accounting - trends and perspectives. *Journal of information systems & operations management*, Vol. 15.1, стр. 104-113.
8. Tsaneva, M. (2019). A practical approach for integrating heterogeneous systems, *Business management*, no. 2, стр. 5-16, 2019.
9. Министерство на финансите (2023). ДДС № 03/03.05.2023 г. - Изискванията за структурата, формата и съдържанието на платежните документи за целите на изпълнение на преводи от/към сметки на бюджетни организации в XML формат съгласно стандарт ISO 20022 кредитен превод. Официална интернет страница. [Online] 2023. https://www.minfin.bg/upload/54839/DDS-03-2023-Article-151_1-ZPF-Budget-Pay-Docs.pdf.
10. ISO (2022). Introduction to ISO 20022 –Universal financial industry message scheme. ISO 20022. [Online] 2022. <https://www.iso20022.org/sites/default/files/2022-02/introtoiso20022.pdf>

РАЗВИТИЕ НА ТУРИСТИЧЕСКИЯ ПАЗАР В БЪЛГАРИЯ ЧРЕЗ ДИГИТАЛИЗАЦИЯ В ТУРИЗМА

Дарина Георгиева Цветанова

Ph.D. Student, Department "Economics of Tourism, University of National and World Economy,
Sofia, email: darina.tsvetanova@unwe.bg

Борис Георгиев Карагеоргиев

Ph.D. Student, Department "Economics of Tourism, University of National and World Economy,
Sofia, email: boris.karageorgiev@unwe.bg

Резюме

Туристическият пазар през последните години все повече се трансформира и все по-ключово място за позициониране на бизнеса заемат иновативните технологии, дигитализацията и дигиталния маркетинг. Важна специфика е по-лесното опериране на индустрията с огромно количество бази данни, чрез които може да се осигури по-качествено, адекватно и бързо обслужване на потенциалните и реални потребители на туристически продукт.

Иновативните технологии и дигитализацията не са просто технология, а начинът, по който те се използват за ангажиране на аудиторията, като атрактивността се съдържа в това, че непрекъснато се дава възможност за разширяване на съществуващи пазари, но и за откриване на нови такива. Според редица автори дигитализацията и иновациите имат все по-нарастващо значение, като същевременно с това и все повече нарастват разходите за реклама в дигитална среда, което е показателно за ключовата му роля в съвременния свят и отношението между бизнеса и всички заинтересовани страни.

Ключови думи: туристически пазар, дигитализация, развитие, туризъм

DEVELOPMENT OF THE TOURISM MARKET IN BULGARIA THROUGH DIGITALIZATION IN TOURISM

Darina Georgieva Zvetanova

Boris Georgiev Karageorgiev

Abstract

In recent years, the tourism market has been increasingly transformed, and innovative technologies, digitization and digital marketing occupy an increasingly key place for the positioning of businesses. An important specificity is the easier operation of the industry with a huge amount of databases, through which it is possible to provide better quality, adequate and fast service to the potential and real users of a tourism product. Innovative technologies and digitization are not just technology, but the way in which they are used to engage the audience, the attractiveness of which is that it constantly enables the expansion of existing markets, but also the discovery of new ones. According to a number of authors, digitization and innovation are of increasing importance, while at the same time the costs of advertising in the digital environment are increasing, which is indicative of its key role in the modern world and the relationship between business and all stakeholders.

Keywords: tourism market, digitalization, development, tourism

JEL: M2, M3, Z3

Увод

Туристическият пазар през последните години все повече се трансформира и все по-ключово място за позициониране на бизнесите заемат иновативните технологии, дигитализацията и дигиталния маркетинг. Важна специфика е по-лесното опериране на индустрията с огромно количество бази данни, чрез които може да се осигури по-качествено, адекватно и бързо обслужване на потенциалните и реални потребители на туристически продукт.

Дигитализацията в туристическия бизнес е един от основните компоненти, които създават възможности за развитието на туристическия продукт, неговата визуализация, адаптивност и приспособимост към изискванията на съвременния турист. **Обект** на изследване е туристическия бизнес, **предмет** на изследване са възможностите за дигитализация и нейното приложение, както и ефекта върху търсенето и предлагането в туризма. **Целта** на доклада се свързва с доказателството, че развитието на туристическия пазар се обуславя от различни фактори, един от които е дигитализацията на бизнеса. **Задачите** доказващи целта се фокусират върху предимствата на новите технологии, развитието на туристическия пазар в България, както и върху дигиталните тенденции на българския туристически пазар.

1. Предимства на новите технологии

Едно от най-големите предимства на новите технологии в помощ на развитието на туристическия бизнес и развитието на туристическия пазар е възможността за постоянен контакт и по-голяма ангажираност на потенциалните туристи, както и проследяването в кой етап от комуникационните процеси могат да се загубят клиенти или кога е възможно да се създадат лоялни туристи. Това е особено ценно за дейността на туроператорите, хотелиерите и ресторантьорите, тъй като на българския туристически пазар има многобройни туристически предприятия, предлагащи сходни продукти и използвайки голямата база данни и инструменти за дигитален маркетинг, като по този начин отделните участници на пазара могат по-правилно да се позиционират.

Понастоящем на българския туристически пазар има над 4000 регистрирани туроператора и турагента, с валидна застраховка „Отговорност на туроператор“ (Национален туристически регистър, Туроператори и турагенти с валидна регистрация, 2023) и над 27 000 места за настаняване с валидна категоризация (Места за настаняване в България с валидна категоризация, 2023) и над 20 000 заведения за хранене и развлечение с валидна категоризация (Места за хранене и развлечение в България с валидна категоризация, 2023) и над 3500 туристически атракции и над 1800 туристически събития и фестивали (Регистър на туристическите атракции, фестивали и събития, 2023).

Развитието на туристическия пазар най-лесно може да се постигне с помощта на изкуствения интелект, който все повече става водещ инструмент в дигиталния маркетинг, осигуряващ конкурентно предимство, въз основа на 6 основни показатели, които са ключови за маркетинговите анализи – правилните данни, правилната роля в точното време, правилният продукт, правилният клиент и правилният канал (Zlateva, 2021:524).

Все повече фирми, особено в туристическия сектор, се обръщат към въвеждането на автоматизирани системи, виртуални асистенти и чат ботове, които използват ефективно в комуникацията със своите потенциални и реални потребители, като едновременно с това все повече се използват различни инструменти, с които се проследява поведението им, в следствие на което сегментирането става по-правилно, а маркетинговите бюджети по-оптимизирани. Трудността, обаче, която се явява пред маркетинговите специалисти по отношение на разпределението на маркетинговите бюджети е по какъв начин да ги разпределят пропорционално между онлайн и офлайн среда, т.к. според нас най-правилното решение би било да се интегрират и разпределят, както в реална среда, така и във виртуална, като внимателно се подберат офлайн каналите, които да се използват.

Важно е да се упомене, че днешните потребители са все по-информирани и взискателни и това води след себе си до промяна в маркетинговата комуникация, както и в мисленето. Днес конвенционалният, традиционен маркетинг и паралелното му използване в едно с дигиталния маркетинг доведе до промяна в начина на комуникация, като от еднопосочна, тя стана

двупосочна и все по-належащо незабавна, за което в огромна помощ са чат-ботовете и виртуалните асистенти. Дигиталното присъствие на фирмите, Интернетa и новите технологии донесат със себе си три важни промени във взаимодействието с клиентите, както в реална среда, така и във виртуална. Тези промени W. Rowan ги обособява по следния начин - за вземане на решение за покупка вече не е важно физическото местоположение на обекта, възприемането на една компания/организация/ започва да се основава не само на техните продукти и услуги, но и на нейната дигитална марка и присъствие (Rowan, 2022).

В средата на туристическия пазар стоят пътуващите/туристите и отношенията с тях на всички субекти и именно тук се крие едно от основните им предимства.



Фиг.1 Туристическата система/ Туристическа среда

Източник: Cooper, Ch. et al., Tourism: Principles and practice – Fourth edition, Prentice Hall, Pearson Education Hall, England, 2008

За да се прилага успешно дигитализацията в туризма и за да има резултат е необходимо да се следва конкретна стратегия, в която да има дефинирани цел, визия, тактика и целева аудитория и да предлага следване на целенасочени действия (Рибов, 2011).

В дигиталната стратегия ключово място заема правилното използване на социалните медии, дигиталната реклама, оптимизацията на търсачки, изграждането на корпоративен сайт, инфуенсър маркетинга, e-mail маркетингът, CRM системите, като всеки един от посочените елементи има различни предимства, но за всичко това ще стане въпрос по-нататък в текста, като изтъкнем техните характеристики в използването за целите на маркетинга в туроператорските агенции. За да е успешна една компания, тя трябва да има ясно формулирани цели и задачи, които да са част от разписана маркетингова стратегия.

Приложението на дигиталният маркетинг позволяват разширяване на обхвата и стесняване на фокуса върху специфични нишови сегменти вътре в широкия пазар (Zlateva, 2021:528), като тук е моментът да се подчертае динамичността и разнообразието по отношение на маркетинговия подход, който се използва при изграждането на отношенията с клиентите.

Маркетинговият подход, който се използва в дигитална среда позволява лесното проследяване на поведението на потребителите, както предоставя възможност за идентификация в кой етап на процеса се намират уникалните потребители, като по този начин се изготвя и предоставя подходящо съдържание за потребителя. Маркетинговият подход е и в основата на планирането по отношение на стратегията, за да се създаде възможност за достигане на желания краен резултат за покупката на туристическия продукт – настаняване, хранене, транспорт или други допълнителни услуги, като посещение на атракция, събитие и други.

Дигиталните тенденции в туризма и дигиталното мобилно поведение на клиентите са тясно свързани с общите тенденции в туризма (Gelter, 2017). Дигиталните технологии и тяхното използване в туризма са свързани с иновативните подходи, а те от своя страна предполагат растеж и по-висока конкурентоспособност за туристическите предприятия, като той е резултат

на обновяването, както посочва Питър Дракър (Drucker, 1999). Това се затвърждава от факта, че БВП се увеличава на база по-големия интензитет на създаване и управление на иновациите. Посочените фактори показват необходимостта от маркетингов анализ на туристическия пазар, като целта му е създаване на конкурентни предимства на конкретната фирма, прилагаща иновации и дигитализация чрез различни канали.

2. Развитие на туристическия пазар в България

Още преди появата на здравната криза по данни от международни изследвания, през 2017 г. 46% от туристите в Европа са организирали пътуванията си онлайн, през 2018 г. потребителите на Интернет са вече над 4 млрд., а на мобилни услуги – повече от 5 млрд. Дигитализацията трансформира осезаемо туристическата индустрия и вече всеки втори турист организира почивката си чрез интернет, половината от пътуванията започват от мобилно устройство и неговото разпространение дигитализацията, иновативността и използването на новите технологии е започнало да придобива популярност в различни сфери от обществения живот, но по време на здравната криза те стават още по-актуални и популярни, заедно с маркетинговото и мениджърско управление, като се превръщат в основен начин различните бизнеси да се запазят на пазара.

Промяната на туристическия пазар посредством дигитализацията се реализира и чрез електронното управление на макро равнище, което е бъдещето. С цел оптимизиране на маркетинговите дейности от изпълнителната власт в България, в лицето на Министерството на туризма, е създадена интернет платформата и мобилно приложение iLoveBulgaria (Национална туристическа платформа iLoveBulgaria), с което да се популяризира бранд България. Платформата съдържа информация за туристически обекти и услуги по области и интереси и спомага за намирането им в реално време и място чрез GPS & QR скенер. Платформата използва най-модерните технологии и е застъпена възможността за директно споделяне на туристически обекти чрез Chat Extension на Вайбър. Данните отчитат 25 000 сваляния на iLoveBulgaria, 16 000 посетители на месец на уебсайта и блога му, 100 000 последователи на Фейсбук страницата и над 366 000 достигнати лица, близо 6500 последователи в Инстаграм. В тази посока се създава и Постоянен съвет за дигитализация в туризма (iNews.bg, 01.11.2023).

Туризмът е свързана система, съставена от сложна дистрибуторска верига, в която различни туристически предприятия си взаимодействат и всички заедно или по отделно използват информационните технологии и иновации, за да останат свързани помежду си, като едновременно с това новите технологии им помагат да разработват и произвеждат нови продукти, да администрират и управляват дейността си, да ръководят маркетинга си и пазарната комуникация, продажбите и резервациите, и да комуникират с партньорите и клиентите си. Дистрибуционната политика избира канали, по които може да бъде продаден туристическият продукт (Евтимова, 2006). Колкото е важна дигиталната среда за туристическите предприятия, толкова Интернет и новите технологии са използвани от туристите. Смарт устройствата, Интернетът и новите информационни технологии присъстват почти постоянно в живота, работа, свободното време, пътуванията, пазаруването и т.н. Този факт променя конюктурата на туристическото търсене на различни продукти, тъй като потребителите стават все по-информирани и изискващи.

В резултат не само на посочените фактори, свързани с дигитализацията, но и на други пазарни фактори извън посочените, осигуряващи конкурентни предимства на туристическите обекти се констатира завишен интерес на потреблението в България през 2023 г. в сравнение с 2022. Според НСИ, през юни 2023 г. в страната са функционирали 3 232 обекта с 10 или повече легла - хотели, мотели, къмпинги, хижи и други места за краткосрочно настаняване, със 138.4 хил. стаи и 310.1 хил. легла в тях. В сравнение с юни 2022 г. местата за настаняване се увеличават с 4.6%, а броят на леглата в тях - с 5.1%. Общият брой на ношувките е 3 337.9 хил., или с 18.2% повече в сравнение със същия месец на предходната година. Броят на пренощувалите лица в местата за настаняване се увеличава с 16.4% в сравнение със същия месец на 2022 г. и достига 975.5 хиляди. Приходите от ношувки достигат 229.2 млн. лв., или с 28.2% повече в сравнение с юни 2022 година (Национален статистически институт, 2023).

3. Дигитални тенденции на туристическия пазар в България

Всичко посочено до тук акцентира върху необходимостта да се анализират дигиталните тенденции, от които туристическото предлагане в България може да се възползва, за да отговори на нуждите и новото поведение на туристите, за по-голямата им нужда от преживявания и качество на продукта, като от друга страна да доведе до устойчиво и възходящо развитие на пазара и през следващите години.

Първата тенденция, която се констатира касае факта, че съвременния потребител става все по-глобален, все повече използва мобилните си телефони и в тази връзка е важно дигитализирането на бизнеса да бъде съобразено и адаптирано към този тип устройства, за да може във всеки един момент да бъде близо до потенциалните си клиенти. На преден план в иновативните подходи, свързани с дигитализацията, които туристическите предприятия трябва да внедрят в маркетинговото си управление трябва да са подчинени на идеята за заедност, възможност за образование и полезност, под формата на геймификация, да предлагат забавление, но същевременно с това да бъдат лесни и достъпни за управление и употреба. Също така дигитализацията не трябва да бъде за сметка на качеството на обслужване, удобството на клиента, което изпитва, а напротив иновациите само трябва да засилят това усещане.

Следващата тенденция се свързва с дигиталните разплащания на потребителите на туристически продукт. Като един от ключовите иновационни елементи в почти всички видове туристически предприятия е начинът на разплащане. Профилът на съвременния турист е силно изменен и той днес търси все по-бързи, удобни и вариативни решения за реализиране на почивката си. Понастоящем потребителите на туристически продукт пазаруват онлайн, използват кредитни, дебитни карти за разплащане и онлайн банкирането, тъй като тези разплащателни методи предполагат спестяване на време в офисите на туристическите компании. Туристическите предприятия, използващи повече и по-разнообразни иновативни методи на разплащане, гарантират по-голяма вероятност за резервации.

Друг иновативен подход свързан с дигиталните тенденции е преминаването от масовото производство към масовото персонализиране в бизнеса, с фокус към маркетинговото управление, насочено по персонален начин към клиента, използвайки потенциала на „големите бази данни“, като все по-водещо е изграждането на взаимоотношения с потребителите, а не толкова акцент върху конкретната продажба. Внедряването на чат-ботове към сайтовете или социалните мрежи на туристическите обекти, също са друга нова тенденция, налагаща се в резултат на факта, че съвременния клиент има нужда от незабавна комуникация и получаване на информация в реално време.

Дигитализацията на туристическия бизнес обхваща всички дейности – дигитален маркетинг, дигитално управление, електронни финанси, счетоводство, управление на човешки ресурси, резервации, продажби, комуникация, като за целта туристическите предприятия, за да бъдат конкурентноспособни и ефективни, е необходимо да следят световните тенденции за развитие на хардуер и софтуер технологиите, отговарящи на техните нужди, както и да правят опити в кратки срокове да бъдат използвани.

Едни от революционните и по-иновативни подходи, които започват да се утвърждават през последните години в туристическото предлагане са създаването на виртуална (VR) и разширена реалност (AR), като именно тази иновация ще набира все повече популярност сред туристическите предприятия, тъй като позволява едновременно сливане на реално и виртуално, минало, настояще и бъдеще. Внедряването на VR и AR в туристическите предприятия са само част от иновативните тенденции, от които могат да се възползват туристическите бизнеси за успешно маркетингово управление и развитие на пазара, както и да създаването на допълнителни и по-специализирани продукти. Все по-често като иновативен подход, започват да се използват 6DoF VR слушалки, VR видео, 360 камери, YouVisit приложението и смартфоните, които успяват да помогнат на потенциалния туристически потребител да пътува до далечни дестинации без да напуска комфорта на дома си. Необходимо е правилно да се разбират технологиите AR и VR, като първата предлага подобряване на изживяването на мястото, а второто ни позволява предварително изследване на дестинацията и продуктите в нея (Янева и кол., 2023:179-190).

Дигиталните тенденции и иновациите на глобално равнище на българския туристически пазар, водещи до неговото развитие, са свързани и с провеждането на Международната конференция за дигитална трансформация в туризма през 2022 година, на която сред приоритетните задачи в областта на туризма са иновациите, които да направят страната ни целогодишно атрактивна туристическа дестинация с разработване на единен дигитален пътеводител, който да посреща всеки турист, посетил България. Подобна виртуална карта на България ще цели да обединят различните туристически продукти – като културни и винени маршрути, фестивали, фермерски пазари, развиващи местните общности, значими природни обекти и знакови забележителности от културно-историческото ни наследство, които представляват интерес за всеки гост на дестинацията (Експерти от осем държави споделят опит на конференция за дигитална трансформация в туризма в София, 25 октомври 2022).

Друга дигитална иновация на туристическия пазар в България, която осигурява неговото развитие е платформата Umni.bgm (AI чатботът: Градският консиерж, свързващ местата и хората, 2023), която през март '23 печели глобално състезание за стартъпи на Световната организация по туризъм на ООН (UNWTO) в категория Иновации и дигитализация за островни дестинации със своя островен AI Консиерж. Иновацията представлява интелигентен дигитален асистент за цяла дестинация, която е реализирана за туристите на град Пловдив като Европейска столица на културата през 2019 г., а е приложена и от други туристически градове в света. Технически AI чатбот за дестинации или така нареченият Градски консиерж обикновено отговаря за предоставянето на информация на посетителите на местна област, град, село, и т.н., както за популярни туристически атракции, ресторанти, възможности за развлечение или други съоръжения и местни услуги, така и за информация за болница, денонощни аптеки, посолства и консулства, прогнози за времето, валутен курс и места за обмяна на валута, бизнес услуги, дори услуги за домашни любимци за тези, които пътуват с домашните си любимци. Тук стои въпросът за необходимостта на подобни платформи, т.к. в гугъл карти подобна информация, също е достъпна за посетителите и туристите в определена дестинация, но може би не е дотолкова систематизирана.

В България през 2021 г. Созопол (Bgtourism, 15/10/2021) е номиниран и е получил награда в категорията Смарт Сити със 145 електронни услуги и мобилно приложение, както и паркинги снабдени с LPR камери, а курортен комплекс Слънчев бряг също отчита редица постижения по повод дигитализацията на паркингите си за разпознаване на номерата на автомобилите в курорта. Всяка година тази класация предизвиква местната власт да активизира внедряването на дигиталните иновации в дестинациите.

Развитието на туристическия пазар в България посредством внедряване на дигитализацията се наблюдава и в реализираните проекти на редица туристически обекти у нас. Например Хотел Jagerhof, (21hoteliers, 17.11.2021) в Пловдив е един от новите лидери с дигитализацията чрез възможността телевизорите в стаите да дава възможност от менюто му да се поръчва рум сървиз, да се избират филми от каталог с нови заглавия, както и гостите да проверяват сметката си. Дигитално да се избират специални възглавници, затъмняващи завеси, ергономични матраци, които гарантират пълноценната почивка на гостите.

Възможностите за устойчив растеж и развитие на България като туристическа дестинация, се отчитат и при нейното популяризиране като четири сезонна дестинация. В тази връзка световната организация по туризъм публикува 64-те най-добри примери за дестинации, които използват силата на туризма за устойчив растеж, сред които е и България. В класацията страната е представена с 2 рекламни видео клипа, които потвърждават вече посоченото – „Find your story” и „Cultural tourism, gourmet cuisine & wine”. В платформата на Министерство на туризма са налични 64 дигитални видео клипа, които представят българския туристически пазар, както и стимулират възможностите за туристическото предлагане и туристическото търсене, като удовлетворяват и новите изисквания на потребителите (Янева и кол., 2023: 172-179).

Ако през 2019 г. разходите за реклама в Интернет са в размер на \$299 милиарда, то през 2021 г. техният размер възлиза на \$521,02 милиарда щатски долара в световен мащаб (www.statista.com, 2022). Това почти двойно увеличение в рамките на 2 години може да се обясни и с желанието компаниите да останат свързани със своите потребители и да придобият по-голяма разпознаваемост, което е по-трудно да се случи в традиционния туристически пазар,

предвид променената пазарна конюнктура за посочения период в световен аспект, а не само в България.

Осъзнавайки предимствата на дигиталния маркетинг, то съвсем резонно се прогнозира, че до 2026 г. разходите за реклама в дигиталния свят ще достигнат до \$ 876 милиарда долара, като с все по-нарастващо значение ще бъде мобилната интернет реклама и това е най-лесният и ефективен начин компаниите да бъдат в непрестанно взаимодействие с всички пазари, потребители и партньори.

Чрез дигитализацията, освен в интернет, а и чрез съвременните информационни и комуникационни технологии се търсят възможности за привличане и запазване на пазарен потенциал, колкото е възможно по-устойчиво, както и изграждане на трайни взаимоотношения. Доказано е развитието на туристическия пазар, обусловено от различни фактори, един от които - дигитализацията на бизнеса.

Заключение

В настоящото изложение е показано развитието на туристическия пазар в България, обусловено от приложението на дигитализацията и иновациите приоритизирани в туристическото предлагане, както на ниво изпълнителна власт, така и на бизнес равнище, прилагани от Министерството на туризма, хотелиерите и туроператорите в България, обусловени от туристическото търсене на все по-мобилни възможности за информация, покупка и потребление на туристически продукт. Анализирани са възможностите за дигитализация и нейното приложение, както и ефекта върху търсенето и предлагането в туризма. Научното изследване представлява анализ на настоящата конюнктура на туристическия пазар в България, без претенции за всеобхватност, дълбочинност и емпирика. Използваните данни имат предимно вторичен характер, в резултат на което са обобщени, анализирани и дискутирани. Резултатите имат теоретико-приложен характер и могат да бъдат в полза, както на бизнеса, така и за научна основа на бъдещи изследвания.

References

1. Евтимова, Н. (2006). Маркетинг в туризма. София: Дионис. (Evtimova, N., 2006. Marketing v turisma, Sofia, Dionis).
2. Експерти от осем държави споделят опит на конференция за дигитална трансформация в туризма в София, Министерство на туризма, 25 октомври 2022. Available at: <https://www.tourism.government.bg/bg/kategorii/novini/eksperti-ot-osem-durzhavi-spodelyat-opit-na-konferenciya-za-digitalna> [Accessed 02.10.2023]. (Eksperti ot 8 darjavi spodeliat opit na konferencia za digitalna transformazia v turizma, Ministerstvo na turizma, 25 oktombri 2022).
3. Национален статистически институт, 2023, Available at: <https://www.nsi.bg> [Accessed 02.10.2023]. (nsi.bg).
4. Национален туристически регистър, Министерство на туризма. Available at: <https://ntr.tourism.government.bg/Registration.nsf/alltato.xsp>, [Accessed 02.10.2023]. (Nazionalen turistiheski registar, Ministerstvo na turizma).
5. Национална туристическа платформа iLoveBulgaria, Available at: <https://ilovebulgaria.eu/> [Accessed 02.10.2023]. (Nazionalna turistiheska platforma iLoveBulgaria).
6. Места за настаняване в България с валидна категоризация, Министерство на туризма. <https://ntr.tourism.government.bg/CategoryzationAll.nsf/mn.xsp> [Accessed 02.10.2023]. (Mesta za nastaniavane v Balgaria s validna kategorizacia, Ministerstvo na Turizma).
7. Места за хранене и развлечения в България с валидна категоризация, Министерство на туризма. Available at: <https://ntr.tourism.government.bg/CategoryzationAll.nsf/zhr.xsp> [Accessed 02.10.2023]. (Mesta za hranene I razvlehenie v Balgaria s validna kategorizacia, Ministerstvo na Turisma).
8. Регистър на туристическите атракции, фестивали и събития, Министерство на туризма. Available at: <http://rta.tourism.government.bg/TAFRegister.aspx?AspxAutoDetectCookieSupport=1> [Accessed 02.10.2023]. (Registar na turistiheskite atrakcii, festivali I sabitia, Ministerstvo na Turisma).

9. Рибов, М. (2011). Конкурентни предимства в туризма, Издателство Тракия, София. (Ribov., M., 2011, Konkurentni predimstva v turisma, Izdatelstvo Trakia, Sofia).
10. Създават постоянен съвет за дигитализация в туризма, iNews.bg, 01.11.2023. Available at: Създават постоянен съвет за дигитализация в туризма - Туризм - inews.bg, [Accessed 02.10.2023]. (Sazdavav postoiainen savet za digitalizacia v turisma, iNews.bg, 01.11.2023).
11. Тихомир Янакиев е кмет на годината в категория „Smart city“, Bgtourism, 15/10/2021. Available at: <https://bgtourism.bg/15102021-sozopol-smart-city-tihomir-yanakiev/>(AI чатботът: Градският консиерж, свързващ местата и хората [Accessed 02.10.2023]. (Tihomir Ianakiev e kmet na godinata v kategoria „Smart city“, Bgtourism, 15/10/2021).
12. Янева, М., Цонев, Н. и кол., (2023). Развитие на туристическите райони на България, чрез моделиране на иновативни подходи при специализирани видове туризъм. Издателство УНСС, София, ISBN 978-619-232-747-7, с.172-179. (Ianeva, M., Conev, N., I kol., Razvitie na turistiheskite raioni v Balgaria hrez modelirane na inovativni podhodi pri sprcializirani vidove turism, Izdatelstvo UNSS, Sofia).
13. Cooper, Ch. et al., (2008). Tourism: Principles and practice – Fourth edition, Prentice Hall, Pearson Education Hall, England
14. Drucker, P. (1999). Innovate or Die. The Economist September 25th-October 1st, 27-34.
15. Gelter, H. (2017). Digital tourism - An analysis of digital trends in tourism and customer digital mobile behaviour for the Visit Arctic Europe project. he future belongs to those who prepare for it today. The Future starts now!
16. Hotel Weekly #16: Хотел Jägerhof, Пловдив, 21hoteliers, 17.11.2021. Available at: <https://21hoteliers.bg/hotels/hotel-weekly-16-hotel-jagerhof-plovdiv/>[Accessed 02.10.2023]
17. Rowan, W. (2022). Digital Marketing: Using New Technologies to Get Closer to Your Customer. London: Kogan Page Business Books, 6. Retrieved from https://www.academia.edu/25592254/Digital_Marketing_Using_New_Technologies_to_Get_Closer_
18. Statista, (2022, June). *Digital advertising spending worldwide from 2021 to 2026(in billion U.S. dollars)*. Retrieved June 10, 2022, from www.statista.com/: <https://www.statista.com/statistics/237974/online-advertising-spending-worldwide/>
19. Umni Bg, Available at: AI чатботът: Градският консиерж, свързващ местата и хората | [Accessed 02.10.2023]
20. Zlateva, D. (2021). Big Data world and the new marketing horizons. *Revista Inclusiones*, 8 num(ISSN 0719-4706), 519-535. Retrieved may 25., 2022, from <http://revistainclusiones.com/carga/wp-content/uploads/2021/03/36-Zatleva-VOL-8-NUM-Esp.-AbrJun-Cristian-Daniel2021INCL.pdf>

ЕЛЕКТРОННИТЕ АДМИНИСТРАТИВНИ УСЛУГИ НА ОБЩИНИТЕ – РЕГИОНАЛНИ АСПЕКТИ И ИЗМЕРЕНИЯ

Асен Божиков

Главен асистент, доктор, катедра „Бизнес информатика“, Стопанска академия „Д. А. Ценов“ – Свищов, e-mail: a.bozhikov@uni-svishtov.bg

Резюме

В доклада е разгледано състоянието на електронните административни услуги, предлагани от общините. Направен е преглед на съществуващи разработки по тематиката и са изведени проблеми и предизвикателства пред дигитализацията на административните услуги. Анализирани са резултатите от проведено емпирично изследване, с фокус върху сходствата и различията, по отношение на протичащата дигитализация в местната администрация в Северна и Южна България.

Ключови думи: е-управление, дигитализация, електронни административни услуги

THE ELECTRONIC ADMINISTRATIVE SERVICES OF MUNICIPALITIES - REGIONAL ASPECTS AND DIMENSIONS

Asen Bozhikov

Abstract

The paper examines the state of administrative e-services offered by local municipalities. It reviews existing developments on the subject and highlights the problems and challenges facing the digitalization of administrative services. The results of an empirical study with a focus on the similarities and differences in the ongoing digitalization of local administration in Northern and Southern Bulgaria are analyzed.

Key words: e-government, digitalization, administrative e-services

JEL: H83, L86

Въведение

През последните няколко години темата, свързана със състоянието на електронното управление (е-управление) и предоставяните електронни административни услуги (ЕАУ), отново възвърна своята актуалност. Причините за това могат да се търсят не само в нормативната рамка, приета на европейско и национално ниво, но също така и във фактори, като разразилата се пандемия от Ковид-19, която се оказа ускорител на процеса на дигитализация, не само в бизнес организациите, но и в публичния сектор. Технологичният напредък и свързаната с него дигитална трансформация доведе до нов тип рационализация в публичния сектор, което намира отражение в промяна на политиките и процесите, които протичат там. За е-управление в България се говори още от началото на новото хилядолетие, но едва след 2014 година се наблюдава съществен напредък и изменения в нормативната и институционална рамки.

Е-управлението се свърза с използване на възможностите на информационните и комуникационни технологии (ИКТ) с цел осигуряване на висока ефективност на дейностите, подобряване на отчетността и прозрачността, съкращаване на разходите и минимизиране на риска от корупция в администрацията (Khan, Krishnan, & Dhir, 2020).

Предоставянето на електронен достъп до информация и услуги на гражданите и бизнеса е съществена част от е-управлението в една страна. Осигуряването на ефективни и висококачествени публични електронни услуги е една от приоритетните области на действие на Националната програма „Цифрова България 2025“ (Министерство на транспорта, информационните технологии и съобщенията, 2019). Основната движеща сила на е-управлението е именно гражданското общество (Павлова, 2017). В тази връзка, трансформацията на начина на предоставяне на публичните услуги не трябва да се разглежда единствено като технологичен процес, а също така и като социален, в който ролята на гражданите е водеща, а техните желания са предпоставка за появата на нови или усъвършенстване на съществуващи услуги, както и за повишаване на административния капацитет, дори и в най-развитите страни (Павлова, 2018). Ползите за гражданите и бизнеса, които произтичат от тези електронни услуги, са свързани с намаляване на времето за изпълнение на услугата, спестяване на разходи и време за придвижване до общинската администрация, а така също и повишаване качеството на услугата.

Поради тази причина се полагат сериозни усилия, както на централно, така и на местно ниво, за преминаване към електронно административно обслужване чрез използване на съвременните ИКТ и фокусиране върху нуждите, очакванията и интересите на потребителите (ДАЕУ, 2021). Тази нова стратегическа визия за е-управление е насочена към постигане на по-висока информираност сред гражданите и бизнеса и по-лесен и бърз достъп до предоставяните от доставчика ЕАУ, а в по-дългосрочен хоризонт се предвижда изграждане и преминаване към мобилно електронно управление (Марков, 2022).

Преглед на съществуващи изследвания относно състоянието на ЕАУ в общините

Общините, като представители на местната власт, също се стремят към предоставяне на ЕАУ и проследяването на достигнатото равнище на дигитализация е от съществено значение за открояване на предизвикателствата, които трябва да се преодолеят в бъдеще. В литературата могат да бъдат открити редица разработки, изследващи състоянието на ЕАУ, предлагани от общините. Въпреки че в повечето от тях регионалният аспект не е застъпен, те дават ясна представа за достигнатата степен на развитие на дигиталните услуги на местно ниво и открояват пречките и предизвикателствата пред дигиталната трансформация на общинската администрация. В следващата таблица са систематизирани основните изводи от разработки, в които е направен анализ на резултатите от автори емпирични проучвания, проведени през 2022 години, относно предлагането на ЕАУ в общините.

Таблица 1: Констатации от проведени емпирични проучвания относно ЕАУ в общините

Брой респонденти	Основни констатации от проучването	Автор/и
109 общини (анкетно проучване по метода на отзовалите се)	<ul style="list-style-type: none"> ▪ Наличие на две групи бариери пред създаването и използването на ЕАУ в общините – външни и вътрешни. ▪ Външните бариери, подредени по низходяща степен на значимост, са: нагласите на гражданите и бизнеса за ползване на ЕАУ и открояващата се съдържаност към заявяване на такива услуги; регулаторните ограничения, свързани с разбиране на съществуващата нормативна база, което води до невъзможността тя да се приложи на практика; координацията на политиката за приложение и използване на ЕАУ, която в повечето случаи не се популяризира достатъчно. ▪ Вътрешните бариери, подредени по низходяща степен на значимост, са: характера на предоставяните услуги; липсата на достатъчно квалифициран персонал, който да обезпечи внедряването и предлагане на ЕАУ. ▪ Отчетени са предимствата за общинските администрации, произтичащи от внедряване на ЕАУ – повишава се динамиката на процеса след дигитализиране на услугата; не се наблюдава съществено дублиране на работата и нарастване на 	(Лазарова, Веселинова, Господинов, & Стоянова, 2023)

	<p>бюрократичната тежест; не се губи контакта с потребителите на услугата, дори тяхната обратна връзка спомага за подобряване на процесите в общинската администрация.</p>	
<p>45 общини (анкетно проучване чрез случайна извадка, стратифициране по големина на адм. единица и тип район)</p>	<ul style="list-style-type: none"> ▪ Сравнително слаб напредък в предлагането на ЕАУ, като малка част от общините позволяват двустранна комуникация с възможност за електронно връчване и плащане. ▪ Наличието на разработена собствена политика, свързана с дигитализацията на общината, все още не е възприето като необходимост. ▪ В общините се работи активно в областта на дигитализацията на процесите и предоставяните административни услуги през последните 3 и повече години. ▪ Възможност потребителите на е-услугите да ползват поне 2 метода за електронна идентификация. ▪ Предизвикателства пред дигитализацията в общините са свързани с: ограничен финансов ресурс; липса на достатъчно квалифицирани експерти; нужда от допълнителна квалификация на персонала и липса на интерес към/невъзможност за ползване на дигиталните услуги от гражданите и бизнеса. ▪ Необходимост от коригиране на честотата на провеждане на анализ на кибер-рисковете. ▪ Разработването на собствени мобилни приложения не е приоритет на общините. 	<p>(Кирилова & Найденов, 2022)</p>
<p>51 общини (анкетно проучване чрез случайна стратифицирана извадка)</p>	<ul style="list-style-type: none"> ▪ Експертите в общините участват активно в обсъждането, анализирането и вземането на решения, свързани с дигитализацията на процеси и услуги в общината, макар че в почти половината от случаите това участие се отбелязва като нерегулярно. Подобна ситуация се наблюдава и по отношение на обмяната на информация за дигитализация на процеси и услуги с експерти от други общини. ▪ Разработена е методика за оценка на дигитализацията на местните административни услуги, като се разглеждат вътрешни и външни фактори на дигитализация. ▪ За външните фактори се отчитат следните констатации: <ul style="list-style-type: none"> ✓ В действащата нормативна база се наблюдават известни пропуски по отношение на отчитане на спецификите на местните административни единици, което поставя под известно съмнение нейната полезност и приложимост за дигитализацията в общините. ✓ Търсенето на ЕАУ от гражданите и бизнеса е сравнително ниско, а значителна част от местните резиденти (особено гражданите) нямат добре развити цифрови умения и компетенции. ▪ За вътрешните се отчитат следните констатации: <ul style="list-style-type: none"> ✓ Технологичната инфраструктура на общините е на сравнително добро ниво от гледна точка на хардуер, софтуер и решения за киберсигурност и защита, но липсва собствен финансов ресурс, който да обезпечи успешната дигитализация. ✓ Липсва достатъчен брой собствени експерти, които да реализират успешно процесите по дигитализация, но е положителен фактът, че сравнително голяма част от служителите там имат нагласа към повишаване на квалификацията, за да могат да работят с новите дигитални решения. ✓ Притеснение буди резултатът, че не повече от половината от предлаганите на потребителя ЕАУ са дигитализирани в резултат на провеждане на предварителен анализ и обмисляне на оптимизацията на работните процеси. 	<p>(Казакова, 2022)</p>

Сходни заключения са направени и при изследване на нивото на дигитализация на общинските услуги, осъществено в рамките на проект на Българската търговско-промишлена палата, а именно - проект BG05SFOP001-2.025-0106-C01 „Дигитална демокрация в действие“. В него обаче анкетни карти са изпратени както на служители в общината (доставчик на ЕАУ), така и на бизнеса и на гражданите (ползватели на тези услуги) с цел получаване на обратна връзка и своевременно откриване на пропуски и несъответствия. Извадката включва 118 общини, а резултатите от проведеното проучване могат да се обобщят в следните няколко аспекта (Българска търговско-промишлена палата, 2022):

- Съществува правопрпорционална зависимост между големината на общината и степента, до която е достигнала дигитализацията на предлаганите услуги.
- Наличието на общинска програма за дигитализация или друг подобен управленски документ предполага, че общината е в по-напреднал етап на дигитализация и предлагане на ЕАУ.
- През последните години дигитализацията в общините е насочена предимно към улесняване на достъпа за ползвателите на ЕАУ, както и за автоматизация на вътрешни работни процеси. От своя страна обаче гражданите и бизнеса не са използвали активно тези услуги, като сред основните причини за това се посочват липсата на достатъчна информираност за предоставяните ЕАУ, недоверието в тях и прекалено ограниченото портфолио от предлагани услуги.
- Пандемията от Ковид-19 се оказва съществен катализатор за създаването на нови административни услуги и въвеждане на нови технологии, позволяващи осъществяването на работа от разстояние за общинските служители.
- Ефектът от дигитализирането на предлаганите административни услуги, както и тяхното качество се оценяват позитивно от всичките три групи респонденти.
- Сред ползите от дигитализирането на административните услуги се посочват: съкращаване на времето за предоставяне/получаване на услугата; повишаване на качеството и по-бързия обмен на данни между участниците в процеса.
- Дигитализирането на административните услуги крие и предизвикателства, сред които са изтъкнати: дигитална неграмотност на гражданите; технологични проблеми; недостатъчни финансови възможности за поддържане на устойчиви ЕАУ; липса на квалифициран персонал, особено ИТ специалисти; нежелание на гражданите за ползване на ЕАУ; ЕАУ не са подходящо проектирани, така че да бъдат използвани и от маргинализирани групи от хора (възрастни, хора със специални потребности, социално-слаби).

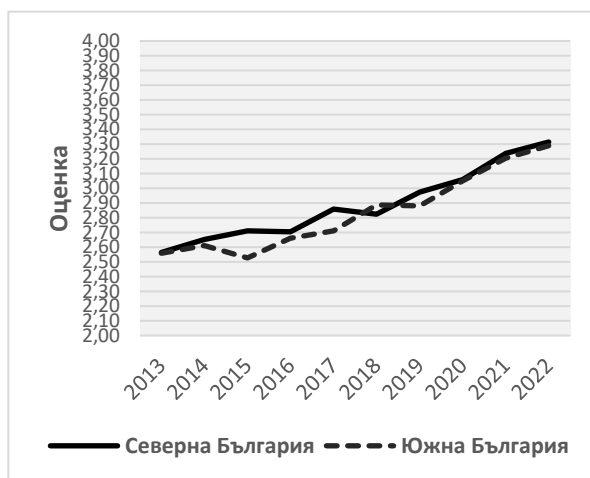
Направеният обзор на разработки, разглеждащи процесите по дигитализация в общините и предоставянето на ЕАУ, дава основание да обобщим, че се наблюдава напредък в тази посока. До известна степен той е резултат и от повсеместната дигитализация, настъпила в резултат на пандемията от Ковид-19 (Калчева, 2020). Въпреки това, все още се отчитат съществени предизвикателства, на които трябва да се обърне сериозно внимание, както от властите на местно, така и на централно ниво. Липсата на интерес на гражданите и бизнеса към ЕАУ не е резултат само от ниската цифрова грамотност на населението, тъй като България е на предпоследно място в ЕС по процент от населението с поне базови цифровия умения с едва 31% (European Commission, 2022). Тази незаинтересованост вероятно е продиктувана и от недостатъчните информационни кампании, както и от липсата на достатъчно ЕАУ с възможност за двупосочна комуникация. В научната литература вече се говори за преминаване от модел на поискване на ЕАУ (pull) към модел на изпращане на ЕАУ (push), което се възприема като проактивна мярка (Scholta & Lindgren, 2023). Проблемът с осигуряването на квалифициран персонал в общините е още по-съществен. Човешкият капитал е фактор от изключително значение и заема важно място в цялостния процес по дигитализация (Кирилов & Казакова, 2022), защото тя не може да се случи само с ИКТ. Необходими са хора с умения и дигитални компетенции, които да анализират и оптимизират действащите работни процеси и да

имплементират новите технологии, за да се предоставят качествени ЕАУ. Заедно с това трябва да се има предвид, че развитието на е-услугите в организациите от публичния сектор има своите специфики и се различава в известна степен от развитието на е-услугите в бизнес организациите (Бонева, 2020). Не на последно място, общинските администрации отчитат, че разполагат със сравнително добра обезпеченост от гледна точка на хардуер и софтуер, но продължава да се наблюдава тенденцията компютрите от висок клас да са разположени предимно в централната администрация без да се взема под внимание обстоятелството, че именно местните администрации са основен доставчик на административни услуги (Министерски съвет, 2022).

Териториални различия при предлагането на ЕАУ между Северна и Южна България

След извършения преглед на състоянието на дигитализацията на общинските административни услуги, в настоящата точка е направен опит да се открият сходствата и различията по отношение на този процес в местната администрация в Северна и Южна България. За целите на анализа са обхванати два масива от данни: общодостъпни данни от независими източници и данни от проведено собствено анкетно проучване.

Интерес за настоящото изследване представлява индикаторът „Развитие на електронното правителство“ от раздел „Администрация“ на регулярното изследване „Регионални профили: показатели за развитие“, което предоставя информация за развитието на областите в България. Споменатият показател дава възможност на местните власти да направят самооценка относно степента и обхвата на предоставяните от тях електронни услуги. Скалата за оценяване е от 1 до 4, а колкото по-висока е оценката, толкова по-добре са развити предлаганите ЕАУ.

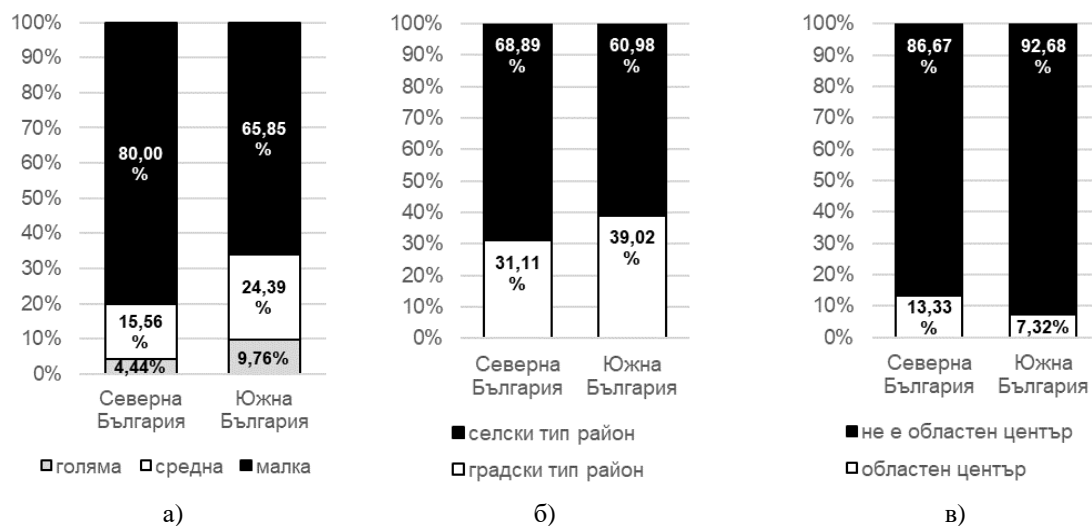


Source: ИПИ, Регионални профили (www.regionalprofiles.bg) и изчисления на автора

Фигура 1: Оценка за развитието на електронното правителство по географски райони

На фиг. 1 е представена оценката за нивото на развитие на електронното правителство за период от 10 години. Видно е, че няма съществени различия при самооценката на предоставяните ЕАУ от местните власти в Северна и Южна България. Този факт не е неочакван с оглед на целенасочено провежданата европейска и държавна политика в тази насока през последните години. Като допълнение към това може да се отчете, че от 2020 година стойностите на този индикатор за двата географски района са почти идентични. Същевременно обаче, изхождайки от първичните данни по области, се открояват известни различия – области като Видин, Плевен и Силистра бележат сериозен напредък при предлагане на ЕАУ, докато такива като Кюстендил, Ямбол и Пазарджик не дават явна индикация за системно повишаване на стойността на показателя (Бошнаков & Казакова, 2022).

За целите на проучване на състоянието на ЕАУ в общините е проведено емпирично изследване в периода януари-февруари 2023 година. Разработената анкета включва петнадесет въпроса, които са обединени в пет групи. Анкетната карта е дигитализирана чрез Google Forms и е разпратена до електронните адреси на всички 265 общини в страната. Получени са 86 коректно попълнените анкети от общински администрации (45 от Северна България и 41 от Южна България), което представлява 33% от генералната съвкупност. Резултатите от проведеното онлайн проучване са разгледани по-детайлно в следващите няколко абзаца.



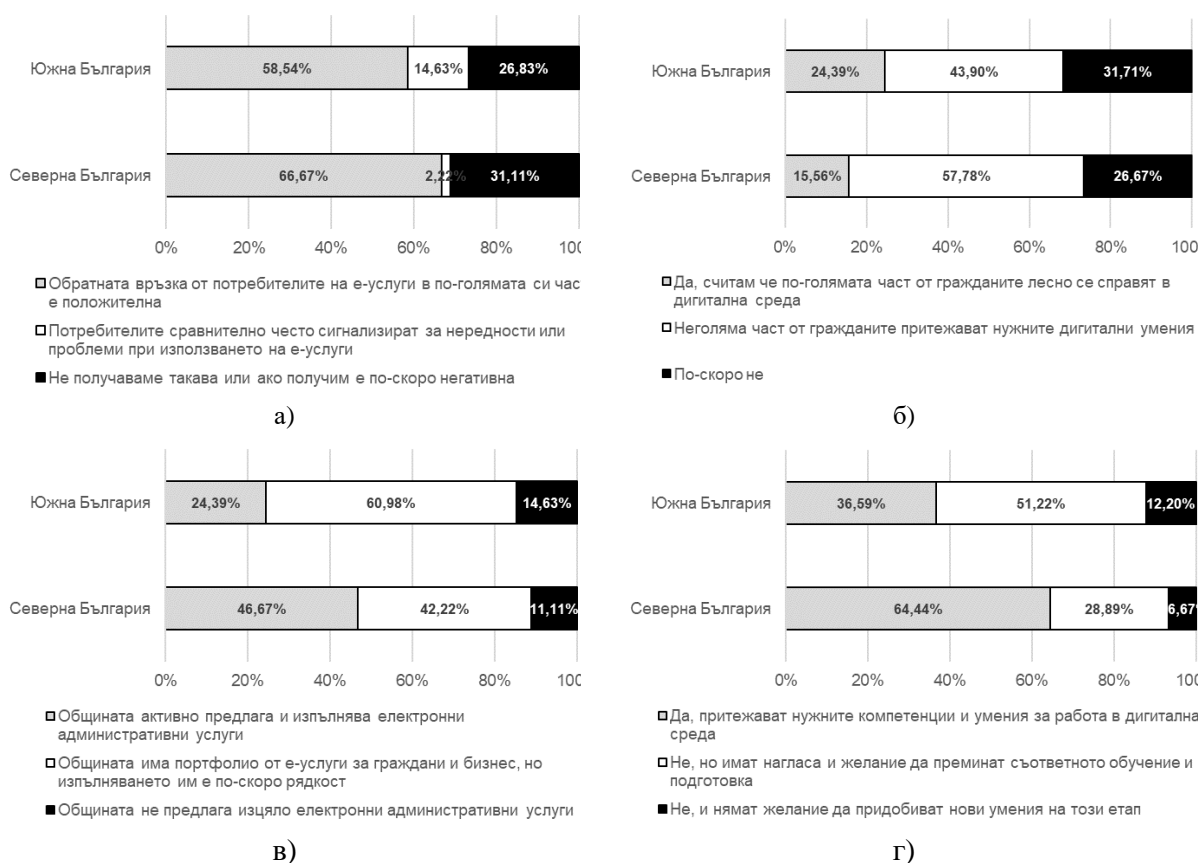
Source: Авторско изследване

Фигура 2: Профил на респондентите по отделни признаци

На фиг. 2 е представена кратка характеристика на профила на респондентите по няколко признака – категория на общината (фиг. 2а), тип район (фиг. 2б) и дали е областен център (фиг. 2в). Групирането на общините по категории се базира на категоризация на Национално сдружение на общините в Република България (НСОРБ), която те използват при методически разработки, номинации и др. В Стратегическия план за развитие на земеделието и селските райони 2023-2027 г. като „селски райони“ се определят общините, в които няма населено място с население над 15 000 души.

Видно е, че в наблюдаваната съвкупност от респонденти присъстват административни единици от всички групи, както малки (до 20 000 жители), така и средни (между 20 000 и 60 000) и големи (над 60 000 жители); както от градски, така и от селски тип район; както от областните центрове, така и от провинциални общини. По този начин наблюдаваната съвкупност обхваща различни категории общини и дава представа за развитието на е-управлението на различни нива и в различни аспекти за общините в Северна и Южна България.

Затвърждава се фактът, установен и в предходни изследвания, че ползвателите на ЕАУ нямат сериозен интерес към тях. За това свидетелстват отговорите на въпроса, свързан с честотата на получаване на заявки за онлайн услуги. В над 70% от случаите респондентите отговарят, че това се случва сравнително рядко, като стойностите имат известно различие - 71,11% за Северна България и 78,05% за Южна България. Въпреки че частта на гражданите и бизнеса, които ползват тези услуги, не е голяма, обратната връзка, която те предоставят в по-голямата си част е положителна т.е. ползвателите са доволни от наличните ЕАУ, като тук се наблюдава по-добър резултат за Северна България (фиг. 3а). Това до някъде може да се обвърже и с резултата от направения по горе анализ на показателя за развитието на електронното правителство, при който оценката за предоставяните ЕАУ в Южна България все още е по-ниска от тази за Северна България. Същевременно ползвателите на ЕАУ в Южна България са по-склонни да сигнализират при нередности или проблеми при използване на съответните електронни услуги.



Source: Авторово изследване

Фигура 3: Териториални различия, свързани с дигитализацията на административните услуги

Получените резултати по отношение на необходимите дигитални умения на гражданите за потребяване на ЕАУ, демонстрират, че такива по-скоро липсват или само малка част от гражданите ги притежават (фиг. 3б). Отчетени са високи стойности, както за Северна (84,44%), така и за Южна България (75,61%), което отново поставя на дневен ред въпроса за необходимостта за повишаване на цифровите умения и компетенции на гражданите, като предпоставка за успешна реализация на е-управление в страната. Интересна е констатацията, че почти една четвърт от служителите от общинските администрации в Южна България са на мнение, че по-голямата част от гражданите могат да се справят успешно с дигиталните услуги.

Анализът на обобщените данни за въпроса, свързан с портфолиото от предлаганите ЕАУ и тяхното изпълнение, дава основание да се твърди, че общините от Северна България активно предлагат и изпълняват ЕАУ, докато това е по-скоро рядкост за тези в Южна България (фиг. 3в). Притеснение буди констатацията, че все още над 10% от анкетираните от двата географски района не предлагат изцяло електронни услуги, което се явява пречка пред успешната реализация на дигитализацията в общинската администрация.

Положителен е фактът, че около 80% от респондентите са отбелязали, че във вътрешната нормативна уредба на съответната община са засегнати въпроси за предоставянето на ЕАУ, а около 14% демонстрират готовност за актуализация на вътрешни документи с цел да бъдат регламентирани и въпросите за предоставяне на ЕАУ. Регистрираните стойности по този въпрос са почти идентични за Северна (77,78%; 13,33%) и Южна България (80,49%; 14,63%).

На този етап преминаването към мобилно електронно управление може да се определи по-скоро като пожелание, отколкото като реалност. Около 30% от анкетираните и от двата географски района са на мнение, че на този етап не е необходим достъп до предлаганите от тях ЕАУ чрез

мобилно приложение. Въпреки това резултатите показват, че е налице желание за включване и ползване на тези услуги през мобилни приложения, а на този етап общините от Северна България са в малко по-напреднала фаза за предоставяне на такива приложения (22,22% срещу 14,63%).

Наличието на квалифицирани кадри, които притежават умения и компетенции за работа в дигитална среда е от съществено значение за качествено предоставяне и изпълнение на ЕАУ от общините. По отношение на този въпрос се наблюдават съществени различия между Северна и Южна България (фиг. 3г). Оказва се, че служителите от общините в Южна България не притежават нужните дигитални умения, но имат желанието и нагласата да преминават съответните обучения и подготовка, които да повишат техните компетенции за работа в дигитална среда. В тази връзка отново трябва да подчертаем същественото значение на въпроса за необходимостта от непрекъснати обучения на служителите в общините и отделянето на достатъчен финансов ресурс за организиране или посещаване на такива обучителни курсове и семинари.

Представените данни и анализи свидетелстват за съществуването на реален напредък по отношение на дигитализацията на административните услуги в общините и развитието на е-управлението в България. ЕАУ са пътя към постигане на по-добро, по-бързо и по-качествено административно обслужване на гражданите и бизнеса от местната администрация. Същевременно пред тяхното активно предлагане и използване все още продължават да стоят два съществени проблема. На първо място, сравнително слабото търсене на тези услуги от гражданите и бизнеса, било то поради незаинтересованост, незнание, недоверие или отсъствие на дигитални умения и компетенции. На второ място, недостигът на квалифицирани експерти в общинската администрация, което от своя страна затруднява нормалното протичане и дори забавя процеса по дигитализация на предлаганите административни услуги. В допълнение към посоченото, докладът установява съществуването на известни регионални различия по отношение на дигитализацията на услугите в общините от Северна и Южна България. Въпреки това, те не са толкова съществени, а наблюдаваните положителни тенденции може да се приемат, като източник на оптимизъм, за движение в правилната посока – към дигитализиране на административните услуги.

References

1. Бонева, М. (2020). Изследване на тенденциите за развитие на електронните услуги. *Годишник на Стопанска академия, СХХІІІ*, 91-118.
2. Бошнаков, В., & Казакова, М. (2022). Използване на дигитални услуги от населението в България: основни предпоставки, тенденции и регионални измерения. *Народостопански архив*(2), 42-54.
3. Българска търговско-промишлена палата. (2022, 12 1). *Аналитичен доклад по резултатите от проведеното изследване за пречките и предизвикателствата пред дигиталната трансформация в работата на общините в полза на гражданите и бизнеса*. Retrieved from https://www.bcci.bg/bulgarian/projects/Doklad_Digitalni_uslugi_v_obshtinite.pdf
4. ДАЕУ. (2021). *Актуализирана Стратегия за развитие на електронното управление в Република България 2019 – 2025 г.* София: ДАЕУ. Retrieved from <https://www.strategy.bg/FileHandler.ashx?fileId=24869>
5. Казакова, М. (2022). *Статистическо изследване на дигитализацията на местните административни услуги*. Докторска дисертация, УНСС, Статистика и иконометрия, София. Retrieved from <https://ras.nacid.bg/dissertation-preview/71305>
6. Калчева, Д. (2020). Е-услугите в публичната администрация и пандемията COVID-19. *ПУБЛИЧНИ ПОЛИТИКИ.bg*, 11(4), 52-61.
7. Кирилов, Р., & Казакова, М. (2022). Човешкият капитал като фактор на успешната дигитализация на местните административни услуги. *Бизнес управление*(2), 52-64.

8. Кирилова, К., & Найденов, А. (2022). Развитие на дигиталните административни услуги в Република България. *Икономически и социални алтернативи*(4), 5-17. doi:<https://doi.org/10.37075/ISA.2022.4.01>
9. Лазарова, Е., Веселинова, Н., Господинов, Ю., & Стоянова, М. (2023). Стратегически аспекти на електронното административно обслужване в общините в Република България. *Алманах научни изследвания "Наука, знание, иновации - синергия за устойчиво развитие"*, 31, 237-267.
10. Марков, К. (2022). Нова стратегическа визия на електронното управление в Република България за периода 2020 – 2025 г. *Икономически и социални алтернативи*(1), 27-34. doi:<https://doi.org/10.37075/ISA.2022.1.03>
11. Министерски съвет. (2022). *Доклад за състоянието на администрацията*. София. Retrieved from https://government.bg/files/common/DSA_2022AR.pdf
12. Министерство на транспорта, информационните технологии и съобщенията. (2019). *Национална програма "Цифрова България 2025"*. София. Retrieved from https://www.mtc.government.bg/sites/default/files/uploads/it/cifrova_bulgariya_2025.pdf
13. Павлова, Д. (2017). Бъдещето на държавната администрация в условията на електронно управление. *Обществото на знанието и хуманизмът на XXI век* (pp. 679-685). София: За буквите.
14. Павлова, Д. (2018). Роля на Е-управлението върху възможностите за участие на гражданите в социалните процеси. *Европейските граждани и интелектуалната собственост* (pp. 335-342). София: За Буквите.
15. *European Commission*. (2022). Retrieved from Bulgaria in the Digital Economy and Society Index: <https://ec.europa.eu/newsroom/dae/redirection/document/88740>
16. Khan, A., Krishnan, S., & Dhir, A. (2020). Electronic government and corruption: Systematic literature review, framework, and agenda for future research. *Technological Forecasting and Social Change*, 167, 1-19. doi:<https://doi.org/10.1016/j.techfore.2021.120737>
17. Scholta, H., & Lindgren, I. (2023). Proactivity in digital public services: A conceptual analysis. *Government Information Quarterly*, 40(3), 1-12.

ДИГИТАЛИЗАЦИЯ ВЪВ ФИНАНСОВИЯ СЕКТОР

Аглика Кънева

Гл. ас. д-р катедра „Финанси”, УНСС

Резюме

В доклада е изследвано въздействието на дигитализацията и технологиите в областта на финансовите пазари, застраховането и частното пенсионно осигуряване. Извеждат се основните ползи от дигиталните технологии и техните приложения.

Ключови думи: дигитализация, финансов сектор, финтех, технология на разпределената счетоводна книга.

JEL: G2, O14, O3.

DIGITALISATION IN THE FINANCIAL SECTOR

Aglika Kaneva

Abstract

In this paper, the impact of digitalisation and technology in the areas of financial markets, insurance, and private pensions has been examined. The major benefits that digital technology and its applications bring have been outlined.

Key words: digitalisation, financial sector, fintech, distributed ledger technology.

Увод

Технологиите и цифровизацията бързо променят начина, по който функционира финансовият сектор. Иновативните приложения на дигиталните технологии за финансови услуги, или финтех, се използват за промяна на интерфейса между финансовите потребители и доставчиците на услуги и спомагат за подобряване на комуникацията с потребителите и повишаване на тяхната ангажираност¹.

Новите и нововъзникващи технологии, които се прилагат във финансовите услуги, включват технологията на разпределената счетоводна книга (DLT), големи данни, интернет на нещата (IoT), изчисления в облак, изкуствен интелект (AI), биометрични технологии и разширена и виртуална реалност. Много от тези технологии са взаимно зависими. Например, изкуственият интелект се развива благодарение на големите данни, изчислителните облаци и все повече на интернет на нещата².

Новите и нововъзникващи технологии и техните приложения във финансовите услуги имат потенциала да повлияят на множество аспекти в областта на финансите, включително нови бизнес модели и продуктови дизайни, конкуренция, оперативна ефективност, посредничество, достъпност, ангажираност на потребителите, скорост, автоматизация, анализи, поверителност и прозрачност и риск за дигиталната сигурност³.

Основната цел на доклада е да се изследва дигитализацията във финансовия сектор.

За да се реализира целта, в доклада се очертават следните конкретни задачи:

1. Анализ на влиянието на дигитализацията върху финансовия сектор.
2. Представяне на новите технологии и техните приложения при финансовите услуги.

¹ See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 3

² See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 13

³ See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 8

3. Изследване на ползите от дигиталните технологии и техните приложения.

Обект на изследването е дигитализацията във финансовия сектор. Предмет на изследването са приложенията на новите технологии при финансовите услуги.

Основната теза на изследването е, че новите технологии и техните приложения значително ще подобрят ефективността на действие на финансовия сектор.

Определения за финтех

Някои определения за финтех включват:

- "нововъзникващи цифрови технологии, прилагани във финансовия сектор" (НКМА)⁴;
- "финансови иновации с помощта на цифрови технологии" (FSB)⁵;
- "иновации в областта на финансовите технологии" (US National Economic Council).

Приложения на дигиталните технологии във финансовите услуги

Технологичният напредък и новите иновации оказват влияние върху цялата верига за създаване на стойност на финансовите продукти и услуги и имат множество приложения. Представената тук рамка класифицира обхванатите приложения в осем отделни категории: плащания, планиране, кредитиране и финансиране, търговия и инвестиции, застраховане, киберсигурност, операции и комуникации.

В таблица 1 се посочват категориите финансови дейности и услуги, в които се прилагат избраните цифрови технологии.

Таблица 1 Приложения на нови технологии във финансовите услуги

Цифрова технология	Финансови дейности и услуги							
	Платежни услуги	Консултантски и агентски услуги Планиране	Инвестиции и търговия	Кредитиране и финансиране	Застраховане	Сигурност	Операции	Комуникации
Технология на разпределената счетоводна книга	X	X	X	X	X	X	X	X
Големи данни		X	X	X	X	X	X	X
Интернет на нещата					X			X
Изчисления в облак				X			X	
Изкуствен интелект		X	X		X			X
Биометрични технологии					X	X		
Разширена /Виртуална реалност		X	X					X

Източник: OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 14

Както е видно от таблица 1, някои цифрови технологии имат широко приложение, докато други остават по-ограничени, но всички те имат потенциала да окажат значително въздействие върху финансовите услуги.

- **Плащания** - плащанията представляват най-основното приложение на цифровите технологии във финансовите услуги, което, макар и да не е ново, се развива заедно с нововъзникващите технологии. Въпреки че цифровите плащания започнаха с физически инструменти (напр. кредитни карти), те все повече преминават във виртуалната сфера. Тези иновативни платежни услуги могат да бъдат разделени на онлайн плащания и мобилни плащания. Онлайн плащанията се определят като платежни нареждания, които се подават с помощта на устройства, свързани с интернет, а мобилните плащания - като

⁴ Hong Kong Monetary Authority, Whitepaper on Distributed Ledger Technology, 11 Nov 2016, p. 5

⁵ FSB FinTech Issues Group Interim Report, 2017

такива, които разчитат на устройства, свързани към мобилна комуникационна мрежа⁶. Следователно онлайн плащанията включват онлайн банкирането, електронна търговия (напр. Amazon) и платежни услуги (напр. PayPal).

Мобилните плащания включват транзакции с мобилни пари чрез мобилни мрежови оператори (напр. плащания чрез SMS) и предплатени карти, свързани към мобилни телефони. Плащанията не се ограничават само до банковия сектор. Приложения за мобилни плащания съществуват и в застраховането, където регистрацията и плащанията по застраховките могат да се извършват с помощта на мобилно устройство⁷.

В други случаи цифровите транзакции се използват в нови бизнес модели, създадени за улесняване на трансграничните плащания, а именно чрез платформи за обмен на валута от типа peer-to-peer, които могат да свързват купувачи и продавачи на валута, за да елиминират разликата в обменния курс.

Иновативните приложения за плащания използват технологията на разпределената счетоводна книга. Технологията на разпределената счетоводна книга е протокол, използван за изграждане на счетоводна система за съхраняване на записи, например такива, свързани със собственост, транзакции или договорни споразумения. Технологията на разпределената счетоводна книга обаче не се контролира централно от една страна или посредник, а споделя отговорността за добавянето на информация и поддържането на счетоводната книга с всички участници. Всеки участник разполага със собствено идентично копие на счетоводната книга и всяко ново допълнение към нея трябва да бъде одобрено и съгласувано от всички участници. Счетоводните книги се формират чрез поредица или "верига" от информационни блокове. Когато транзакцията бъде одобрена от участниците, във веригата на транзакциите се образува нов блок. Тази информация се записва трайно в счетоводната книга и не може да бъде подправена⁸. Технологията на разпределената счетоводна книга, наричана още блокчейн, което е най-често използваната ѝ форма, е технология за бази данни, която позволява създаването, сигурното прехвърляне (с окончателност) и съхранение на информация⁹.

Първото приложение за плащане, което стана възможно благодарение на тази технология, беше криптовалутата биткойн, а оттогава технологията се използва и за други криптовалути като етер. Криптовалутите могат да се използват като обикновена валута и да се управляват с цифрови портфейли, съхранявани в смартфон. Всички транзакции се записват постоянно в блокчейна, а нова валута може да се генерира от "миньори", които успеят да решат необходимия алгоритъм.

Интелигентните договори са по-нова разработка в областта на използването на технологията на разпределената счетоводна книга за плащания, но все още не са предизвикали същите опасения като другите форми на плащане, свързани с технологията на разпределената счетоводна книга. Тези споразумения са самопринудителни и автоматично изпълняват транзакция, когато са изпълнени определени условия. Такива споразумения могат да се използват например за улесняване на суапови плащания. Те се използват и в застрахователния сектор за автоматично изплащане на застрахователни обезщетения при настъпване на застрахователно събитие, например закъснял полет¹⁰. Разработват се и допълнителни приложения за застраховането, като например автоматични плащания за суапове за природни бедствия или оптимизиране на плащанията за международен фронтинг за каптивни застрахователи¹¹. Интелигентните договори могат да се комбинират и с технологията на интернет на нещата, за да се свързват автоматично свързаните устройства със съответната застрахователна полица¹². Интернет на нещата се отнася

⁶ See FinCoNet, Online and mobile payments: Supervisory challenges to mitigate security risks, September 2016, p. 15

⁷ See OECD, Technology and innovation in the insurance sector, 2017, p. 16

⁸ Hong Kong Monetary Authority, Whitepaper on Distributed Ledger Technology, 11 Nov 2016, p. 9

⁹ See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 11

¹⁰ See OECD, Technology and innovation in the insurance sector, 2017, p. 20

¹¹ See Allianz, Blockchain technology successfully piloted by Allianz Risk Transfer and Nephila for catastrophe swap, Press Release, June 15, 2016, <https://commercial.allianz.com/news-and-insights/news/blockchain-piloted-allianz-risk-transfer.html>

¹² See Norton Rose Fulbright, The future of smart contracts in insurance, September 2016,

до многобройните свързани устройства, които събират информация за движението и други сензорни данни за обекти във физическия свят, и се очаква да представляват все по-голям източник на големи данни. Интернет на нещата може да предостави богата информация за поведението на хората; по този начин получените данни могат да се използват за по-добро адаптиране на продуктите, профилиране на риска и ценообразуване¹³.

Като се имат предвид нейните приложения, технологията на разпределената счетоводна книга има потенциала значително да подобри ефективността на операциите във финансовия сектор. Въпреки това неизменността на базовия код и последващата необратимост на транзакциите биха могли да създадат потенциални проблеми за финансовите транзакции, тъй като в крайна сметка точността на базовия код все още е изложена на човешка грешка¹⁴.

- **Консултации и планиране** - възходът на цифровите платформи предоставя многобройни възможности за ангажиране и подпомагане на потребителите при тяхното финансово планиране. Уебсайтовете и онлайн курсовете могат да улеснят достъпа на потребителите до финансова информация и обучение. Консултациите и планирането също са улеснени от разширената реалност, която може да се комбинира с игровизация, за да се подобри учебният опит¹⁵. Разширената и виртуалната реалност предоставят на потребителите нови начини за възприемане или взаимодействие със заобикалящата ги среда. Разликата между двете е, че разширената реалност предоставя подобрен изглед на действителния физически свят, в който се намират хората, докато виртуалната реалност създава симулиран свят¹⁶.

Обучението също може да стане по-интерактивно чрез игровизация, което може да помогне за развитието на компетентността и увереността на потребителите в управлението на финансите им. Бюджетните инструменти и инструментите за пенсионно планиране могат да помогнат на потребителите да планират по-добре своите разходи и спестявания. Други приложения, като например инструменти за самозадължаване, които насърчават потребителите да спестяват, могат да им помогнат да се справят със собствените си поведенчески предразсъдъци. Такива цифрови инструменти в подкрепа на финансовото образование са широко разработени¹⁷.

Дигиталните платформи може да се използват и, за да помогнат на потребителите да следят финансите си. В някои страни са разработени пенсионни "табла", които дават възможност на бенефициентите да се запознаят с всички свои източници на пенсионни доходи и да предприемат съответните действия, като например да коригират размера на вноските си¹⁸.

Цифровите платформи могат също така да помогнат на потребителите да сравняват финансови продукти и да вземат решение за инвестиране в тях. Разработени са платформи за сравнение, които помагат на потребителите да избират застрахователни продукти, ипотечни кредити и инвестиционни/спестовни продукти. И двете се прилагат най-вече в областта на финансовото образование и услугите за финансово планиране.

Консултациите и планирането също са улеснени от разширената реалност, която може да се използва за подобряване на учебния опит чрез геймификация¹⁹. Разширената реалност може да помогне за по-лесното възприемане на сложни масиви от данни чрез визуализация на данни. За потребителите добавената реалност се използва, за да помогне на инвеститорите да визуализират своите инвестиции и да улеснят инвестиционните си решения²⁰.

<https://www.nortonrosefulbright.com/en/about/our-firm>

¹³ See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 13

¹⁴ See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 12

¹⁵ See 6. G20, OECD, INFE, Ensuring Financial Education and Consumer Protection for All in the Digital Age, 2018, pp. 52-54

¹⁶ See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 14

¹⁷ See G20, OECD, INFE, 2018, pp. 52-54

¹⁸ See OECD, Technology and pensions: The potential for FinTech to transform the way pensions operate and how governments are supporting its development, 2017, pp. 7-8

¹⁹ See G20, OECD, INFE, 2018, p. 52

²⁰ See Virtual and Augmented Reality to Disrupt Banking and Finance, October 6, 2023, <https://www.oxagile.com/article/virtual-augmented-reality-disrupt-banking-finance/>

Друго приложение на дигиталните технологии в сферата на консултирането и планирането е под формата на робоконсултанти, които започват да използват технологии с изкуствен интелект за обслужване на клиенти и за предоставяне на инвестиционни съвети въз основа на активността по сметката на дадено лице. Подобластите на изкуствения интелект могат да се фокусират върху редица различни аспекти на човешката интелигентност, включително разпознаване, разбиране, учене, решаване на проблеми, аргументиране и вземане на решения²¹. Изкуственият интелект често се използва във връзка с машинното обучение, при което машините се обучават с исторически данни, за да разпознават модели и да класифицират нови данни. Чрез усъвършенствани алгоритми машината може да научи модели с нов опит, за да подобри работата си. Въпреки това машината не се обучава напълно самостоятелно; по-скоро процесът на обучение изисква значително ниво на човешки принос, за да се гарантира, че данните се интерпретират правилно²².

Изкуственият интелект се използва и за подкрепа на потребителите, като чатботовете могат интерактивно да отговарят на въпроси, които потребителите могат да имат относно техния продукт или услуга.

Приложения с виртуална реалност, подобни на тези с добавена реалност, са разработени и в приложения за финансови консултации, които предоставят виртуални консултации с банкови или застрахователни експерти²³.

- **Инвестиции и търговия** - дигиталните технологии се използват и за създаване на нови и/или по-ефективни начини за достъп и оптимизиране на търговията и инвестициите. Например платформите за директна търговия и инвестиции улесняват достъпа до пазарите както за институционалните инвеститори, така и за потребителите на дребно. За институционалните инвеститори тези платформи намаляват зависимостта от маркет мейкърите за целите на търговията. За инвеститорите на дребно търговията и инвестирането могат да се извършват на много по-ниска цена, отколкото чрез посредник, а някои платформи дори предлагат готови професионално разработени портфейли²⁴. Друг пример са социалните платформи за търговия. Те могат да позволят на инвеститорите автоматично да копират стратегиите за търговия на търговците, които са избрали да следват.

На базово ниво прилагането на алгоритми в търговията и бързината, с която могат да се изпълняват сделките, позволиха високочестотна търговия. Алгоритмичната търговия обаче вече се разширява с нови приложения.

Платформите за роботизирани съвети предлагат услуги за управление на инвестиции и портфейли, които могат автоматично да търгуват, за да поддържат желаните рискови профили на портфейлите или да реализират инвестиционни загуби за данъчни цели²⁵. Платформите за роботизирани съвети могат също така да използват алгоритми, за да препоръчат определена инвестиционна стратегия предвид профила или риска на инвеститора. Подобни услуги, предлагани на фирми, могат да им помогнат да управляват риска на портфейла или да оптимизират управлението на активите и пасивите.

Изкуственият интелект има и потенциални приложения за оптимизиране на инвестициите чрез разпознаване на модели и прогнозиране на това кои инвестиции ще бъдат с високи резултати в бъдеще. Технологиите вече се прилагат при хедж фондове²⁶. Някои хедж фондове вече използват изкуствен интелект в своите инвестиционни модели. Изкуственият интелект се използва и за

²¹ See Rao A. S., Five Myths and Facts about Artificial Intelligence, Predictive Analytics and Futurism, Issue 14, December 2016, pp. 14-17

²² See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 13

²³ See MetLife, MetLife launches virtual reality in India, March 8, 2017, <https://www.metlife.com.au/blog/blog-home/industry-insights/launch-virtual-reality-in-india/>

²⁴ See OECD, Robo-advice for pensions, 2017, p. 7

²⁵ See OECD, Robo-advice for pensions, 2017, p. 7-10

²⁶ See Satariano A., Silicon Valley Hedge Fund Takes on Wall Street With AI Trader, Bloomberg, 6 February 2017, <https://www.bloomberg.com/news/articles/2017-02-06/silicon-valley-hedge-fund-takes-on-wall-street-with-ai-trader>

интегриране на променливи, свързани с околната среда, социалната сфера и корпоративното управление, в инвестиционната стратегия²⁷.

- **Кредитиране и финансиране** - приложенията на цифровите технологии в операциите по отпускане на кредити и финансиране спомагат за предоставянето на кредити на лица или предприятия, които преди това са имали затруднен достъп до традиционния кредитен пазар.

Появяват се нови бизнес модели за отпускане на заеми и финансиране, най-вече платформи от типа "peer-to-peer". Тези платформи от типа "peer-to-peer" могат да използват един от няколко различни модела. При платформите, базирани на дарения и възнаграждения, физическите лица, които финансират контрагента, директно даряват средства, например за социална кауза или в замяна на бъдещо възнаграждение (обикновено продукт или услуга, генерирани от финансирания проект). При платформите, базирани на заеми, физическите лица очакват да получат възвръщаемост на инвестицията си. При платформите, базирани на дялово участие, лицата получават дял в компанията, която набира средства.

- Големите данни също така дават възможност за по-добра оценка на рисковете, свързани с кредитирането. В практиката се появяват нови методи за изчисляване на кредитни точки въз основа на нетрадиционни променливи, като например социални мрежи, които усъвършенстват тези точки, за да представят по-добре рисковите характеристики на дадено лице. Цифровизацията на ежедневните дейности драстично увеличи обема на наличните данни, създавайки изключително големи и сложни масиви от данни, обикновено наричани "големи данни". Такива данни се получават не само от текстови или цифрови форми, но и от изображения, видео- и аудиоклипове, както и от данни, генерирани от комуникационни и други устройства (напр. смартфони, персонални компютри, свързани с интернет). Бързият напредък в информационните технологии вече позволява обработката и анализа на такива големи масиви от данни. Големите масиви от данни могат да се използват във всяка точка от веригата на стойността на финансовите продукти - от създаването до продажбата. Три понятия, тясно свързани с големите данни, са интернет на нещата, изчислителни облаци и изкуствен интелект. Интернет на нещата е източник на големи данни, изчислителните облаци улесняват обработката и съхранението на големи масиви от данни, а изкуственият интелект е усъвършенстван начин за анализиране и използване на големи данни²⁸. Услугите, базирани на облак, осигуряват рентабилен и сравнително лесно мащабируем капацитет за обработка и съхранение на данни при поискване. Облачните технологии значително увеличиха капацитета на финансовите институции да събират и анализират данни, като по този начин улесниха развитието на анализа на данни и различните му приложения²⁹.

В друго приложение на големите данни обработващата плащания фирма Worldpay започна да отпуска заеми на МСП в партньорство с Liberis, базиран в Лондон небанков кредитор за бизнеса, предлагайки необезпечени парични аванси въз основа на прогнозни продажби на карти. Worldpay е в състояние да анализира данните за продажбите, преминаващи през нейната система, за да определи предварително каква сума са в състояние да върнат клиентите и за какъв период от време, като по този начин може да изплаща авансово само суми, които са в рамките на тези ограничения³⁰.

- **Застраховане** - приложенията на дигиталните платформи и новите технологии трансформират и застраховането. Първо, те променят начина, по който се осъществява достъпът до застраховки и тяхното разпространение. Например, доставчиците вече предлагат достъп до застраховки чрез мобилни устройства. Появяват се и платформи за

²⁷ See GlobeNewswire, Sustainalytics Launches ESG Signals, December 14, 2016, <https://www.globenewswire.com/en/news-release/2016/12/14/897471/34801/en/Sustainalytics-Launches-ESG-Signals.html>

²⁸ See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 12-13

²⁹ See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 13

³⁰ See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 17

взаимно застраховане, чрез които хората могат да създадат своя собствена група от лица, с които да обединят риска³¹.

С най-голямо въздействие върху застраховането обаче може да окаже използването на технологиите за подобряване на застраховането и ценообразуването на риска. Големите данни и подобреният анализ на данни, включително изкуственият интелект, увеличават броя на променливите, които могат да бъдат взети предвид при определянето на цената на дадена полица. Въпреки това, макар че това може да повиши прецизността, то също така може да прекрачи границите на целта на застраховането за обединяване на рисковете и да доведе до изключване от застраховане на рискове, считани за "лоши рискове".

Повишената прецизност и ефикасност на подписването на застрахователни договори позволява също така да се предлагат нови видове продукти. Застрахователни договори могат да се издават за много кратки периоди от време, например за конкретни пътувания с автомобил или за краткосрочно наемане на жилище.

Дигиталните технологии променят и начина, по който могат да се сключват застраховки. Например технологията за разпознаване на лица може да се използва за оценка на здравословното състояние и възрастта на дадено лице при сключване на застраховка "Живот". Сензорите и камерите в автомобилите могат да определят поведението на водача на пътя и тази информация може да се използва за определяне на премията за автомобилна застраховка.

В сферата на здравеопазването изкуственият интелект може да се използва и за анализ на снимки, за да се идентифицират определени медицински състояния или здравето на лицата. Успешно е приложен, например, за откриване на рак на кожата по снимка, което е сравнимо с резултатите на дерматолозите³².

- **Киберсигурност** - въпреки че повишената зависимост от дигиталните технологии може да увеличи риска от компрометиране на киберсигурността, дигиталните технологии предоставят и множество възможности за подобряване на сигурността на цифровите финансови услуги. Криптирането на данни за защита на цифрово съхранявани данни се усъвършенства с развитието на технологиите. Биометричната технология може да се използва за подобряване на проверката на самоличността и удостоверяването на автентичността, за да се намали рискът от откраднати пароли или фалшифицирани транзакции³³. Биометричните технологии се основават на разпознаването на физиологични или поведенчески характеристики и могат да се използват за удостоверяване на самоличността чрез откриване на уникални за отделните лица характеристики. Техниките, които сега се използват за проверка, включват сканиране на пръстови отпечатьци, удостоверяване на автентичност чрез глас, разпознаване на лица, сканиране на ириса и разпознаване на походката. Биометричните технологии представляват голямо подобрение в сферата на сигурността в сравнение с проверката чрез пароли. Въпреки това тези технологии все още се разработват и сигурността се подобрява, за да се намали рискът от компрометиране на биометричната информация³⁴.

Анализът на данни може да се използва за откриване на нередовни модели и за установяване на наличието на измама. Технологиите на разпределената счетоводна книга биха могли да увеличат прозрачността на транзакциите, като ги направят по-лесни за проследяване и контрол, а също и да намалят риска от фалшифицирани транзакции³⁵.

- **Операции** - за всички видове финансови институции технологиите имат потенциала значително да подобрят ефективността на процесите и ефикасността, с които те работят, и да намалят общите режимни разходи. Ежедневните процеси могат да бъдат рационализирани чрез повишена автоматизация. Функциите, свързани със спазването на

³¹ See OECD, *Enhancing the Role of Insurance in Cyber Risk Management*, OECD Publishing, Paris, 2017

³² See Esteva A., Kuprel B., Novoa R. A., Ko J., Swetter S. M., Blau H. M., Thrun S., *Dermatologist-level classification of skin cancer with deep neural networks*, *Nature*, vol. 542, 2017, pp. 115–118

³³ See OECD, *Technology and innovation in the insurance sector*, 2017, p. 18

³⁴ See OECD, *Financial Markets, Insurance and Private Pensions: Digitalisation and Finance*, 2018, p. 13-14

³⁵ See OECD, *Technology and innovation in the insurance sector*, 2017, p. 18

нормативните изисквания, могат да бъдат по-ефективни, а за улесняване на спазването на нормативните изисквания се използват технологични приложения, които често се наричат "RegTech". Такива приложения могат да улеснят регулаторните изисквания за отчетност или други процеси на съответствие, например чрез използване на технологии за разпределена счетоводна книга или друга цифрова база данни, за да се улеснят изискванията за познаване на клиента (KYC). Разработват се и приложения за улесняване на функциите по управление на риска³⁶.

- **Комуникации** - новите технологии променят и начина, по който доставчиците на финансови услуги общуват със своите клиенти. Те променят начина, по който се предлагат финансовите продукти. Онлайн рекламите често са насочени към профила на зрителите, който се определя от тяхното онлайн поведение и навици на сърфиране. След като потребителите се ангажират с даден продукт или услуга, редовните комуникации могат да бъдат адаптирани индивидуално към тях, например чрез напомняне с текстово съобщение за вноска в спестовен план или пенсионен фонд или за плащане на сметки. Функциите за подпомагане на потребителите също могат да бъдат трансформирани с помощта на технологии, като например използването на чат ботове или сесии с консултант във виртуална реалност³⁷.

Приложенията на тези нови дигитални технологии оказват значително влияние върху ключови аспекти на финансовите дейности и услуги³⁸. Цифровизацията създава иновации, намалява разходите за клиентите и фирмите и може да насърчи по-голяма конкурентоспособност и избор за предприятията и домакинствата. Пример за това е областта на плащанията. С разрастването на електронната търговия клиентите все по-често купуват и продават стоки и услуги онлайн. И все повече плащания на дребно се извършват чрез мобилни телефони³⁹.

Заклучение

Непрекъснатият напредък в областта на телекомуникациите и компютърните технологии е важен фактор за трансформацията на финансите. Технологичният напредък значително подобри качеството и скоростта на обработка и спомогна за намаляване на разходите за информация и другите разходи за сключване на сделки. Тези промени се отразиха както на доставчиците, така и на потребителите на финансови продукти и услуги⁴⁰.

Дигиталните технологии навлизат все по-повече във финансите, както и в икономиката като цяло. Приложенията на блокчейн са особено обещаващи, особено когато става въпрос за използването им във връзка с други дигитални технологии като изкуствен интелект, големи данни и интернет на нещата.

Блокчейн или в по-широк смисъл технологиите за разпределена счетоводна книга ще се превърнат в един от ключовите двигатели на интернет на бъдещето с потенциал за децентрализация на цифровите приложения и управлението на данни в интерес на гражданите и потребителите. Ще се появят нови икономически възможности за стартиращите предприятия в областта на децентрализираното финансиране, но и за традиционните фирми от финансовия сектор, които се стремят към иновации⁴¹.

Блокчейн или изкуственият интелект създават нова динамика в предоставянето на финансови услуги и въвеждат нови участници в сферата на финансовите услуги. Технологичните фирми могат да достигнат до огромни потребителски мрежи за миг⁴².

³⁶ See OECD, Technology and innovation in the insurance sector, 2017, p. 18

³⁷ See OECD, Technology and innovation in the insurance sector, 2017, p. 18

³⁸ See OECD, Technology and innovation in the insurance sector, 2017, p. 19

³⁹ See Holle L., Digitalisation of financial markets – a key priority, VIEWS - The EUROFI Magazine, Helsinki, 2019, pp. 172-173

⁴⁰ See OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018, p. 13

⁴¹ See Bury C., EU leadership in digital technologies. Can blockchain and FinTech be the EU's game changers?, VIEWS - The EUROFI Magazine, Helsinki, 2019, pp. 173-174

⁴² See Holle L., 2019, pp. 172-173

Цифровите технологии откриват големи възможности, но също така поставят съществени предизвикателства, а крайните им ефекти ще зависят до голяма степен от приемането на подходящи публични политики на национално и международно равнище⁴³.

Международните организации и правителствата са изцяло ангажирани с оценката на предизвиканите от технологиите промени в сферата на финансовите услуги и с определянето на подходящите политически мерки, необходими за справяне с тези промени⁴⁴.

References

1. Allianz, Blockchain technology successfully piloted by Allianz Risk Transfer and Nephila for catastrophe swap, Press Release, June 15, 2016, <https://commercial.allianz.com/news-and-insights/news/blockchain-piloted-allianz-risk-transfer.html>
2. Bury C., EU leadership in digital technologies. Can blockchain and FinTech be the EU's game changers?, VIEWS - The EUROFI Magazine, Helsinki, 2019
3. Dedola L., Ehrmann M., Hoffmann P., Lamo A., Paz Pardo G., Slacalek J., Strasser G., Digitalisation and the economy, ECB Working Paper Series No 2809, Frankfurt, 2023
4. Esteva A., Kuprel B., Novoa R. A., Ko J., Swetter S. M., Blau H. M., Thrun S., Dermatologist-level classification of skin cancer with deep neural networks, Nature, vol. 542, 2017
5. FSB FinTech Issues Group Interim Report, 2017
6. FinCoNet, Online and mobile payments: Supervisory challenges to mitigate security risks, September 2016
7. G20, OECD, INFE, Ensuring Financial Education and Consumer Protection for All in the Digital Age 2018
8. GlobeNewswire, Sustainalytics Launches ESG Signals, December 14, 2016, <https://www.globenewswire.com/en/news-release/2016/12/14/897471/34801/en/Sustainalytics-Launches-ESG-Signals.html>
9. Holle L., Digitalisation of financial markets – a key priority, VIEWS - The EUROFI Magazine, Helsinki, 2019
10. Hong Kong Monetary Authority, Whitepaper on Distributed Ledger Technology, 11 Nov 2016
11. MetLife, MetLife launches virtual reality in India, March 8, 2017, <https://www.metlife.com.au/blog/blog-home/industry-insights/launch-virtual-reality-in-india/>
12. Norton Rose Fulbright, The future of smart contracts in insurance, September 2016, <https://www.nortonrosefulbright.com/en/about/our-firm>
13. OECD, Enhancing the Role of Insurance in Cyber Risk Management, OECD Publishing, Paris, 2017
14. OECD, Financial Markets, Insurance and Private Pensions: Digitalisation and Finance, 2018
15. OECD, Robo-advice for pensions, 2017
16. OECD, Technology and innovation in the insurance sector, 2017
17. OECD, Technology and pensions: The potential for FinTech to transform the way pensions operate and how governments are supporting its development, 2017
18. Rao A. S., Five Myths and Facts about Artificial Intelligence, Predictive Analytics and Futurism, Issue 14, December 2016
19. Satariano A., Silicon Valley Hedge Fund Takes on Wall Street With AI Trader, Bloomberg, 6 February 2017, <https://www.bloomberg.com/news/articles/2017-02-06/silicon-valley-hedge-fund-takes-on-wall-street-with-ai-trader>
20. Virtual and Augmented Reality to Disrupt Banking and Finance, October 6, 2023, <https://www.oxagile.com/article/virtual-augmented-reality-disrupt-banking-finance/>

⁴³ See Dedola L., Ehrmann M., Hoffmann P., Lamo A., Paz Pardo G., Slacalek J., Strasser G., 2023, p. 34

⁴⁴ See OECD, Technology and innovation in the insurance sector, 2017, p. 30

ИЗКУСТВЕНИЯТ ИНТЕЛЕКТ – ОТ ПЕРЦЕПТРОН ДО CHATGPT

Ваня Лазарова

Доцент, доктор, катедра „Информационни технологии и комуникации“ УНСС,
e-mail: vlazarova@unwe.bg

Резюме

В тази разработка, ще се опитаме да проследим пътя на изкуствения интелект от неговата поява, до най-новите му проявления. Ще разгледаме най-значимите литературни източници, свързани с науката за изкуствения интелект. Ще се постараме да опишем основните етапи, по които се е стигнало до съвременното състояние на изкуствения интелект. Също така освен чисто информационни въпроси, ще засегнем и някои въпроси от други научни области, тъй като развитието на изкуствения интелект е свързано с много области от човешкото знание освен информационните системи и технологии - философията, когнитивните науки, статистиката. В разработката ще се фокусираме върху някои ключови за развитието на изкуствения интелект теории и експерименти.

Ключови думи: *изкуствен интелект, перцептрон, генеративни модели*

ARTIFICIAL INTELLIGENCE - FROM PERCEPTRON TO CHATGPT

Vanya Lazarova

Abstract

In this paper, we will try to trace the path of artificial intelligence from its emergence to its latest manifestations. We will look at the most significant literary sources related to the science of artificial intelligence. We will try to outline the main stages by which the modern state of artificial intelligence has been reached. Also, in addition to purely informational, we will touch on some philosophical questions, since the development of artificial intelligence is related to philosophy. In the paper, we will focus on some key theories and experiments for the development of artificial intelligence.

Key words: artificial intelligence, Perceptron, generative AI models

JEL: O3, Z0

Изкуствен интелект (ИИ) и връзка с човешки интелект

Счита се, че за първи път понятието Изкуствен Интелект (ИИ) е употребено през 1956 г. от Джон Маккарти на семинар, проведен в Дортмут колеж в САЩ, където се събрали учени интересувани се от използването на компютри за наподобяване на човешкия интелект.

Използването на понятието ИИ е в два аспекта - за означаване на способността на компютърните системи да извършват дейности, подобни на човешкия интелект; за означаване на научно-приложната област за изследване и изграждане на системи с ИИ. В този доклад ще се спрем на използването на ИИ при изграждане на системи.

През 2004 г. Джон Маккарти (McCarthy, 2004), дефинира понятието така: „Това е наука и инженерни технологии за създаване на интелигентни машини, и по-специално интелигентни компютърни програми. Това е свързано със задачата за използване на компютри за наподобяване на човешкия интелект, но ИИ не трябва да се ограничава само с методи, които са биологично наблюдаеми.“

Дефиницията показва, че начинът изкуственият интелект да се опише, е чрез човешкия интелект, макар да не трябва да се разчита само на биологичните характеристики. Тъй като интелектът отразява само умствените възможности на даден човек, обикновено се използва

понятието интелигентност, което е по-широко, има множество проявления и включва възможността да се чувства, да се разбират взаимовръзки между събития, да се вземат решения.

Компютрите придобиват огромни възможности - изчислителна мощ, милиони операции за секунда, технически характеристики, които ги направят по-бързи от човека. Но дали са по-интелигентни? Маккарти задава въпроса: „Какво е интелигентност?“, защото без отговора на този въпрос, дефиницията не е пълна. Остава неясно коя компютърна програма да се счита за интелигентна.

Интелигентността е понятие, което традиционно се свързва с човека и неговите способности да мисли и да се учи от опита, да решава проблеми и взема решения, да се адаптира към околната среда и да притежава разбиране и контрол върху собствените мисловни процеси. Човешката интелигентност се проверява чрез различни тестове, които формират коефициент за интелигентност (IQ), измерител, предложен още през 1912 г. от Уилям Стърн (Стърн, 1912).

Компютрите, съвкупност от взаимобвързани хардуер и софтуер, работят, като програмите (софтуера) дават инструкции на хардуера как да реагира на въведената информация. Тези инструкции могат да бъдат от прости, до много комплексни – да решават задачи, да дават указания за движение, да съхраняват, обработват и визуализират данни за стоки и услуги, да предсказват въз основа на изчисления върху данни и пр.

Какво ще ни накара да опишем една компютърна програма като интелигентна?

За първи път отговор на този въпрос се опитал да даде Алън Тюринг, през 1950 г., в публикацията си „Компютърни машини и интелект“ (Turing, 1950). Той предложил, вместо да се дава отговор на въпроса дали една компютърна програма е интелигентна, да се подходи по друг начин, чрез тест. Идеята му се състои това: наблюдател човек, да прецени от разговор, чрез размяна на текст, дали срещу себе си има друг човек или машина, без да ги вижда. Компютърната програма ще издържи „теста на Тюринг“, ако разпитваният не е способен да я различи от човек. Тюринг е считал, че ако поне 30% от тестваните не разпознаят, че това е програма, тя минава теста.

Днес вече сме убедени във възможностите на ИИ, но както всяка наука, и тази преминава през върхови моменти на развитие и моменти на спад. Има един дълъг период през 70-те до средата на 80-те години, когато в тази област са правени малко изследвания, в сравнение с останалите сфери на науката.

През 1957 г. Франк Розенблат прави фундаментално изследване и предлага "Перцептрон: възприемащ и разпознаващ автомат" (F., 1957) като клас изкуствени невронни мрежи, имитиращи аспекти на мозъка и рецепторите на биологичните системи. На база на това изследване през 1967 той изгражда компютър Mark 1 Perceptron базиран на невронна мрежа, която се „обучава“ на принципа проба - грешка.

През 1969 Марвин Мински и Сиймор Пейпърт публикуват книга, озаглавена „Перцептрони“, която е забележителната разработка върху невронните мрежи. Тази разработка за известно време, е главният аргумент срещу бъдещи проекти за изследване на невронни мрежи. (Minsky, 1969) Мински и Пейпърт атакуват Перцептрона на Розенбалт, като посочват, че неговата двуслойна структура, не може да обхване всички асоциации в една невронна мрежа.

Единствената област от ИИ, която се развива успешно през 70-те и началото на 80-те години са експертните системи. Те са предназначени да решават проблеми в някоя специфична област от човешкото познание. Днес те се наричат от някои учени „предишна ера на ИИ“, именно защото техния бум е по времето, когато беше замряло развитието на другите области на ИИ.

Експертните системи са изградени на основа на факти, извлечени от знанията на експерти от конкретна област, от които чрез логически правила се достига до ново знание. Типичен пример за такава система е MYCIN - експертна система за диагностициране на бактериални болести. MYCIN е разработена в началото на 70-те години на миналия век в Станфордския университет от Едуард Шортлиф (Shortliffe & B.G., 1975). Програмата е била разработвана в продължение на 5 години. Било необходимо да се натрупат и съхранят като факти, знанията на експертите в областта на бактериалните болести. Въз основа на фактите, чрез съчетаване на логически правила, се достига до заключение. Ето един прост пример от тази експертна система:

Ако Микроорганизмът е грам-отрицателен,

Морфологията на организма е пръчица и
Организмът е аеробен.

То Данните сочат, че организмът е бактериод.

Развитието на експертните системи днес не е спряло, макар че сега те са изградени съвсем на други принципи.

През 1987 година Дейвид Румелхарт и Джеймс Макклеланд публикуват фундаменталния си труд „Паралелно разпределена обработка“ ((Rumelhart & McClellan, 1987), с който дават отново силен тласък на развитието на ИИ. Защо все пак хората са по-умни от компютрите, въпреки че отделни компютърни програми са способни да надминат човешкото изпълнение на дадена задача? Отговорът, според тези учени, се крие в мощната паралелна архитектура на човешкия ум. Те описват тази архитектура, чрез теория на познанието, наречена конекционизъм. Теорията предполага, че умът е съставен от голям брой елементарни единици, свързани в невронна мрежа. Психичните процеси са взаимодействия между тези единици, които се подбуждат и възпират една друга в паралелни, а не в последователни операции. В този контекст знанието вече не може да се разглежда като съхранявано в локализирани структури; вместо това се състои от връзки между двойки единици, които са разпределени в цялата мрежа. На тази теория основно се дължи напредъка в последно време на мрежите за дълбоко обучение (DL Deep Learning), и на големите езикови модели (Large Language Models LLM).

През 1995 г. Ръсел и Норвиг публикуват учебник (Russell and Norvig, 1995), наречен „Искусствен интелект: модерен подход“. Тази книга става настолна за много поколения студенти, а през 2020 година, претърпява 4-то издание. Според наблюденията на университета в Бъркли (1547 Schools Worldwide That Have Adopted ИИМА (berkeley.edu) от този учебник учат повече от 1500 университета в света. В България, според този сайт, от учебника се ползват общо 5 университета.

Джон Сърл, професор във Философския факултет в Бъркли и дългогодишен изследовател на ИИ, създава понятието „силен изкуствени интелект (СИИ)“ (Strong Artificial Intelligence SAI). Машина със силен изкуствен интелект ще има психически състояния като тези на хората, например, може да разбере разказани истории както човек ги разбира. Засега силният ИИ съществува само в теоретичен аспект - компютър има интелект, равен или надминаващ човешкия. Такъв ИИ ще има самосъзнание; ще може да предвижда; да се самообучава и да взема самостоятелно решения. Професор Сърл, създава мисловен експеримент, наречен „Китайска стая“, за да демонстрира невъзможността, според него, за конструиране на СИИ.

Същността на мисловния експеримент се състои в следната постановка. В една стая стои човек, който не знае китайски език и през една дупка в стената, комуникира писмено с китаец на китайски. Китаецът праща през дупката въпросите на човека, а той чете предварително написани инструкции на английски, в които са посочени отговорите на всички възможни въпроси на китаеца. Когато получи йероглифите, търси в инструкцията измежду всички възможни съчетания, кое да избере за отговор. Напомняме, че експериментът е мисловен, теоретичен. Един външен наблюдател ще заключи, че човекът в стаята знае китайски. Според така проведен експеримент, китаецът може никога да не разбере, че събеседникът му не знае китайски. А пък човекът вътре в стаята може никога да не проумее за какво е бил разговорът.

Ако в стаята вместо човек, имаше компютър, в който предварително са въведени инструкции за отговори и реакции на всички възможни въпроси и ситуации, то този компютър ще изглежда като човек, но в никакъв случай не може да се отъждестви с човек.

Експериментът „Китайска стая“ ще илюстрираме като използваме изкуствен интелект (Фиг. 1). Компанията Майкрософт работи в посоката за създаване и обучение на приложение на ИИ, което генерира изображения, въз основа на текст Image Creator from Microsoft Bing. Текстът, който задаваме на ИИ да илюстрира е: „John Searle's science experiment called the chinese room with Chinese instructions and an outside observer“ Тази илюстрация е най-близо до представата за мисловния експеримент на проф. Сърл.



Нарисувал: Image Creator from Microsoft Bing <https://www.bing.com/images/create/john-searle27s-science-experiment-called-the-chines/64bb8536c57746d1b544a94166da8440?id=3%2b6bnLNh1ksOiVU1MqnfQw%3d%3d&view=detailv2&idpp=genimg&FORM=GCRIDP>

Фиг. 1. Експеримент „Китайска стая“.

Този експеримент, наречен от някои учени „най-значимия мисловен експеримент на 20 век“ не е оборен теоретически, има много противници, които смятат, че е възможно да се създаде свръхинтелект. Основните критики се отнасят до това, че експериментът е само мисловен, не може да бъде проведен експериментално. На което професор Сърл отговаря, че технически невъзможно не означава, че е логично невъзможно (Searle, 1980).

Генеративен изкуствен интелект (Generative Artificial Intelligence GAI)

Изкуственият интелект е научна област, която съчетава компютърни науки и набори от данни, за решаване на конкретни проблеми. ИИ се обучава - това е процес на избор за ниво верни резултати и постепенното достигане до това ниво. Един от най-новите клонове в дълбокото обучение е разработването на специфичен вид невронна мрежа, наречена Трансформър. Това е невронна мрежа, която е проектирана да генерира последователности от свързани елементи от данни (като изречение например). За първи път Трансформърът е бил предложен през 2017 г. в статията „Вниманието е всичко, от което се нуждаете“ (Vaswani, et al., 2017), резултат от усилията на редица учени. От 2017 г. до създаването на приложението ChatGPT времето е само 5-6 години.

Генеративен изкуствен интелект е наричан този изкуствен интелект, който е в състояние да генерира съдържание, което преди това не е съществувало в пространството – текст, изображения, книги, филми, дори части от научни трудове.

Стъпвайки на съществуващи технологии като Трансформър и големите лингвистични модели (LLM), моделите, използващи GAI са обучавани върху огромни обеми от данни (например цялата История на България и например всички речи и произведения на Стефан Стамболов) как да предскажат следваща дума в дадено изречение. Анализирайки хиляди съчетания от думи, могат да предположат, че след фразата „Не щеме злато, искаме...“, най-вероятното продължение е „свобода и човешки правдини“. (Следвайки известното стихотворение на Стамболов „Не щеме ний богатство,/ не щеме ний пари,/ а искаме свобода, човешки правдини!“ заедно с фразата „щеме“, която очевидно се свързва с някакъв специфичен изказ.). GAI може да създаде ново стихотворение в стила на Стамболов, което не съществува до този момент, само въз основа на подаден на вход първи ред, Примерите могат да бъдат върху различни данни, но главната идея на генеративния изкуствен интелект е, че може даден модел

да се „научи“ да генерира статистически вероятни резултати, когато крайният потребител му даде задача.

Генеративните модели се използват от години в статистиката за анализ на числени данни. Възходът на дълбокото обучение обаче направи възможно разширяването им извън числата; до текст, изображения, реч и други сложни типове данни. Бъдещето е в тези модели, които се обучават върху широк набор от данни, които могат да се използват за различни задачи, с минимална фина настройка. Системите, които изпълняват конкретни задачи в един конкретен домейн, отстъпват място на генеративен ИИ, който учи по-общо и работи в различни домейни и върху различни проблеми. Базовите модели, обучени на големи масиви от данни и после фино настроени за набор от приложения, движат тази промяна.

Най-популярните примери на такъв тип приложения с GAI днес са ChatGPT, Bing AI, Bard и др. Силата на тези приложения е в това, че в резултат на анализа на огромни обеми от текстови данни и изображения, могат да съставят смислен отговор на който и да е въпрос, да генерират съдържание, което преди това не е съществувало.

Същността на GAI моделите, както посочихме, но нека пак да повторим, защото е много важно е, че те са вероятностни модели, генерират статистически най-вероятното съчетание на изрази, което следва от поставената им задача, но не и непременно 100% вярното.

На българския телевизионен екран в момента може да се гледа една популярна игра „Попитаме 100 човека...?“, в която играещите се опитват да отгатнат какво са отговорили тези 100 човека на въпроса, а не кой е правилния отговор. Постановката на тази игра много прилича на постановката, при която човек задава въпрос на някое от приложенията с генеративен изкуствен интелект. GAI ще даде най-вероятния отговор, който е извлечен от анализ на коментарите, блоговете, уикитата, чатовете и пр., но не задължително 100% верния отговор. Много вероятно е също така, при липса на достатъчно данни по даден въпрос, генеративния изкуствен интелект да си „измисли“ изрази, факти, т.е. да генерира информация, която изглежда най-правдоподобна като отговор на даден въпрос. Това явление има име - „халюцинации на GAI“. Компаниите, които разработват генеративни модели, винаги много ясно декларираат отказ от отговорност, като предупреждават многократно ползващите моделите да не се доверяват изцяло на тях. Днес много хора не са в състояние да разберат, че срещу тях няма мислеща като човек машина, нито свръхинтелект. Засега.

Реални опасности от развитието на ИИ

Сред учените се води неразрешен и до днес диспут „Може ли да се създаде свръхинтелект, който да чувства като човек?“. Към този въпрос се връщат най-често учените от когнитивните науки, за които е важно дали може да се направи ИИ, който да наподобява човешкия.

Дали въобще някога ИИ ще може да изпитва истински емоции, без да ги симулира? Това е въпрос, който засяга науката за човешката психика и мозък. Засега не е ясно. Има много неща, които не знаем за възможностите на ИИ. Учените изследват развитието на ИИ и предупреждават, че в бъдеще може да се стигне до състояние на „сингулярност“ (Kurzweil, Ray, 2005), точка от която ИИ, може да се развива самостоятелно, без участието на човек и няма да може да се върне обратно.

Но дори и сегашното развитие на ИИ е такова, че доведе до спорове, дискусии относно вредите, които може да нанесе на хората и дали тези вреди са съпоставими с ползите от него.

По-горе в текста споменахме за „халюцинациите“ на генеративния изкуствен интелект – състояние, при което програмата генерира измислици и неверни изречения и твърдения. Това състояние не е умишлено, вече подчертахме, че засега AI само симулира емоции и съпричастност, но това води до заблуда много хора, което е вредно и особено опасно. Проблемът е, че софтуерът с изкуствен интелект може да заблуди всеки - от високо образовани специалисти до хора с лабилна психика, готови да посегнат на живота си.

Много популярен стана случаят от юли 2022 г., със софтуерния инженер на Гугъл, който твърдял, че чатботът, с който работи е личност и има съзнание. Сензацията толкова се разширила, че било необходимо да се намесят психолозите на компанията, да проведат и

разпространят диалог с чатбота, за да покажат, че това е просто програма, която имитира човек – чатботът казал че има семейство, деца и пр.

Защо е толкова трудно да се разпознаят „халюцинациите“ на AI?

Проблемът идва най-вече от ограниченото ни познание по даден проблем. Ако проблемът е илюстриран от AI, ние лесно ще видим грешките и несъответствията и ще се разберем, че дадена картинка е направена от AI. Увеличете многократно Фиг. 1, по-нагоре от този текст, и ще видите едно гротескно изображение на човек - несъразмерно око, недовършени ръце, изобщо не особено умел опит на генеративния изкуствен интелект да направи илюстрация на Джон Сърл и то такава, че да не повтаря никоя негова снимка. Все пак изображението трябва да е уникално. Но вече системите за генериране на изображения дотолкова се усъвършенстваха, че наистина могат да създадат изображение, което да е поне толкова достоверно, колкото би го рисувал художник.

Как да разпознаем дали даден текст съответства на истината, при положение, че нямаме достатъчно познания по дискутирания въпрос. Това е трудно. Например, ученици трябва да напишат есе за Че Гевара или Васил Левски... Имат някакви познания по въпроса, но не дотолкова, че да преценят дали е вярно всичко, което им е написал чатбота.

Ето един реален пример... Google разработва свой чатбот, наречен Bard, който е в експериментален етап. При зададен въпрос: „Как е загинал Васил Левски“, чатботът отговаря с няколко смислени изречения, че „Левски е един от най-великите български революционери и национални герои“. Заедно с това „халюцинира“ версията, че е „обесен на Витоша, като при обесването му е използван механизмът на бялата лястовица“... Как се е намесил разказът на Йовков в този въпрос, никой не може да отговори. Но пък написаното звучи толкова достоверно, че заблуждава дори хора, които знаят от училище българската история. Ако този текст, беше попаднал при чужденец, който нищо не знае за Левски, но все пак е чувал нещо по първата част от текста, че това е наш национален герой, ще повярва и на останалата част. Текстът можеше да попадне в ученическо есе на китайски, немски, японски и пр. и тази небивалица бързо да се разпространи... В даден момент можеше да съществува в толкова версии в интернет, че да бъде приета за много вероятна от следващия чатбот, който анализира данните по този въпрос.

Заклучение

Само за около 70-80 години, ИИ претърпя огромно развитие. Времето от научните открития, до тяхното прилагане в практиката става все по-кратко.

Проблемите свързани с потенциалния риск от създаването на изкуствен суперинтелект най-често се negliжират от разработчиците на ИИ. Причината е, че създаването на системи с общ изкуствен интелект (Artificial General Intelligence), е централна стратегическа цел на водещите компании в областта на ИИ, като Microsoft, OpenAI, Meta, Google, Amazon, IBM. Да се приеме сериозно риска от създаването на изкуствен суперинтелект би означавало да се забави в определена степен създаването на системи с общ ИИ чрез въвеждане на допълнителни правила, контролни функции и други ограничения.

ИИ трябва да служи на хората, за подобряване на техните качества. Като преподаватели във висши училища, сме длъжни да запознаваме студентите с опасностите от безконтролното развитие на изкуствения интелект. Чрез съвместни разработки между преподаватели и студенти (Hristov Georgi, Markova M. (2022); Kovacheva M. (2020) трябва да насочваме тяхната креативност към полезните приложения на ИИ, към използването има главно за облекчаване и повишаване качеството на човешкия труд.

Хората от своя страна трябва да знаят, че срещу тях засега стоят обучени машини. Може би в близко бъдеще това ще се промени, но във всички случаи, човекът трябва да запази контрол върху машините.

Последната дума относно жизнено важни решения, свързани с ИИ, трябва да се взема от хората.

References

1. F., R. (1957). The Perceptron: A Perceiving and Recognizing Automaton.
2. McCarthy, J. (2004). WHAT IS ARTIFICIAL INTELLIGENCE? (Stanford University). Stanford: Stanford University.
3. Minsky, M., & Papert, S. (1969). Perceptrons: An Introduction to Computational Geometry. Cambridge: MIT Press.
4. Rumelhart, D. E., & McClelland, J. L. (1987). Parallel Distributed Processing. Explorations in the Microstructure of Cognition: Foundations. MIT Press.
5. Russell, S. J., & Norvig, P. (1995). Artificial Intelligence: A Modern Approach. Prentice Hall.
6. Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences([https://web-archive.southampton.ac.uk/cogprints.org/7150/1/10.1.1.83.5248.pdf](https://web.archive.southampton.ac.uk/cogprints.org/7150/1/10.1.1.83.5248.pdf)).
7. Shortliffe, E., & B.G., B. (1975). A model of inexact reasoning in medicine. Mathematical Biosciences. 23 (3–4), pp. 351–379.
8. Turing, A. M. (1950). Computing Machinery and Intelligence.
9. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Polosukhin, I. (2017). Attention Is All You Need. Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
10. Wiki. (1966). Wikipedia. Извлечено от ELIZA: <https://en.wikipedia.org/wiki/ELIZA>
11. Стърн, У. (1912). Die psychologischen Methoden der Intelligenzprüfung: und deren Anwendung an Schulkindern [Психологическите методи за тестване на интелигентността]. От W. Stern, Монографии на образователната психология, бр. 13. Лайпциг.
12. IBM. What is artificial intelligence? <https://www.ibm.com/topics/artificial-intelligence>
13. Stanford University. Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/entries/chinese-room/#toc>
14. Kurzweil, Ray (2005) The singularity is near: When humans transcend biology.
15. Kovacheva M. (2020) Storing Big Data in NoSQL databases compared to SQL -advantages and problems. In: 10th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education – 2020, Sofia.
16. Hristov Georgi, Markova M. (2022) Reporting of R&D Costs Under IAS 38 – The Case of Biopharmaceutical Companies. 4/2022, Научни трудове на УНСС.

ARTIFICIAL INTELLIGENCE AND TAX ADMINISTRATION IN BULGARIA

Georgi Emilov Hristov

Ph.D. candidate, Department of Information Technologies and Communications, University of National and World Economy
e-mail: georgi.hristov@unwe.bg

Abstract

This report explores the transformative potential of Artificial Intelligence (AI) technologies in tax administration in Bulgaria, with a focus on bridging the gap between existing tax fraud detection tools and the incorporation of business characteristics, such as abnormal financial status. It underscores the importance of integrating AI technologies to harmonize both structural and financial characteristics in tax fraud detection. The report presents a critical overview of existing tax fraud detection tools, deriving key insights on how such technologies can be improved.

Key words: integration of AI, tax administration, tax fraud

Introduction

Artificial Intelligence (AI) has emerged as a transformative force across various sectors, revolutionizing the way businesses and governments operate. In the context of tax administration, AI has the potential to streamline processes, enhance efficiency, and improve compliance, thereby contributing to the economic growth and stability of nations. In recent years, the Bulgarian tax administration bodies, predominantly the National Revenue Agency (NRA), have recognized the significance of digitalization, but the implementation of AI technologies remains a complex process which is still due.

The integration of AI into tax administration not only promises to increase revenue collection but also to simplify complex taxation processes, minimize errors, and foster a climate of transparency and trust between the government and its citizens.

This report aims to provide an overview of how different AI technologies are implemented in different tax administration bodies around the world. It explores the difficulties of such implementation and proposes a tool for enhancing tax fraud detection.

Research gap

This report serves as a critical bridge over an existing gap in the current landscape of tax administration and fraud detection. Presently, the tools and methodologies employed for detecting tax fraud predominantly focus on structural characteristics, such as the organization's structure or transactional patterns. While these aspects are undoubtedly essential for identifying potential irregularities, they often fall short in capturing a more holistic view of tax fraud. This limitation becomes glaringly evident when one considers the inherent dynamism and adaptability of fraudulent activities in the modern economic landscape.

The gap that this report addresses centers around the need to incorporate the business characteristics of tax fraud, specifically factors like abnormal financial status, into the detection process. Traditional methods may not sufficiently recognize signs of tax evasion or fraud that are *rooted in the financial* health or behavior of an entity. Such behaviors may include abnormal fluctuations in revenue, expense patterns, or financial ratios that deviate from industry norms and benchmarks.

By delving into this critical gap, this report seeks to shed light on how Artificial Intelligence can be harnessed to bridge the divide between structural and business characteristics in tax fraud detection. AI technologies have the capacity to analyze vast and complex financial datasets, identifying anomalies

and patterns that human auditors might overlook. By taking into account abnormal financial status alongside structural information, AI-powered solutions have the potential to revolutionize tax administration, providing a more comprehensive and accurate assessment of potential fraud.

Literature review

The use of AI in tax administration can have significant benefits. It can improve taxpayer compliance by incorporating information and communication technology (ICT) into the tax administration system (Djafri et al., 2023). The application of e-tax systems, AI, chatbots, and biometric identification can streamline processes such as registration, filing, payment, and identification, making it easier for taxpayers to comply with their obligations (Djafri et al., 2023). Additionally, AI can be used to monitor entrepreneurial activity and identify cases of tax evasion, both by regulatory authorities and by entrepreneurs themselves.

Furthermore, AI can contribute to macroeconomic stability and economic growth. Tax administration reforms, including the integration of AI, can address challenges such as low tax revenue and corruption in the tax bureaucracy (Herbert et al., 2018). By improving tax policy and administration, AI can help generate higher tax revenue and reduce opportunities for corruption, leading to economic stability and growth.

Artificial intelligence (AI) has the potential to revolutionize tax administration by improving efficiency, enhancing tax collection and management, and reducing tax evasion. Several studies have explored the application of AI in tax administration and its impact on various aspects of the tax system.

Zhang (2023) suggests that the construction of an AI-based environmental protection tax system can improve tax collection and management, tax payment service, and tax management (Zhang, 2023). By leveraging AI technology, tax authorities can optimize the design of the tax system and enhance the efficiency of tax administration processes. This can lead to more effective tax collection and a reduction in tax evasion.

Another study highlights the digitalization of tax administration communication as a result of AI technology (Ihnatišínová, 2021). AI creates new digital communication channels, enabling more efficient and paperless tax administration. This can streamline communication between taxpayers and tax authorities, leading to faster and more accurate processing of tax-related information.

Furthermore, the integration of AI technology and tax risk management can establish an intelligent tax system (Huang & Zhang, 2022). By using AI algorithms such as Bagging and Support Vector Machines (SVM), tax authorities can analyze large volumes of data to identify potential tax risks and anomalies. This can help in detecting tax evasion and improving tax compliance.

Jiang (2022) presents an intelligent tax planning platform based on the current situation analysis, and proposes a new tax planning algorithm based on the advanced technology of big data and artificial intelligence, which helps enterprises to carry out intelligent tax planning under the premise of legal compliance, reduce company labor costs, and increase the freelancers' revenue.

However, there are certain challenges that need to be addressed when implementing such technologies. One such challenge is the need for administrative capacity and accountability. Studies have shown that countries like Bulgaria and Romania face obstacles in terms of administrative capacity and accountability, which can hinder the successful implementation of reforms, including those related to AI (Wegener et al., 2011). Therefore, it is crucial for Bulgaria to address these challenges and build a strong administrative foundation to support the integration of AI in tax administration.

To fully harness the potential of AI in tax administration, Bulgaria should consider the experiences and best practices of other countries. In the following section, we provide several examples of successful implementation of AI technologies in tax administration processes.

Examples of implementation of AI in tax administration

The government of Malta has implemented a system which alerts the authorities when a tax payer's income does not correspond to the value of assets in his possession (Zammit, 2023). While a good example of combating tax fraud and money laundering, this case poses the question of personal rights and security.

The Canadian government has announced TaxGPT (<https://taxgpt.ca/>), a chatbot which assists taxpayers in the process of understanding and filing their income returns. Of course, relying solely on a chatbot to do your taxes is never a good option, but it is still a great support tool.

The Netherlands also implemented a self-learning algorithm which categorizes citizens into risk groups in effort to stop childcare benefits fraud. It is also a good example of how terribly bad such implementation can go if not done right. Thousands of Dutch families have been wrongfully categorized and punished over the years which led to a need for revision of the tool. (Goujard and Manancourt, 2022). Such system has been successfully implemented in Italy.

Australia claims to have identified over \$530 million in unpaid tax bills and prevent \$2.5 billion in fraudulent claims using AI models, including deep learning and natural language models.

China has implemented reforms in tax collection and administration, using AI and blockchain technology to reshape its core processes (Wang & Chen, 2018). Similarly, Russia has utilized AI in the tax sphere to identify cases of tax evasion and ensure compliance.

By studying and adapting the successful approaches, Bulgaria can develop a comprehensive strategy for integrating AI into its tax administration system.

Challenges in implementing AI in the Bulgarian tax administration

There are several non-neglectable challenges in implementing AI technologies in the tax administration process in Bulgaria.

Wegener et al., 2011 presents the need for administrative capacity and accountability so that reforms can be pushed through successfully. As he goes to show, Bulgaria and Romania did not have that capacity then. The country's political and socioeconomic transformations are still unfinished, which hinders the adequate completion of the Europeanization process (Wegener et al., 2011). These challenges can impact the implementation of AI in tax administration, as they require a strong administrative capacity and a robust accountability framework.

Another issue is lack of infrastructure and technology. Upgrading and modernizing existing IT infrastructure to support AI implementation can be costly and time-consuming. This includes procuring powerful hardware, software, and networking resources. Moreover, integrating the complex AI systems with the already in-place legacy systems require skillful professionals in the field. This leads to the next issues – skill gap.

Building a team with the necessary AI expertise can be challenging. Recruiting and training data scientists, machine learning engineers, and AI specialists may require significant effort. In addition to that, they also must be skilled in tax, audit, and accounting which makes the task even harder.

As we saw in the aforementioned examples, ethical considerations are always an issue which has to be taken care of. Tax administration AI systems must be designed with ethical considerations in mind, avoiding bias in algorithms and ensuring fairness in decision-making are crucial aspects for people to trust the technology. Building and maintaining public trust in AI-powered tax administration is critical. Ensuring that the public understands how AI is being used and its benefits is important for acceptance.

To successfully implement AI in the Bulgarian tax administration, a comprehensive strategy that addresses these challenges is essential. This strategy should involve collaboration among stakeholders, including government agencies, IT experts, data scientists, and the public, to ensure that AI systems are effective, transparent, and trusted.

Conclusion

The implementation of AI in tax administration in Bulgaria holds immense potential for transforming the efficiency, effectiveness, and transparency of the country's tax system. This report has explored the various facets of integrating AI into tax administration, highlighting both the opportunities and challenges associated with this endeavor.

Bulgaria, like many other countries, faces significant challenges in modernizing its tax administration system. The complexities of tax laws, the need for efficient revenue collection, and the demand for improved taxpayer services all converge to create a pressing need for innovative solutions. AI technologies, including machine learning, data analytics, and automation, offer a promising path forward.

As Bulgaria progresses in its journey toward AI-driven tax administration, it is imperative that the government and tax authorities collaborate with technology experts, data scientists, and other stakeholders. This collaboration will help in building a strong foundation for AI implementation, ensuring that it aligns with the unique needs and requirements of Bulgaria's tax system.

In conclusion, embracing AI in tax administration is not only a technological improvement but also a strategic necessity for Bulgaria. By harnessing the power of AI, Bulgaria can create a more efficient, fair, and taxpayer-friendly tax system that not only bolsters revenue collection but also enhances the overall economic and business environment in the country.

References

1. Ihnatišínová, D. (2021). Digitalization of tax administration communication under the effect of global megatrends of the digital age. *SHS Web of Conferences*, 92, 02022. <https://doi.org/10.1051/shsconf/20219202022>
2. Joshi, A., Prichard, W., & Heady, C. (2014). Taxing the informal economy: the current state of knowledge and agendas for future research. *The Journal of Development Studies*, 50(10), 1325-1347. <https://doi.org/10.1080/00220388.2014.940910>
3. Kochanova, A., Hasnain, Z., & Larson, B. (2018). Does e-government improve government capacity? evidence from tax compliance costs, tax revenue, and public procurement competitiveness. *The World Bank Economic Review*, 34(1), 101-120. <https://doi.org/10.1093/wber/lhx024>
4. Wegener, S., Labar, K., Petrick, M., Marquardt, D., Theesfeld, I., & Buchenrieder, G. (2011). Administering the common agricultural policy in bulgaria and romania: obstacles to accountability and administrative capacity. *International Review of Administrative Sciences*, 77(3), 583-608. <https://doi.org/10.1177/0020852311407362>
5. Huang, W. and Zhang, J. (2022). Artificial intelligence technology and tax risk management innovation.. <https://doi.org/10.1117/12.2641091>
6. Jiang, C. (2022). Research and implementation of tax planning platform based on big data and artificial intelligence.. <https://doi.org/10.1117/12.2658261>
7. Lu, Y. (2022). The influence of public mental health based on artificial intelligence technology on the teaching effect of business administration major. *Journal of Environmental and Public Health*, 2022, 1-11. <https://doi.org/10.1155/2022/5353889>
8. Zhang, J. (2023). Optimization of the environmental protection tax system design based on artificial intelligence. *Frontiers in Environmental Science*, 10. <https://doi.org/10.3389/fenvs.2022.1076158>
9. Zammit, C. (2023). New AI-powered software to be deployed to catch tax cheats. *Times of Malta*, <https://timesofmalta.com/articles/view/finance-ministry-deploy-ai-powered-software-catch-tax-cheats.1030430>
10. Clothilde, G., Manancourt, V., (2022). Dutch scandal serves as a warning for Europe over risks of using algorithms, <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/>

НАМАЛЯВАНЕ НА РЕКЛАМНИТЕ ИЗМАМИ В ДИГИТАЛНИЯ МАРКЕТИНГ С ПОМОЩТА НА ИЗКУСТВЕН ИНТЕЛЕКТ

Ивона Велкова

Асистент, доктор, катедра „Информационни технологии и комуникации“, УНСС
e-mail: ivonavelkova@unwe.bg

Резюме

Непрекъснатата дигитализация на глобалната икономика поставя началото на свързаност и възможности за бизнеса да се ангажира с потребителите. Въпреки това, тя също така довежда и до различни заплахи в онлайн пространството. Една от тези заплахи е рекламната измама, която застрашава целостта и ефективността на онлайн рекламните кампании в все по-дигитализираната икономика. Изкуственият интелект (ИИ), оборудван с възможности за анализ на данни в реално време и усъвършенствани алгоритми за идентифициране на модели, предлага възможности за ефективно реагиране срещу този вид атаки. Чрез анализа на множество от данни, ИИ може да разграничи истинските потребителски взаимодействия в дигиталното пространство от измамните, да маркира аномалии и да се адаптира бързо към развиващите се измамни тактики. Този документ се представя под формата на решение за смекчаване на рекламните измами, със специфичен акцент върху тяхното въздействие и ролята на ИИ в справянето с това предизвикателство и насърчава създаването на по-сигурна и устойчива дигитална икономика.

Ключови думи: изкуствен интелект, заплахи, дигитален маркетинг

REDUCING ADVERTISING FRAUD IN DIGITAL MARKETING WITH ARTIFICIAL INTELLIGENCE

Ivona Velkova

Abstract

The continuous digitization of the global economy brings about connectivity and opportunities for businesses to engage with consumers. However, it also introduces various threats in the online space. One of these threats is ad fraud, which jeopardizes the integrity and effectiveness of online advertising campaigns in an increasingly digitized economy. Artificial intelligence (AI), equipped with real-time data analysis capabilities and advanced algorithms for pattern recognition, offers effective ways to counteract such attacks. Through the analysis of extensive datasets, AI can distinguish genuine user interactions in the digital realm from fraudulent ones, identify anomalies, and adapt rapidly to evolving fraudulent tactics. This document presents an approach to mitigating ad fraud, with a specific focus on its impact and the role of AI in addressing this challenge, promoting the creation of a more secure and resilient digital economy.

Keywords: artificial intelligence, threats, digital marketing

JEL: M31, O33

Въведение

Дигитализацията на глобалната икономика представлява промяна в начина, по който се провеждат икономическите дейности, обменът на данни и достъпът до пазарите. През последните десетилетия, бизнес средата свидетелства за изключителна трансформация в начина, по който компаниите рекламират своите продукти и услуги. Тази промяна се движи от

напредъка в технологиите като Интернет, изкуствен интелект (ИИ), разпространението на мобилни устройства и растежа на платформите за електронна търговия. Процесът на дигитализация довежда до повишена свързаност и взаимозависимост между различните индустрии и потребителите в глобален мащаб [1], [2].

Интегрирането на технологиите не само фундаментално трансформира сектори като икономиката, здравеопазването и образованието, но също така очертава ключови роли за редица направления, включително и дигиталния маркетинг. Последният се използва в икономиката като незаменим компонент на съвременните маркетингови стратегии, използвани от компаниите по целия свят. За да може бизнесът да достигне до клиентите той използва дигиталната реклама, която е ключов компонент на дигиталния маркетинг. Тя предлага на бизнеса възможността да се свързва с целевата си аудитория в реално време, предоставяйки персонализирано съдържание и решения. Дигиталният маркетинг включва широк спектър от онлайн дейности, насочени към достигане и ангажиране с целеви аудитории чрез дигитални канали като уебсайтове, социални мрежи, електронна поща и различни форми на онлайн съдържание. В резултат на това онлайн рекламата се превръща в неразделна част от маркетинговите кампании и ключов двигател на приходите в дигиталната икономика [3], [4].

Въпреки че използването на дигиталната реклама донася множество предимства за бизнеса, то внася и нови предизвикателства като част от тях са трафик от ботове (интернет-трафик, който е генериран от автоматизирани програми), измами с данни, рекламни измами и др. Последните обхващат различни практики, целящи манипулиране на онлайн рекламни кампании, като например фалшиви кликове, измами при показвания и стекинг на реклами [5]. Рекламните измами имат вредни ефекти както върху бизнеса, така и върху потребителите. За бизнеса, те довеждат до финансови загуби, тъй като плащат за измамни взаимодействия, които не предоставят истинска стойност. Освен това, този вид измами развалят репутацията на марката, тъй като потребителите могат да я асоциират с измамни дейности. Проучване на Juniper на установява, че 22% представляват значителна част от общите разходи за онлайн реклама на компаниите, които са отрицателно повлияни от рекламни измами, което допълнително подчертава значителната финансова тежест, която налага на бизнеса [6]. За потребителите, рекламните измами могат да доведат до лошо онлайн преживяване, излагане на потенциално вредно съдържание и намаляване на доверието в дигиталната реклама.

За да се предотвратят тези предизвикателства, използването на напреднали технологии като изкуствен интелект (ИИ) се явява ключова стратегия. ИИ-базираните решения предоставят възможности за по-ефективно и активно откриване и предотвратяване на измамните дейности, тъй като имат способността да анализират огромни количества данни в реално време. Това означава, че те могат бързо да идентифицират и маркират подозрителни дейности, когато се появят, което позволява незабавна намеса, както и откриване на сложни модели и аномалии в данните [5].

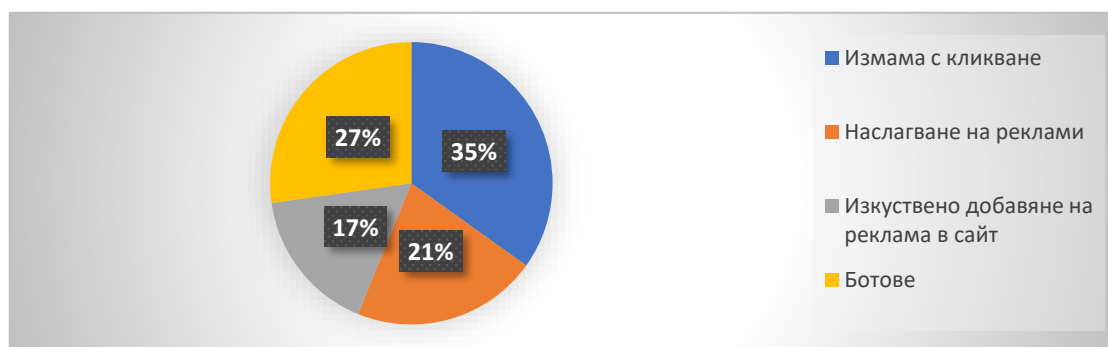
Същност и видове рекламна измама

Разпространението на дигиталния маркетинг поставя началото на нова ера на рекламата, позволявайки на бизнеса да достигне до глобалната аудитория. Дигиталната реклама е маркетингова стратегия, която използва дигитални канали като социални медии, мобилни приложения и платформи за популяризиране на продукти, услуги или съдържание. Много от онлайн рекламите позволяват на фирмите да се насочват към конкретни демографски данни, интереси и ключови думи, като гарантират, че рекламите им достигат до аудитория, интересуваша се от техните продукти или услуги. Достигайки до правилните хора, вероятността кликванията да се превърнат в продажби или потенциални клиенти, се увеличава. Освен с всички предимства при използването на дигиталните реклами, съществуват и измами в този аспект [5].

Сферата на рекламните измами обхваща спектър от хитрости, като някои от най-разпространените методи включват внедряването на „clickbots“ – автоматизирани програми, щателно проектирани да взаимодействат с реклами, като по този начин създават разходи за рекламодателите, без да кулминира в желаното преобразуване на продажбите [7]. Друга често срещана рекламна измама се нарича „наслагване на реклами“, практика, при която отделно

рекламно пространство се продава многократно на отделен уебсайт, което води до наслагване на реклами [8]. Тази измама може да създаде илюзията за увеличен брой импресии (основен показател, който измерва колко често реклама се показва пред потенциални зрители) на рекламодателите, но въпреки това ограничава визуалното излагане на потребителите на уебсайтовете само до най-горната реклама в стека. Тези импресии включват автоматизирани ботове и злонамерени участници, които генерират фалшиви изгледи и кликания. Тези измамни дейности могат да подкопаят ефективността и целостта на дигиталните маркетингови кампании [9].

Според проучване има различни видове рекламни измами, категоризирани спрямо техния обхват, разпространение, причинени щети и т.н.. В диаграмата по-долу са представени някои от най-популярните видове рекламни измами, спрямо тяхното разпространение и значение в сферата на измамите с дигитална реклама по данни от 2022г. [10].



Диаграма 1. Видове рекламни измами.

Видове рекламни измами

- Измама с кликване** - включва повтарящи се и често автоматизирани кликания върху онлайн реклами с намерението за изкуствено увеличаване на честотата на кликване. Както се вижда на Диаграма 1, тя е сред най-популярните измами. Тази практика не включва истински потребителски интерес, а се изпълнява, за да заблуди рекламодателите да повярват, че рекламните им кампании са по-успешни, отколкото са в действителност. Този вид измама е свързана с рекламата с плащане на клик (pay-per-click - PPC), която включва поставяне на реклами в страниците с резултати от търсачките (напр. Google Ads), където рекламодателите наддават за конкретни ключови думи и техните реклами се показват, когато потребителите търсят по тези думи. PPC е модел на онлайн реклама, при който рекламодателят плаща на платформата, където иска да се визуализира рекламата му, въз основа на броя кликове, които получава. Когато потребителите взаимодействат с реклама, като кликнат върху нея, рекламодателите се таксуват с такса. Чрез изкуствено увеличаване на разходите, намаляване на ефективността на кампанията и причиняване на неудовлетвореност за целевия бизнес, хората, реализиращи измамата, се стремят да спечелят конкурентно предимство. Това може да бъде особено пагубно по време на пиковите часове като празничните сезони, когато конкуренцията за рекламно пространство е жестока [4], [11].
- Измама с наслагване на реклами** - практика в дигиталната рекламна индустрия, при която множество дисплейни реклами са умишлено наслоени или подредени една върху друга в рамките на едно и също рекламно разположение на уеб страница или мобилно приложение. Това може да доведе до показване на няколко реклами едновременно в едно рекламно пространство, често без знанието или съгласието на потребителя. Това се прави, за да се увеличи изкуствено броят на рекламните импресии или кликания, като по този начин се генерират приходи за издателя (предоставящият платформата за визуализация на рекламата) и разходи за рекламодателя, дори когато потребителят може да не вижда или да не се ангажира с рекламите. Те често използват автоматизирани скриптове или ботове, за да

създават фалшиви рекламни взаимодействия, като генерират приходи, без да предоставят истинска стойност на рекламодателите [8].

- **Изкуствено добавяне на реклама в сайт** – при този вид измама рекламите се добавят в уебсайтове или уеб съдържание без съгласието на собственика на сайта или легитимната рекламна мрежа. Тези неоторизирани реклами често са с ниско качество и могат да отнемат приходи от законните издатели. Това изкуствено на реклами може да наруши потребителското изживяване чрез инжектиране на натрапчиви или неподходящи реклами, което води до отрицателно въздействие върху качеството на съдържанието [12], [13].

Осигуряването на положително потребителско изживяване е от съществено значение за запазване на доверието на потребителите. Натрапчивите реклами, бавно зареждащите се страници и неуместното насочване могат да отблъснат потребителите и да ги накарат да използват рекламни блокери, което може да намали приходите на уебсайтовете и приложенията.

Използване на технологии с ИИ за предотвратяване на рекламни измами

В днешно време дигиталната рекламна индустрия обработва милиарди точки от данни ежедневно, което прави почти невъзможно ръчното откриване на измамни дейности поради големия обем. Изкуственият интелект (ИИ) притежава капацитета бързо да анализира огромни набори от данни, използвайки усъвършенствани алгоритми за машинно обучение, за да разпознае измамни модели и тенденции, които могат да убегнат от традиционните инструменти [5]. ИИ изпълнява идеята за разработване на машини или софтуер, които могат да имитират човешкия интелект и действия. Той включва системи, базирани на правила, които се придържат към предварително определени инструкции, както и системи, управлявани от данни, които подобряват своята производителност чрез учене от модели [14], [15].

Възможностите на ИИ са наблюдение в реално време, поведенчески анализ, откриване на ботове, разпознаване на шаблони, анализ на съдържанието, прогнозиране на измами, автоматизирано докладване и блокиране на измамен трафик. Освен това алгоритмите на ИИ могат да бъдат обучени да откриват автоматизирани ботове и да ги различават от истинския човешки трафик, използвайки сложни модели за машинно обучение, за да идентифицират поведението на ботове, характеризиращо се с бързи, повтарящи се кликания, което се наблюдава в реално време и това помага за незабавно реагиране на аномални модели. Съществуващи решения са софтуери с използване на ИИ за откриване и блокиране на измамен трафик от ботове като White Ops, Integral Ad Science (IAS), MOAT на Oracle, Google Ads Click Fraud Protection, ClickCease и др., които откриват необичайно високи честоти на кликане, идващи от конкретен IP адрес, което показва потенциална измама с кликания. В резултат на това се намаляват и рекламните разходи за изкуствено създаден трафик и се подобрява ефективността на рекламните кампании [16].

ИИ имат възможност за предотвратяване на измами с импресии, като ИИ следи видимостта на рекламите в реално време и преценява дали импресиите са наистина видими от реални потребители. Той постига това чрез откриване на аномалии и нередности в разположението на рекламите. Съществуващи решения са DoubleVerify, Pixalate и др. ИИ може да идентифицира кога дадена реклама се показва в скрито или малко рекламno поле, което я прави практически невидима за потребителите. В такива случаи ИИ разпознава и докладва тези случаи като измама с импресии. Този проактивен подход помага на рекламодателите да избегнат плащането за импресии, на които липсва истинска видимост, като в крайна сметка подобрява техните рекламни кампании и предотвратява финансови загуби [17].

ИИ алгоритмите са умели в откриването на рекламната измама „наслагване на реклами“. Усъвършенстваните модели за машинно обучение на ИИ разглеждат отчетените рекламни импресии спрямо действителната видимост, разпознавайки несъответствията, които показват подредждането на рекламите. Така ако една уеб страница отчете 1000 рекламни импресии, но ИИ разпознае, че само една реклама е била наистина видима за потребителите, това разкрива измамата с наслагване на реклами. Това предоставя възможност на рекламодателите да

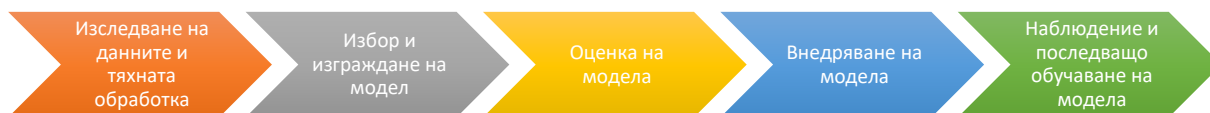
поддържат точността на кампаниите и да предотвратяват измамни практики. Съществуващо решение е системата Integral Ad Science (IAS) [18].

Рекламодателите все повече внедряват управлявани от ИИ модели, които проследяват всички действия на потребител. Така се гарантира, че правилните реклами получават заслуги за реализациите, елиминирайки възможността на измамниците да си приписват заслуги за неспечелените реализации.

Подход за анализ на рекламни данни, свързан с откриване на измамни дейности с ИИ

Анализът на рекламните данни е незаменим инструмент за получаване на знания, особено когато става въпрос за откриване и борба с измамни дейности. Използвайки различни технологии за извличане и анализ на данни, рекламодателите и фирмите могат да откриват ценни знания, скрити в огромни масиви от данни. Справянето с нарастващото предизвикателство на рекламните измами обаче изисква нещо повече от конвенционален анализ на данни.

Поради това се предлага подход, който съчетава предварителна обработка на данни, разработване на модел, непрекъснат мониторинг и експертни познания за откриване на измамни дейности в рекламни данни с помощта с използване на Python и алгоритми на ИИ. Този подход може да бъде приспособен към всеки бизнес и множества от данни. Предложеният подход се състои от стъпките, представени на Фигура 1.



Фигура 1. Стъпки за подход за анализ на данни с ИИ.

На първа стъпка от Фигура 1 се изследват данните. Те могат да бъдат от различни източници, включително кликания върху реклами, информация за устройството, IP адреси, низове на потребителски агенти и всяка подходяща контекстуална информация. Необходимо е да се познават данните, с които се работи – брой записи, типове данни, липсващи стойности и др. Данните, с които ще бъде извършено верифицирането на подхода, са от TalkingData AdTracking, пуснат на Kaggle през 2017 г. [19]. Множеството от данни съдържа над 100 000 записа за кликания върху мобилни реклами. Целта му е да предвиди дали потребителят ще инсталира приложение, след като щракне върху реклама и дали ще разграничи измамните кликания от легитимните. Данните са заредени и анализирани във формат DataFrame на Python. DataFrame е структура за данни, която се използва в програмния език Python, в библиотеката за анализ на данни и манипулация "pandas". DataFrame е двумерна таблица, която прилича на таблица с редове и колони. Тя позволява лесно съхранение и манипулация на данни в структуриран и организиран формат. След като данните бъдат подготвени с pandas, могат да бъдат използвани и други библиотеки като scikit-learn, TensorFlow, Keras, които са специализирани за внедряване на ИИ алгоритми.

```

import pandas as pd

data = pd.read_csv("train_sample.csv") #прочита файла
# показване на първите няколко реда от DataFrame, за да разбере структурата на данните
data.head()

#Резултат
ip app device os channel click_time attributed_time is_attributed
0 87540 12 1 13 497 2017-11-07 09:30:38 NaN 0
1 105560 25 1 17 259 2017-11-07 13:40:27 NaN 0
2 101424 12 1 19 212 2017-11-07 18:05:24 NaN 0
  
```

3	94584	13	1	13	477	2017-11-07 04:58:08	NaN	0
4	68413	12	1	1	178	2017-11-09 09:00:09	NaN	0

Наборът от данни е зареден и структурата на предоставената извадка е:

- ip: IP адресът на щракването, което е цифрово представяне, потенциално анонимно или хеширано от съображения за поверителност.
- app: Идентификационният номер на приложението за маркетинг, кодиран като цяло число.
- device: ID на типа устройство на мобилния телефон на потребителя, кодирано като цяло число.
- os: ИД на версията на операционната система на мобилния телефон на потребителя, кодиран като цяло число.
- channel: Идентификационният номер на канала на издателя на мобилна реклама, кодиран като цяло число.
- click_time: Времето на щракването в UTC.
- attributed_time: Времето, когато приложението е изтеглено, ако е изтеглено след щракване върху реклама. Това поле съдържа NaN (не число) за записи, при които не е имало изтегляне, което е обичайно за такива набори от данни, където „положителното“ събитие (изтегляне на приложение) е рядко в сравнение с броя кликания.
- is_attributed: Целевата променлива, указваща дали приложението е изтеглено (1) или не (0) след щракване върху рекламата.

От визуализацията се вижда, че attributed_time има липсващи стойности, което има смисъл, тъй като не всички кликания водят до изтегляния. Колоната is_attributed е двоична и ще бъде етикетът за прогнозно моделиране. Повечето от показаните тук записи не са приписани (стойността е 0), което предполага, че наборът от данни може да е дисбалансиран с много повече неизтегляния, отколкото изтегляния.

На следващо място следва да се провери общият размер на набора от данни:

```
dataset_size = data.shape
```

Оценяване на разпределението на целевата променлива is_attributed, за да се потвърди дали наборът от данни е небалансиран:

```
target_distribution = data['is_attributed'].value_counts(normalize=True)
```

```
#Резултат
((100 000, 8),
 0 0,99773
 1 0,00227
```

Наборът от данни съдържа 100 000 записа и 8 колони. Разпределението на целевата променлива is_attributed потвърждава, че наборът от данни е силно небалансиран:

- 0 (not attributed): 99,773% от записите
- 1 (attributed): 0,227% от записите

Това ниво на дисбаланс е типично при проблеми с откриването на измами, при които събитието, което има значение (в този случай изтегляне на приложение след щракване върху реклама) е рядко в сравнение с несъбитията. Такъв дисбаланс създава предизвикателства за прогнозното моделиране, тъй като повечето алгоритми за обучение предполагат равен брой примери за всеки клас или поне по-балансирано разпределение. За справяне с това могат да се използват няколко техники сред които са: повторно семплиране - чрез свръхсемплиране на малцинствения клас, недостатъчно семплиране на мнозинствения клас, или комбинация от двете; присвояване на по-висока тежест на малцинствения клас по време на обучението;

алгоритми за откриване на аномалии или ансамбъл методи, които са комбинирането на прогнозите на няколко модела.

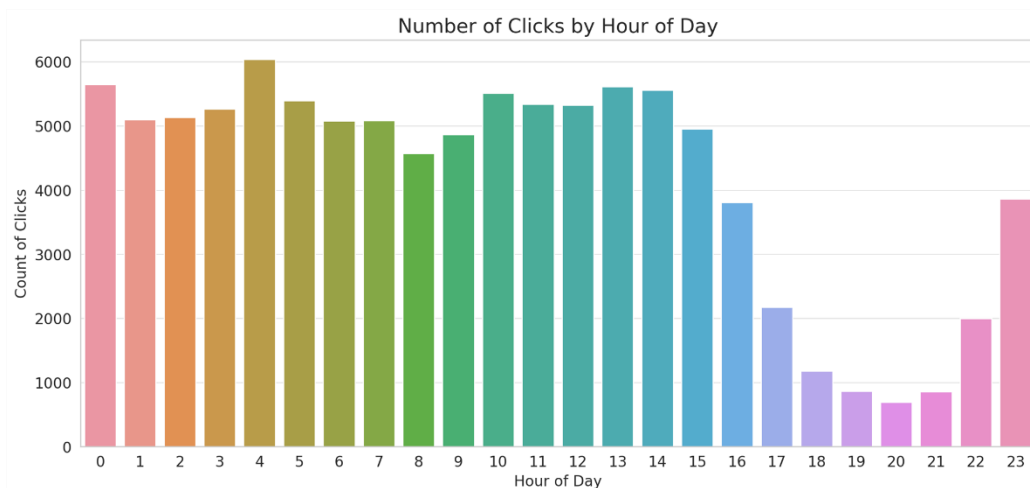
```
# проверка на типовете данни и наличието на липсващи стойности
data_info = pd.DataFrame(data.dtypes, columns=['Data Type'])
data_info['Non-Null Count'] = data.count()
data_info['Dtype'] = data.dtypes
data_info['Memory Usage'] = data.memory_usage(deep=True)
# обобщение на числовите стойности
numerical_summary = data.describe()
```

Възможно е след получаването на данните да са необходими допълнителни стъпки. Те се фокусират върху преобразуването на типа данни, анализирането на колоната „attributed_time“ и общата предварителна обработка на данни.

За извършване на анализ на данните, последните могат да бъдат визуализирани чрез графика. Кодът за нея в Python, е:

```
import matplotlib.pyplot as plt
import seaborn as sns
# задаване на стил на графиката
sns.set_style("whitegrid")
plt.figure(figsize=(14, 6))
sns.countplot(x='hour_of_day', data=data)
plt.title('Number of Clicks by Hour of Day')
plt.xlabel('Hour of Day')
plt.ylabel('Count of Clicks')
plt.show()
plt.figure(figsize=(14, 6))
# ограничава се хистограмата до времеви разлики от по-малко от 1 час (3600 секунди)
sns.histplot(data=data.loc[data['click_time_diff'] < 3600, 'click_time_diff'], bins=50, kde=False)
plt.title('Distribution of Click Time Differences (less than 1 hour)')
plt.xlabel('Time Difference (seconds)')
plt.ylabel('Frequency')
plt.show()
```

Резултатът се визуализира на Диаграма 2.



Диаграма 2. Графика, показваща броя на кликванията по часове.

Диаграма 2 показва броя на кликванията в различните часове на деня. От нея става ясно, че активността на кликванията варира, като в определени часове активността е по-висока. Това може да се дължи на модели на поведение на потребителите или може да показва периоди, в които измамната дейност е по-разпространена.

Тази хистограма показва честотата на разликите във времето на кликване в секунди. Има висока честота на кратки разлики във времето, която намалява бързо с увеличаването на разликата във времето. Този модел предполага, че има много случаи, при които щракванията от един и същ IP се появяват в бърза последователност, което може да е показателно за автоматизирано щракване или активност на бот.

Следващите стъпки при анализа на данните могат да бъдат:

- **Инженеринг на функции:** Създаване на нови функции, които биха могли да уловят по-добре моделите, свързани с измама, като функции, базирани на времето (напр. кликвания по време на деня) или функции за взаимодействие (напр. комбинации от IP и приложение).
- **Визуализация на данни:** График на данните за визуализиране на разпределения, връзки и потенциални аномалии.
- **Предварителна обработка:** Кодиране на категорични характеристики, ако е необходимо, функции за мащабиране, обработка на липсващи стойности и справяне с дисбаланса на класа.
- **Избор на модел:** Избор на подходящ модел или набор от модели за проблема с небалансираната класификация.
- **Обучение и оценка на модела:** Обучение на модела(ите) и оценяването им с помощта на подходящи показатели като площта под ROC кривата (AUC), кривата на прецизност-припомняне, матрица на объркване и др.
- **Настройка на модела:** Коригиране на хиперпараметрите на модела за подобряване на производителността.
- **Валидиране:** Използване на кръстосано валидиране, за да се гарантира, че моделът се обобщава добре към всички данни.

Предложеният подход може да бъде разширен и бъдещата работа по него е свързана с интегриране на машинно обучение в Python върху различни множества от данни за откриване на модели, зависимости и знания.

Заклучение

В постоянно се развиващия се свят на дигиталния маркетинг едно от предизвикателствата, пред което се изправят маркетолозите, е всеобщата проблематика на измамите в рекламата. Съществуват различни видове рекламни измами, някои от които генерират огромни разходи на рекламодателя, други носят изкуствено увеличение на нивото на

ефективност от дадена кампания, както и такива които подправят домейна на легитимни сайтове с измамни. При този вид измама рекламодателите вярват, че техните реклами се показват на законни сайтове и заплащат по-висока такса за разполагане на техните реклами, когато в действителност те се озовават в измамни такива. Докато бизнесите отделят значителни бюджети за рекламни кампании, заплахата от измамните дейности продължава да подкопава ефективността и възвръщаемостта на тези усилия. Конвенционалните методи често не успяват в срок със справянето с тези проблеми и водят до значителни финансови щети. Следователно прилагането на ИИ предлага убедително решение за борба с рекламните измами. Алгоритмите с ИИ демонстрират капацитет за бърз анализ на огромни набори от данни в реално време и целесъобразно откриване на аномални модели и аномалии. Това не само възстановява доверието в дигиталната реклама, но и издига цялостния интегритет на индустрията.

Литературни източници

- [1] O. Tomyuk and O. Avdeeva, "Digital transformation of the global media market: in search for new media formats," *Econ. Consult.*, vol. 37, Mar. 2022, doi: 10.46224/ecoc.2022.1.2.
- [2] "What is Artificial Intelligence and How Does AI Work? TechTarget," Enterprise AI. Accessed: Oct. 29, 2023. [Online]. Available: <https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence>
- [3] A. de Lucas Ancillo and S. Gavrilava Gavrilava, "The Impact of Research and Development on Entrepreneurship, Innovation, Digitization and Digital transformation," *J. Bus. Res.*, vol. 157, p. 113566, Mar. 2023, doi: 10.1016/j.jbusres.2022.113566.
- [4] N. Gohil and A. Meniya, *A Survey on Online Advertising and Click fraud detection*. 2020.
- [5] S. Mathur and S. Daniel, "It's Fraud! Application of Machine Learning Techniques for Detection of Fraudulent Digital Advertising," *Webology*, vol. 19, no. 1, pp. 2475–2490, Jan. 2022, doi: 10.14704/WEB/V19I1/WEB19166.
- [6] N. Agius, "\$84 billion of ad spend lost due to ad fraud in 2023," Search Engine Land. Accessed: Oct. 28, 2023. [Online]. Available: <https://searchengineland.com/ad-spend-lost-ad-fraud-2023-432610>
- [7] "What is click fraud? | How click bots work," Cloudflare. Accessed: Oct. 29, 2023. [Online]. Available: <https://www.cloudflare.com/learning/bots/what-is-click-fraud/>
- [8] "What is Ad Stacking? [+How to Prevent it]." Accessed: Oct. 29, 2023. [Online]. Available: <https://edgimesh.com/blog/ad-stacking>
- [9] "Infographic: Where Ad Fraud is Rife," Statista Daily Data. Accessed: Oct. 29, 2023. [Online]. Available: <https://www.statista.com/chart/11347/where-ad-fraud-is-rife>
- [10] "Ad Fraud Statistics (2023) - Business of Apps." Accessed: Oct. 29, 2023. [Online]. Available: <https://www.businessofapps.com/ads/ad-fraud/research/ad-fraud-statistics/>
- [11] M. Chen, V. S. Jacob, S. Radhakrishnan, and Y. U. Ryu, "Can Payment-per-Click Induce Improvements in Click Fraud Identification Technologies?," *Inf. Syst. Res.*, vol. 26, no. 4, pp. 754–772, Dec. 2015, doi: 10.1287/isre.2015.0598.
- [12] S. Mittal, R. Gupta, M. Mohania, S. K. Gupta, M. Iwaihara, and T. Dillon, "Detecting Frauds in Online Advertising Systems," in *E-Commerce and Web Technologies*, K. Bauknecht, B. Pröll, and H. Werthner, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006, pp. 222–231. doi: 10.1007/11823865_23.
- [13] S. Arshad, A. Kharraz, and W. Robertson, "Identifying Extension-Based Ad Injection via Fine-Grained Web Content Provenance," in *Research in Attacks, Intrusions, and Defenses*, F. Monrose, M. Dacier, G. Blanc, and J. Garcia-Alfaro, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 415–436. doi: 10.1007/978-3-319-45719-2_19.
- [14] N. Makkineni, A. Ciripuram, S. N, S. Subhani, and V. Kakulapati, "Fraud Detection of AD Clicks Using Machine Learning Techniques." Rochester, NY, Jun. 20, 2023. doi: 10.2139/ssrn.4486834.
- [15] I. Buch and M. Thakkar, "AI in Advertising," 2021. doi: 10.13140/RG.2.2.19912.24323.
- [16] M. Aljabri and R. M. A. Mohammad, "Click fraud detection for online advertising using machine learning," *Egypt. Inform. J.*, vol. 24, no. 2, pp. 341–350, Jul. 2023, doi: 10.1016/j.eij.2023.05.006.

- [17] A. Siu, “How the growth of click and impression farming are getting worse with AI,” Digiday. Accessed: Oct. 30, 2023. [Online]. Available: <https://digiday.com/media-buying/the-growth-of-click-and-impression-farming-are-getting-worse-with-ai/>
- [18] J. R. Alzghoul, E. E. Abdallah, and A. S. Al-khawaldeh, “Fraud in Online Classified Ads: Strategies, Risks, and Detection Methods: A Survey,” *J. Appl. Secur. Res.*, vol. 0, no. 0, pp. 1–25, 2022, doi: 10.1080/19361610.2022.2124328.
- [19] “TalkingData AdTracking Fraud Detection Challenge.” Accessed: Nov. 04, 2023. [Online]. Available: <https://kaggle.com/competitions/talkingdata-adtracking-fraud-detection>.

ВЪЗМОЖНОСТИ И ПРЕДИЗВИКАТЕЛСТВА ПРЕД БАНКИТЕ ОТ ПРИЛОЖЕНИЕТО НА ИЗКУСТВЕН ИНТЕЛЕКТ ПРИ УПРАВЛЕНИЕТО НА КРЕДИТНИЯ РИСК

Кирил Аначков

PhD student at Department of "Finance and Insurance", VUZF, e-mail: anachkov.kiril@gmail.com

Резюме

Целта на изследването е да се анализират възможностите и предизвикателствата, които се откриват пред банките от приложението на изкуствен интелект при управлението на кредитния риск. Изборът на тема е продиктуван, от все по-ключовото значение, което започва да има изкуственият интелект върху банките и потенциала, който има тази нова технология - да революционизира изцяло начина, по който да се оценява и управлява кредитния риск. В структурно отношение изследването се състои от увод, две части, заключение и използвана литература. Изследването е част от дисертационен труд. Методологията в тази част се състои от: теоретичен, сравнителен и критичен анализ. В дисертационния труд се прилага и емпиричен анализ.

Ключови думи: Изкуствен интелект, банки, кредитен риск, подобрене.

JEL: G2, O33, D81.

OPPORTUNITIES AND CHALLENGES FOR BANKS FROM THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN CREDIT RISK MANAGEMENT

Kiril Anachkov

Abstract

The goal of the research is to analyze the opportunities and challenges, that appear before banks through the application of artificial intelligence in credit risk management. The choice of topic is dictated, of increasingly key importance, which artificial intelligence is starting to have on banks and the potential, that this new technology has - to completely revolutionize the way credit risk is assessed and managed. In structural terms, the research consists of an introduction, two parts, conclusion and references. The research is part of a dissertation work. The methodology in this part consists of: theoretical, comparative and critical analysis. In the dissertation work is also applied and empirical analysis.

Keywords: Artificial intelligence, banks, credit risk, improvement.

Увод

Изкуственият интелект (ИИ) има потенциала, да революционизира изцяло начина, по който да се оценява и управлява кредитния риск. Иновативните инструменти и техники, с които разполага, могат да подобрят оперативната ефективност, да доведат до по-бързо вземане на решения, като същевременно водят до по-точни оценки и съответно до намаляване на риска, посредством прилагането на усъвършенствани анализи на огромни количества данни и алгоритми за машинно обучение. Това е и една от причините за постоянно нарастващия интерес към тази тема, която е изключително значима, както за финансовата теория, така и за финансовата практика. Събирането и използването на данни от алтернативни източници и изграждането на нетрадиционни модели за кредитен скоринг има потенциала да осигури по-обхватен поглед за кредитните институции върху кредитоспособността на дадено лице, особено за тези с ограничена кредитна история. Използването на пълния потенциал на изкуствен

интелект може да доведе и до автоматизиране на част от процесите при управлението на кредитния риск. Това от своя страна ще намали човешките грешки и ще подобри скоростта при анализирането на даден кредитополучател и съответно ще подобри клиентското изживяване. Предвид специфичните особености на дейността си, банките постоянно са изправени пред предизвикателства, свързани с риск от различен характер, чието разрешаване до голяма степен зависи от наличието на подходяща подготовка в това число: политики, процедури и практики за редуциране на риска. Именно приложението на изкуствен интелект предоставя новаторски възможности пред банките, но също така открива нови предизвикателства и опасения относно поверителността на данните и съответствието с нормативните изисквания. Наред с правни проблеми като тези, съществуват и отделни етични въпроси, произтичащи от събирането и прилагането на данни от алтернативни източници на информация, по-специално данни, произхождащи от активност в социалните мрежи и доколко е уместно използването им. В изследването ще бъдат разгледани и анализирани всички тези фактори.

1. Възможности пред банките от приложението на ИИ при управлението на кредитния риск

Една от доста обсъжданите теми сред банките през последните години е именно, как изкуственият интелект, може да бъде приложен за подобряване на управлението на кредитния риск. Постоянната дигитализация на процеси в банките и поради огромната перспектива от използването на тази нова технология е важно да се анализира какви положителни резултати, може да доведе приложението ѝ за банките, но и какви рискове крие използването ѝ. Тази тема е нова за финансовата теория и практика и затова свидетелства липсата на задълбочени изследвания, които да отговорят на конкретните въпроси.

Процесът по точното оценяване на вероятността даден кредитополучател да не погаси задълженията си е едно от най-големите предизвикателства пред банките. Правилната оценка на кредитния риск е от първостепенно значение за успешното функциониране на дадена банка. Неправилната оценка на кредитния риск, може да доведе до ипотечни кризи, като тази през 2008г. Поради тази причина банките непрекъснато търсят начин за усъвършенстване на съществуващите кредитни модели, които прилагат, като отделят все по-значително количество ресурси, за тяхното надграждане, за да могат да оценят по-точно риска и да предвидят кредитоспособността на даден кредитополучател възможно най-точно. Водещи банки в света, като JPMorgan и HSBC от няколко години започнаха да изследват най-различни варианти за използването на пълния потенциал на ИИ при управлението на кредитния риск, тъй като приложението му може да доведе до подобрене чрез:

- **Намаляване на измамите при кандидатстване за кредит** – С приложението на ИИ, може да се анализират големи масиви от данни в реално време, като същевременно има възможността, да разпознава различни модели и аномалии, които могат да представляват измама и да ги идентифицира навреме. Това от своя страна предоставя по-големи възможности на банките за ограничаване на измамите.
- **Автоматизиране на процеси по управление на кредитния риск** – Автоматизацията на част от процесите при кандидатстване за кредит, ще доведе до редуциране на човешките грешки при вземане на решения чрез предварително дефинирани правила, като същевременно ще доведе и до оптимизиране на времето за одобрение на даден кредит, а това от своя страна и до подобряване на клиентското изживяване.
- **Подобряване на моделите за кредитен скоринг** – Иновативните инструменти и техники, с които разполагат алгоритмите базирани на ИИ, могат да анализират огромни количества масиви от данни от различни източници на информация в реално време, което да позволи да се вземат много по-точни и правилни решения. В същото време приложението на тази технология, може да доведе до намаляване на разходите и времето за оценка. ИИ разполага с възможността да анализира и такива данни, които са от алтернативни източници или са неструктурирани. С използването на данни от алтернативни източници на информация при изграждането на нетрадиционни модели за кредитен скоринг, банките ще имат възможността да

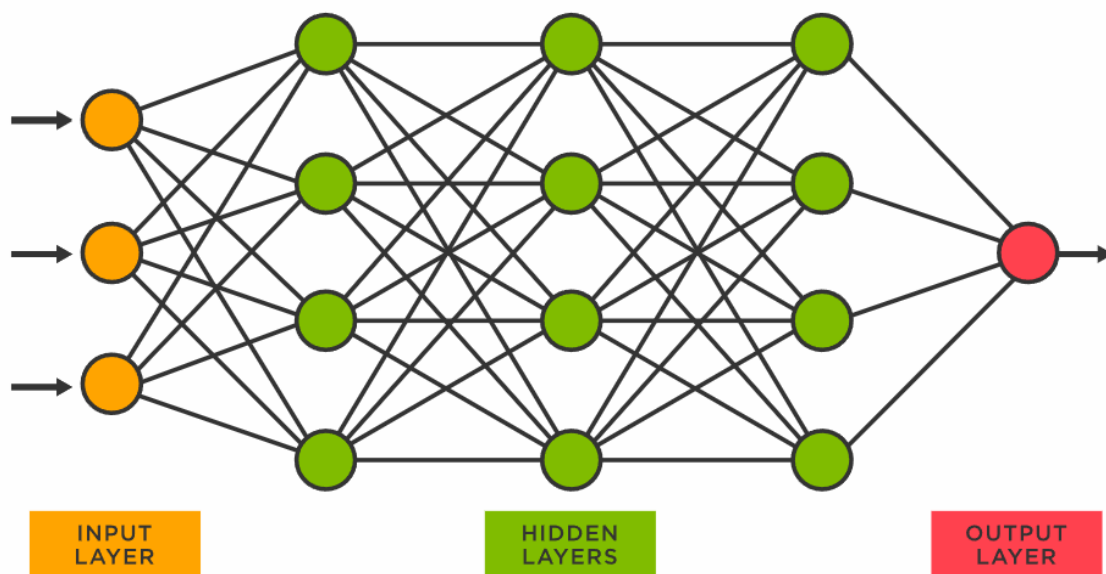
осигурят по-обхватен поглед върху кредитоспособността на дадено лице, особено при тези кандидати, които са с ограничена кредитна история. Това от своя страна съответно има възможността да доведе до намаляване на риска от неизпълнение.

- **Наблюдение на кредита, след отпускането му** – Изкуствения интелект предоставя възможността да се използва и като инструмент, който помага за рационализиране на неефективни процеси и също така за подобрене на наблюдението на клиентски акаунти след отпускане на кредита. Използвайки широк набор от информация и изчислителната мощност, която притежават алгоритмите за ИИ могат да генерират предупредителни сигнали много по-рано от традиционните модели на банката, като същевременно позволява да се извършва непрекъснат мониторинг на кредитоспособността на даден кредитополучател в реално време. Според публикация на Claybaugh (2020), ИИ разполага също с възможността да предвиди вероятността даден кредитополучател да стане просрочен 60 дни, преди това да се случи, като същевременно може да наблюдава всички кредитополучатели с еднакъв фокус. Традиционно кредитните мениджъри се фокусират върху големите кредити и пренебрегват наблюдението на по-малките. Това означава, че така наречените по-малки кредитополучатели, на които не им се извършва постоянен мониторинг са изложени на по-голям риск от просрочие, като шансовете за възстановяване намаляват с времето. Именно с приложението на ИИ може да се осигури едновременното наблюдение, както на големи кредитополучатели, така и на малки с еднакъв фокус, което от своя страна ще доведе до точна представа за ситуацията на всеки отделен клиент в реално време.
- **Създаване на по-точни модели за бъдещото поведение на даден кредитополучател** – Използването на огромни масиви исторически данни и с помощта на техниките на машинното обучение, алгоритмите базирани на ИИ могат да разработят много по-точни модели за вероятността от неизпълнение или просрочие на дадена експозиция, което ще позволи на банките да вземат навременни и по-информирани решения. Развитието и усъвършенстването на тази технология има капацитета да предоставя и стратегии за намаляване на риска на даден портфейл.

Изкуствения интелект е в състояние да провежда моделиране на контрола на риска въз основа на използването на стотици бази данни и по-този начин да оцени много по-добре кредитополучателя, като бъдат изработени профили на клиентите и бъдещото им поведение и това да доведе до редуцирането на кредитния риск в дадената банка. В свое изследване Bezv и Domanska (2022) посочват, че традиционният анализ на кредитния риск често разчита на сложни статистически модели, които изграждат връзки между различните характеристики под формата на математически уравнения. На свой ред изкуствения интелект използва методи за машинно обучение (ML), които могат да се учат от данни, без да изискват програмиране, базирано на правила, както е при традиционните модели, при които трябва да бъдат ясно дефинирани правила преди това. Това от своя страна предоставя гъвкавост, адаптивност и голям потенциал за усъвършенстване на моделите за машинно обучение, пред традиционните модели. В резултат на тази гъвкавост методите за машинно обучение могат по-добре да пасват на моделите за данни от традиционните такива. В изследването се описват основните модели, които банките използват или могат да приложат в близко бъдеще и това, да доведе до подобрене в тяхното оценяване на кредитния риск:

- **Логистична регресия** - Представлява контролиран алгоритъм, който се прилага основно за моделиране на вероятността за случване на събитие с помощта на различни типове данни. При използването му първоначално се задава набор от независими променливи. Като се прогнозира зависимата променлива от данни, чрез анализирането на връзката между независимите променливи. Може лесно да определи променливите използвани за класификация.

- **Линейна регресия** – Статистически регресионен метод, който принадлежи към контролираното машинно обучение и е един от най-прилаганите алгоритми при прогнозен анализ. Той се използва за оценка на данните и установяване на определена връзка между две или повече променливи. Неговата основна цел е въз основа на независимите променливи, да предскаже резултата от бъдещо събитие. Линейната регресия може да бъде проста или множествена, в зависимост от броя на независимите променливи. Алгоритъмът за линейна регресия показва линейна връзка между една или повече независими променливи (x), в зависимост от това дали е проста или сложна регресия и зависимата променлива (y). Целта е да се състави най-подходяща права линия, въз основа на данните, която описва корелацията между независимите променливи със зависимата. Когато стойността на независимата променлива нараства, стойността на зависимата също се увеличава.
- **Поддържаща векторна машина (SVM)** – е метод на машинното обучение, който се използва за създаване на регресия, класификация и откриване на отклонения. Този метод се използва главно за решаване на проблеми с класификацията. Видовете опорно векторни машини, могат да бъдат - линеен и нелинеен, в зависимост от типа данни. В поддържаща векторна машина големия марж се счита за добър. Основната цел при използването на този модел е да намери най-добрата линия или граница на решение, това е така наречената хиперравнина, която разделя различните класове данни и спомага за правилното решение на даден казус. Целта е да се намери най-близката точка до хиперравнината от сините и червените точки от двата класа. Най-близките точки до хиперравнината от двата класа данни са опорните вектори.
- **K-Nearest Neighbors Algorithm (KNN)** – Този алгоритъм се базира на техниките за контролирано обучение и е един от най-лесно изграждащите се алгоритми за машинно обучение. Той се използва основно за решаване на проблеми с класификацията и за регресия. Работи на принципа да идентифицира най-близкия „съсед“. Целта е да се определи кои точки от данни са най-близки съседи до определена точка, като трябва да се изчисли разстоянието от тази точка до другите точки от данни. Тези показатели за разстояние помагат да се формират граници за вземане на решения, които разделят точките на запитване в различни региони. Може да помогне на банките да оценят риска от заем за организация или физическо лице. Използва се за определяне на кредитоспособността на кредитоискателя (What is the k-nearest neighbors algorithm?, 2023).
- **Random forest algorithms** – Използва се за решаване на проблеми с класификацията, посредством събиране на голям набор от исторически данни в така наречените „дървета на решения“, в които целта е да се оптимизира бъдещия процес на анализ, като се постигне един резултат.
- **Невронни мрежи (NNs)** – Невронните мрежи, известни също като изкуствени невронни мрежи (ANN) или симулирани невронни мрежи (SNN), са подгрупа на машинното обучение и са в основата на алгоритмите за дълбоко обучение. Името и структурата им са вдъхновени от човешкия мозък, имитирайки начина, по който биологичните неврони сигнализират един на друг (What is a neural network ?,2023). Те използват неврони в слоеста структура или взаимосвързани възли, наподобяваща човешки мозък. Тяхната цел е компютрите да бъдат научени да обработват огромни количества данни, по начин подобен на човешкия мозък. Това предполага огромен брой на възловите слоеве, през които данните преминават в многоетапния процес на разпознаване на образи. Невронната мрежа е изградена от няколко слоя невронни, като информацията се предава между различните слоеве. На фигура 1. е илюстрирана дълбока невронна мрежа.



Източник: What is a neural network?, (Spotfire, 2022).

Фигура 1. Невронна мрежа

Всички посочени алгоритми, позволяват работа с големи и оптимизирани масиви от данни и създават подходяща среда за използване на нелинейни модели на различни променливи. В свое изследване Basrai и Ali (2021) посочват, че към настоящия момент банките използват традиционни модели за кредитен риск, за да предскажат категорични, непрекъснати или двоични променливи на резултата, тъй като ML моделите са трудни за тълкуване и не са лесно проверими за регулаторни цели. Независимо от това, те могат да се използват за оптимизиране на параметрите и подобряване на процеса на избор на променливи в съществуващите регулаторни модели. Базираните на изкуствен интелект техники за решения могат да доведат до лесно проследими и логични правила за вземане на решения, въпреки че имат нелинейни знаци. Техниките за обучение без надзор могат да се използват за изследване на данните за традиционно моделиране на кредитния риск, докато методите за класификация като поддържащи векторни машини могат да предскажат ключови характеристики на кредитния риск, като вероятността за неизпълнение или загуба при неизпълнение. Например, при оценка на кредитния риск, изкуствения интелект може да бъде приложен за усъвършенстване на кредитния скоринг чрез модели базирани на поведението на даден клиент при покупки в интернет, разрешени от кандидата данни от неговия мобилен телефон и чрез анализ на поведението му при погасяването на други задължения. Това несъмнено ще допълни традиционния кредитен модел и може ефективно да оцени готовността на кредитополучателя за изплащане на задълженията му. Основните софтуерни приложения в областта на моделирането на кредитния риск, с което се извършва всичко това са: SAS, R и Python.

2. Рискове и предизвикателства, които могат да възникнат пред банките от приложението на ИИ при управлението на кредитния им риск

За да бъде цялостен анализа е важно, да се разгледа и какви предизвикателствата и рисковете крие пред банките използването на изкуствен интелект при управлението на кредитния им риск. За да не бъдат допуснати грешки при използването му, той трябва да бъде напълно оценен и да се анализират възможните негативни ефекти, които могат да се появят при приложението му. Рискът от предоверяване в алгоритмите за изкуствен интелект при вземане на решения съществува и трябва да се внимава с това. Необходимо е да се контролира използването на алгоритми; теоретично да се доказва безопасността и стабилността на алгоритмите в различни сценарии и да се намали до минимум използването на алгоритми с голяма волатилност, за да не се стигне до проблеми за банките, които ги прилагат. Това е голямо предизвикателство пред ИИ и до голяма степен поражда съмнения в приложението му, поради трудно обясними резултати

при част от взетите решения. В научни изследвания е доказано, че за непараметрични и нелинейни модели, съществуват алгоритмични рискове свързани с вземане на решения, в така наречената „черна кутия”, при която има риск, да не може да бъде обяснена връзката между входът - изходът на модела, т.е. съществува риск от прекъсване в модела и данните да не са достоверни. Следователно има голяма вероятност алгоритмичните черни кутии и алгоритмичните отклонения да са реални предизвикателства и при приложението на ИИ за изработване на профил на кредитополучателя и бъдещото му поведение.

Освен предизвикателствата от така наречените черни кутии при използването на ИИ може да се появят и други рискове. До този момент те са били в обсега на вземане на решение от човек, а не на технологично ниво. Това са следните видове риск, които могат да окажат ефект върху кредитната дейност на банката:

- Рискове свързани с данните – включва: ниско качество на данни; проблеми с поверителност и защита на данните и необективни данни, които ще доведат до грешки при вземане на решение за даден кредитополучател.
- Рискове в производителността – включва: работа над техническите възможности и отклонение в производителността. Тук е важно да се отбележи, че както при традиционните модели за оценка на кредитния риск, проблемите с производителността могат да възникнат при не добро калибриране на модела или некачествени данни. В случая, обаче сложността на модела изграден на база ИИ прави по-трудно да се оцени дали в резултатите на модела има допуснати грешки.

Подобно на традиционните статистически модели, които банките използват, моделите базирани на ИИ също могат да допуснат грешки, което от своя страна да доведе до непредвидими последици. Много е важно изграждането на даден модел, да бъде извършено по правилен начин и моделът да бъде добре структуриран. Важно е също как ще бъде приложен, тъй като неправилното му използване от своя страна може да доведе до резултат, който да е ненужен или дори да окаже негативен ефект на банката, която го прилага. При конструкцията на алгоритъма трябва да има ясно определена цел, която да бъде приведена в съответствие с бизнес проблема. При конструиране на даден алгоритъм, ако не са взети под внимание всички фактори, има голяма вероятност моделът да е неточен. По отношение на набора от данни, несъмнено ИИ превъзхожда традиционните статистически модели, но тук е много важно как ще бъде структурирана цялата тази информация и как ще бъде калибриран модела. Преди модела да бъде внедрен в дадена банка, задължително трябва да бъде щателно тестван, като се определи дали е изграден правилно, подходящ ли е за предназначението му и да се направят няколко теста как работи в сравнение с традиционен статистически модел. Правилното използване на сложни модели е от ключово значение за вземането на правилни решения.

Освен разгледаните рискове приложението на изкуствен интелект поражда и опасения относно поверителността на данните и съответствието с нормативните изисквания. Регулаторите особено следят банките дали спазват всички изисквания за поверителност на личните данни. От една страна, банките желаят да измерват риска възможно най-точно чрез използването на колкото е възможно по-голям набор от информация, но от друга страна съществува и нуждата от защита на личните данни на клиентите. В изследване на Sadok, Sakka и Maknoui (2022) се посочва, че конструирането на моделите за анализ и оценка на риска базирани на изкуствен интелект и обработващи все по-големи масиви от данни най-вероятно ще събират информация и от алтернативни канали и ще трябва да бъдат възложени на външен субект. Това от своя страна излага банката на риск в случай на нарушаване на протоколите за поверителност. Възлагането на външни изпълнители на оценка на риска също така повдига свързания въпрос за гарантиране, че данните, прехвърлени от банката към оценителя на риска, няма да продължат да се съхраняват допълнително след завършване на разработването на модела за оценка на риска.

Наред с правни проблеми като тези, съществуват и отделни етични въпроси, произтичащи от прилагането на данни от алтернативни източници на информация, по-специално данни, произхождащи от активност в социалните мрежи и доколко е уместно използването им. Например при използването на данни от социални мрежи: Клиент, чийто възможност за

отпускане на кредит е по-нисък, защото при равни други условия той или тя има лоши платци като контакти в социалните мрежи. Този вид резултат *сам по себе си* не е оправдан от правна гледна точка, но очевидно поставя етичен проблем. Какви съображения могат да помогнат при вземането на решения относно социалната приемливост на тези употреби на данни от социални мрежи? Изследването на Óskarsdóttir et al. (2019) на тема „The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics“ предлага решение, точно за това предизвикателство, вдъхновено от производствените табла, като се добави нулева стойност на сегменти от данни, които биха поставили в неравностойно положение кредитополучателите при изготвянето на кредитния модел (Sadok, Sakka and Maknoui, 2022). Въпреки предизвикателствата и рисковете, които съществуват пред банките от използването на изкуствен интелект, те могат да бъдат преодоляни напълно, стига да не се взимат прибързани решения. С правилна, подготовка, координация и решения между различните отдели, предизвикателствата могат да бъдат минимизирани. Приложението на тази нова технология е необходимо, тъй като освен за подобряване на кредитния риск, използването и анализирането на големи масиви информация, ще позволи на банките да разбират по-добре нуждите на своите клиенти, да съкратят времетраенето на част от процесите и съответно това да доведе до подобряване на клиентското преживяване.

Заклучение

От направения анализ, следва да се отбележи, че с приложението на изкуствен интелект при управлението на кредитния риск, могат да бъдат решени проблеми, които обикновено са сложни и нелинейни и е трудно да бъдат разрешени с традиционен статистически модел. Това е и основната причина, поради която банките искат да внедрят и използват пълния потенциал на изкуствен интелект при управлението на кредитния риск. Използването му ще донесе голям напредък за банките и ще улесни процеса на работа, чрез подобряване на оценката на кредитния риск, откриването на измами, мониторинга и събираемостта, но не трябва да се прибързва с пълната интеграция на тази технология без да са изследвани всички ефекти от приложението и. Най-голямата опасност от нейното използване е, именно от това да се направят прибързани изводи, без изобщо да са запознати подробно с това как работи и всичките ефекти, които могат да последват от използването ѝ.

В близко бъдеще се очаква изкуствения интелект да играе основна роля за успешното управление на кредитния риск в банките и да доведе до подобряване на управлението на кредитните експозиции и по-добро позиционирането на пазара на банките, които съумеят да го приложат. Въпреки, че в изследването е анализирано единствено приложението на ИИ при управлението на кредитния риск, използването му несъмнено ще окаже влияние и върху другите дейности на банките. Важно е да се отбележи, че банките, които успеят по-бързо и правилно да приложат интелигентни системи за подпомагане на вземането на решения, базирани на технологията за изкуствен интелект чрез използването на усъвършенствани модели, включително и такива от алтернативни източници на информация, ще имат предимство пред конкурентите си, които залагат на традиционните статистически модели и ще постигнат по-големи положителни резултати.

References

1. Basrai, A. & Ali, S., 2021. Artificial Intelligence in Risk Management, KPMG, [online], available at: < <https://kpmg.com/ae> Accessed [09 October 2023];
2. Bevez, R. & Domanska, O., 2022. Artificial Intelligence (AI) for Credit Risk Management in Banking, Avenga, [online], available at: < <https://www.avenga.com> Accessed [03 August 2023];
3. Claybaugh, T., 2020. Leveraging Artificial Intelligence Throughout the Credit Management Process, MSCCM, [online], available at: < <https://www.msccm.com> Accessed [09 June 2023]
4. Óskarsdóttir, M., Bravo, C., Sarraute, C., Vanthienen, J., & Baesens, B. (2019). The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network

- analytics. *Applied Soft Computing*, 74, 26–39., [online], available at: < <https://doi.org/10.1016/j.asoc.2018.10.004> Accessed [08 October 2023];
5. Sadok,H., Sakka, F. & Maknouzi, M., 2022. Artificial intelligence and bank credit analysis: A review, *Cogent Economics & Finance*, 10:1 [online], available at: < <https://www.tandfonline.com/doi/full/10.1080/23322039.2021.2023262> Accessed [08 October 2023];
 6. What is a neural network?, 2023., IBM, [online], available at: < <https://www.ibm.com/topics/neural-networks> Accessed [02 October 2023]
 7. What is a neural network?, 2022., Spotfire, [online], available at: < <https://www.spotfire.com/what-is-a-neural-network> Accessed [08 October 2023];
 8. What is the k-nearest neighbors algorithm?, 2023., IBM, [online], available at: < <https://www.ibm.com> Accessed [09 September 2023].

ГОЛЕМИ ДАННИ В ЗДРАВЕОПАЗВАНЕТО

Любен Зюмбилски

Ph.D. candidate, Department of Information Technologies and Communications, University of National and World Economy, e-mail: lzyumbilski@unwe.bg

Резюме

Големите данни играят решаваща роля в трансформацията на съвременните здравни грижи. В доклада са представени ключовите начини, по които големите данни се използват в здравеопазването. Тези начини включват прогнозите по отношение на медицината и диагностиката, персонализираните здравни грижи, управлението на болнични и здравни заведения, научни изследвания и обработка на образи, мониторинг на общественото здраве и управление на хронични заболявания. Големите данни предоставят ценни инструменти за подобряване на качеството на здравни грижи и научните изследвания в сферата на здравеопазването.

Ключови думи: дигитализация, здравеопазване, данни, здраве, големи данни.

BIG DATA IN HEALTHCARE

Luben Zyumbilski

Abstract

Big data is playing a crucial role in the transformation of modern healthcare. This abstract presents the key ways big data is being used in healthcare. Paper includes predictive medicine and diagnostics, personalized healthcare, hospital and healthcare facility management, research and imaging, public health monitoring, and chronic disease management. Big data provides valuable tools to improve the quality of healthcare and healthcare research.

Key words: digitization, healthcare, data, health, big data.

Въведение

Здравеопазването е един от секторите, в които обработката на големи данни играе ключова роля за подобряване на качеството на медицинските услуги, оптимизиране на процесите и разработване на иновации в медицинската практика.

Големите данни се явяват ключов компонент в модерното здравеопазване. Този вид данни представлява огромни обеми информация, които се събират от различни източници. Обработват се с помощта на технологични инструменти и се анализират за целите на здравната наука и практика. Големите данни предоставят значителни възможности за подобрене на качеството на медицинските грижи и процесите в здравеопазването.

Същност и приложение на големите данни в здравеопазването

Големите данни имат широк спектър на приложение в здравеопазването и са важен инструмент за подобряване на грижата за пациентите и управлението на здравната система. В този доклад по-нататък ще разгледаме някои от основните приложения на големите данни в тази област:

1. Информационни системи и електронни здравни досиета (Electronic Health Records - EHR)

Те играят съществена роля в използването на големите данни в здравеопазването. Събират, съхраняват и управляват медицинската информация за пациентите в цифров формат, като предоставят по-лесен и бърз достъп до данните за лечение и здравно състояние. Ето някои ключови аспекти и предимства на информационните системи и електронните здравни досиета:

- **Централизиран достъп:** EHR съхраняват информацията за пациентите на едно централно място, което улеснява лекарите, медицинския персонал и здравните специалисти да получават бърз и лесен достъп до важни медицински данни.
- **Интеграция на данни:** Информационните системи могат да интегрират различни видове медицинска информация, включително лабораторни резултати, прегледи, рентгени, рецепти, история на заболяването и други. Това позволява на лекарите да видят цялостната картина за пациента и да вземат по-добри решения за диагностика и лечение.
- **Споделяне на данни:** EHR системите могат да спомогнат за бързото и сигурно споделяне на информация между различни здравни заведения и специалисти. Това е от особена важност при трансфер на пациенти между болници, сътрудничество на различни специалисти и консултации от разстояние.
- **Избягване на двойни /дублирани/ изследвания:** Информационните системи могат да предотвратят излишни и дублиращи се медицински изследвания, като позволят на лекарите да видят предишните процедури и резултатите от тях.
- **Използване на големи данни:** EHR системите събират огромно количество медицински данни, които могат да бъдат използвани за анализ и статистика. Това позволява на изследователите да извличат ценна информация за здравето на населението, ефективността на терапиите и други медицински тенденции.
- **Безопасност и точност:** Внедряването на EHR системи намалява риска от човешки грешки при създаване и интерпретация на медицинската информация. Това води до по-голяма точност и безопасност при грижата за пациентите.
- **Спестяване на време и ресурси:** Електронните здравни досиета съкращават необходимостта от ръчно създаване и съхранение на хартиени документи, като подпомагат оптимизацията на работния процес в здравеопазването.

2. Анализ на данни за диагностика

По-точната диагностика е един от основните аспекти на използването на големи данни в здравеопазването и представлява ключова предимство. Ето няколко начина, по които големите данни допринасят за по-точната диагностика:

- **Анализ на медицински изображения:** Медицински изображения като резултати от рентгенографии, КТ и МРТ се използват широко за диагностика на заболявания. Големите данни позволяват използването на алгоритми за компютърно виждане и машинно обучение, които могат да анализират милиони изображения и откриват скрити патологии, което би било трудно или невъзможно за обработка от човешки ресурс.
- **Интегриране на множество източници на данни:** Големите данни позволяват комбинирането на информация от различни източници като лабораторни резултати, клинични записи, генетични данни и данни от сензорни устройства. Този комбиниран подход предоставя по-цялостна картина за здравословното състояние на пациента и помага за по-добра диагностика.
- **Прогностични модели:** Големите данни могат да се използват за създаването на прогностични модели, които могат да помогнат при ранното откриване на заболявания. Тези модели анализират данни от множество пациенти и предсказват риска от развитие на конкретни заболявания при индивида.
- **Геномика и молекулярна медицина:** Големите данни са от съществено значение в геномиката и молекулната медицина, където се анализират генетичните данни и молекулярни маркери, свързани с конкретни заболявания. Това помага да се постави точна диагноза и да се избере персонализирано лечение.
- **Постоянен мониторинг:** Устройства и сензори, свързани с интернет, позволяват постоянен мониторинг на здравето на пациентите. Тази информация може да предостави реални данни

за състоянието на здравето и да предупреди за промени в него, което подобрява възможността за своевременна диагностика и лекарска намеса.

Пълната диагностика, осигурена чрез големите данни, е от съществено значение за пациентите, тъй като тя води до по-бързо лечение, по-малко грешки и по-добро управление на заболяванията. Това в резултат предоставя по-добра грижа и повишава шансовете за успешно лечение.

3. Мониторинг на пациентите

Мониторинг и проследяване на пациентите с използване на големи данни представлява важен компонент от съвременните здравни грижи. Този подход позволява непрекъснато наблюдение на здравето на пациентите, предоставя информация за тяхното състояние и активност и дава възможност за бърза реакция при нужда. Ето някои ключови аспекти и приложения на мониторинга и проследяването на пациентите чрез големи данни:

- **Устройства и сензори:** Пациентите могат да носят различни видове устройства и сензори, като смарт часовници, монитори за сърдечна дейност, глюкомери и други. Тези устройства събират данни за пулс, кръвно налягане, нива на глюкоза и други биометрични показатели.
- **Телемедицина:** Телемедицината позволява на лекарите да провеждат консултации с пациентите от разстояние, като използват видеокомуникации и други технологии. Това е особено полезно при проследяване на пациенти с хронични заболявания или при предоставяне на специализирана медицинска грижа в отдалечени райони.
- **Аларми и предупреждения:** Устройства и сензори могат да генерират автоматични аларми и предупреждения, които се изпращат до пациентите и доставчика на здравни услуги. Например, при рязко увеличение на сърдечната честота или намаляване на нивата на кислород в кръвта, пациентът може да получи предупреждение. Лекуващият лекар- също.
- **Следене на хронични заболявания:** Мониторингът е от съществено значение за пациентите с хронични заболявания като диабет, белодробни заболявания, сърдечни проблеми и много други. Той предоставя постоянна обратна връзка и информация на пациентите и техните лекари, което помага за контролиране на заболяванията и предотвратяване на спешни случаи.
- **Следене на физическата активност:** Устройства, които се използват за мониторинг на физическата активност и фитнес показателите на пациентите. Тези данни могат да бъдат използвани за предотвратяване на липсата или намалена активност и подпомагане за водене на здравословен начин на живот.
- **Събиране и анализ на данни:** Данните, събрани чрез мониторинг, се съхраняват и анализират чрез големи данни. Това предоставя възможност за извличане на ценна информация от данните, като предупреждаване за здравни рискове и тенденции.

Мониторингът и проследяването на пациентите чрез големи данни допринасят за по-ефективната грижа и подобряват качеството на живот на пациентите, като им предоставят по-голям контрол и информация относно техните здравни състояния. Този подход е особено полезен за пациентите с хронични заболявания и за тяхното дистанционно следене и грижа.

4. Персонализирано лечение

Персонализираното лечение, известно още като прецизно лечение или индивидуализирана медицина, е подход в здравеопазването, който се стреми да предостави лечение, базирано на индивидуалните характеристики и нужди на пациентите. Този подход се основава на използването на големи данни, геномика и други технологии, които позволяват на лекарите да адаптират лечението към конкретния пациент, вместо да следват един общ протокол. Този подход е значим напредък в медицината и има няколко ключови аспекта:

- **Геномика:** Геномиката е наука, която изучава генетичната информация на индивида. С използването на геномни анализи, лекарите могат да определят специфични генетични предпоставки и мутации, които могат да повлияят на здравето на пациента и на отговора му на определени лекарства. Тази информация позволява персонализирано подбиране на лечение и лекарства, които са по-ефективни и с по-малко странични ефекти.
- **Фармакогеномика:** Фармакогеномиката се занимава с изучаване на това как генетичните фактори влияят на реакцията на индивида към лекарства. Тази информация помага да се изберат правилните лекарства и да се предотвратят опасни реакции.
- **Подход, базиран на данни:** Големите данни се използват за анализ на голям обем медицинска информация и съпоставяне на данните за множество пациенти. Този подход позволява идентифицирането на общи тенденции, ефективни лечебни стратегии и най-добри практики.
- **Подобрено следене и мониторинг:** Пациентите, получаващи персонализирано лечение, често се подлагат на по-редовно проследяване и мониторинг. Това включва периодични изследвания и тестове, които се използват за следене на реакцията на пациента към терапията.
- **Борба с хронични заболявания:** Персонализирането на лечението е особено полезно при хронични заболявания като диабет, сърдечни болести, белодробни заболявания и много други. Това позволява на пациентите да получат грижа, която е съобразена с тяхната индивидуална история и състояние.

Персонализираното лечение представлява напреднал и иновативен подход към медицината, който се стреми да подобри резултатите на лечението и да намали страничните ефекти и неефективния избор на терапии. С използването на големи данни и иновативни технологии, персонализираният подход към здравеопазването става все по-реален и обещаващ за бъдещето на медицината.

Предизвикателства

При използването на големи данни в медицината се появяват някои предизвикателства, които трябва да бъдат своевременно решени.

1. **Сигурност на данните:** За да се използват големите данни в здравеопазването, трябва да бъде гарантирана сигурността на личните здравни данни на пациентите. Съществуват и правни регламенти в тази насока, които трябва да се спазват.

2. **Липса на стандартизация:** Медицинската информация често се събира в различни формати и от системи, което създава необходимост от стандартизация на данните.

3. **Етични въпроси:** Събирането и използването на медицински данни, трябва да става при зачитане на етичните норми и правата на пациентите.

Заклучение

Големите данни играят ключова роля в подобряването на здравеопазването, като предоставят възможности за по-добра диагностика, лечение и превенция на болести. Въпреки предизвикателствата, свързани със сигурността, стандартизацията и етиката, използването на големи данни в здравеопазването обещава да допринесе за по-ефективната и персонализирана грижа за пациентите, както и за по-голямата възможност за изследване на здравето и заболяванията. Разбира се, всичко това би оказало положително влияние върху мениджмънта и всички икономически нива в структурите на здравеопазването.

Literature

Davenport, T., Kalakota, R. (2019) The potential for artificial intelligence in healthcare. *Future Healthc Journal*, 2019 Jun; 6(2): 94–98.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181/#S0002title>

Hristov Georgi, Markova M. (2022) Reporting of R&D Costs Under IAS 38 - The Case of Biopharmaceutical Companies. 4/2022, Научни трудове на УНСС.

Pranav Rajpurkar, Emma Chen, Oishi Banerjee, Eric J. Topol. (2022) AI in health and medicine.

Nature Medicine | VOL 28 | January 2022 | 31–38 |

https://drive.google.com/file/d/1TURwltbv5B1Zg0DbE4bHO_tJNfh4GHc/view

КАТАЛОГИЗИРАНЕ НА ДАННИ В СРЕДА НА ГОЛЕМИ ДАННИ

Гено Стефанов

главен асистент, доктор, катедра Информационни технологии и комуникации, УНСС,
e-mail: genostefanov@unwe.bg

Резюме

С непрестанното нарастване на обемите от данни и в среда на Големи данни, правилното и навременно каталогизиране на данните играе важна роля в тяхната обработка и откриване. Съвременните среди за Големи данни неминуемо имат някакъв вид каталогизиране на данните под формата на различни по вид технологии и инструменти. Целта на настоящия доклад е да анализира възможностите и предизвикателствата пред каталогизирането на данни в контекста на Големи Данни. Освен това ще се разгледат конкретни техники и технологии въведени от Amazon Web Services(AWS) - AWS Catalog и AWS Glue Crawler.

Ключови думи: Големи Данни, Каталогизиране на данни. Big Data, Data Catalogs.

DATA CATALOGING IN BIG DATA ENVIRONMENT

Geno Stefanov

Abstract

With the ever-increasing volumes of data and in a Big Data environment, proper and timely cataloging of data plays an important role in its processing and discovery. Modern Big Data environments inevitably have some form of data cataloging in the form of various technologies and tools. The purpose of this report is to analyze the opportunities and challenges of data cataloging in the context of Big Data. In addition, specific techniques and technologies provided by Amazon Web Services (AWS) - AWS Catalog and AWS Glue Crawler - will be analyzed.

Key words: Big Data, Data Catalogs.

1. Въведение

Каталогизирането на данни в областта на Големите данни се отнася до централизираната система за съхранение и управление на метаданни, която съдържа информация за различните набори от данни в екосистемата на организацията. Каталогизирането на данни може да се представи като създаване на инвентар на данните, с помощта на който може да се търсят, оценяват и използват данните от потребителите. Каталогът за данни има решаващо значение за улесняването на управлението на данни, търсенето на данни и използването им в големи и сложни среди с данни, каквито са Големите данни.

2. Каталогизиране на данни

В основата на каталозите за данни са метаданните. Според стандарта ISO 11179 метаданните се дефинират като данни, които описват и дефинират други данни. Всеки потребител на компютър използва метаданни: запазването на документ в папка и даването му на определено име означава активно присвояване на атрибути на метаданни като име на файл и път до файла. Тези метаданни

могат по-късно да се използват за проследяване на файла, когато е необходимо. Има различни видове групи метаданни [2], [3]:

- описателни метаданни: насочени към подобряване на откриваемостта и разбирането на съдържанието

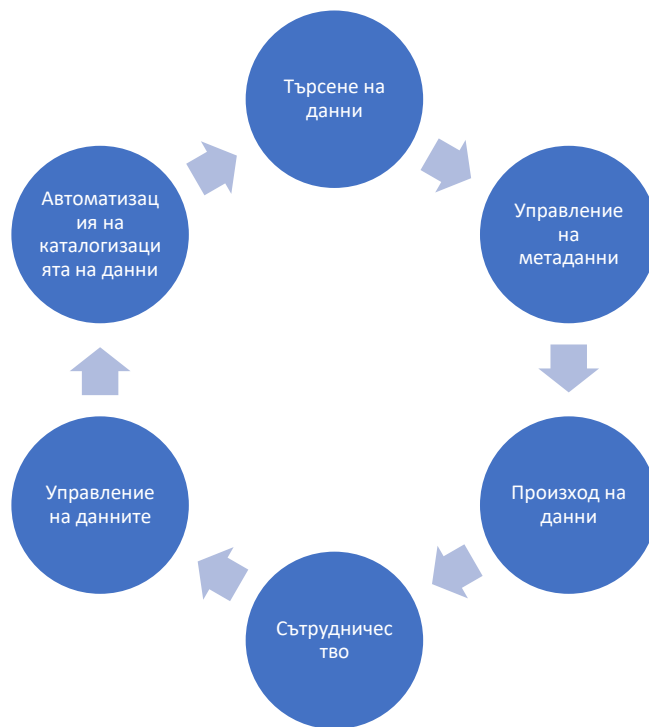
- административни метаданни: могат да бъдат разделени на множество подкатегории като технически, съхранение и разрешения. Тези метаданни дават контекст по отношение на техническия аспект като файлов тип и декодиране, чувствителност и атрибути за достъпност.

Някои автори [1] описват каталогът за данни като „Каталогът за данни поддържа инвентаризация на активи с данни чрез откриване, описание и организиране на набори от данни. Каталогът предоставя контекст, за да позволи на анализаторите на данни, специалисти по данни, администратори на данни и други потребители на данни да намерят и разберат подходящ набор от данни с цел извличане на бизнес стойност,„.

Тъй като каталогизирането на данните разчита на метаданни, качеството на метаданните е от изключително значение. Множество проучвания са фокусирани върху това как може да се оцени качеството на метаданните [4], [5] и подчертават, че липсата на качествени метаданните може да доведе до намалена възможност за търсене и достъпност [4]. Тъй като метаданните се оказват толкова ключов елемент от използваемостта на каталога за данни, всички инструменти за каталогизиране на данни се очаква да имат някакъв вид модул за управление на метаданни в своя софтуер.

Основни характеристики и ползи от каталогизирането на данни в областта на Големите данни включват:

1. Управление на метаданни: Каталогът за данни съхранява детайлни метаданни за наличните данни в Големите данни, включително информация за източниците на данни, произхода на данните, качеството на данните, схемата и други. Тези метаданни имат за цел да подпомогнат разбирането на съдържанието и контекста на данните, което улеснява оценката им за актуалност и надеждност.
2. Търсене на данни: Каталогизирането на данни позволява лесно и бързо търсене и изследване в каталога на данни, за да се открият подходящи набори от данни за анализи или репорти. Това може значително да намали времето, отделено за търсене на данни и достъп до тях.
3. Произход на данни: Информацията за произхода на данните в каталога показва началото и трансформациите, през които данните преминават по време на обработката. Това помага да се разбере как данните се трансформират и преобразуват, което допринася за по-доброто разбиране на данните.
4. Сътрудничество: Каталогите за данни често предоставят функции за сътрудничество, които позволяват добавянето на анотации, коментари, оценки и рецензии към наборите от данни. Това насърчава сътрудничеството между потребителите на данните, анализаторите на данни и лицата, отговорни за управлението на данните.
5. Управление на данните: Каталогите за данни подпомагат усилията по управление на данните, като налагат политики за достъп до данните, гарантират съответствие с регулациите за данните и осигуряват видимост върху това, кой достъпва и използва данните.
6. Автоматизация на каталогизацията на данни: В средите с Големи данни от изключително значение е да има подходящи инструменти за автоматичното каталогизиране на данните в Каталог на данни. Това ще подобри възможностите за бързо откриване на връзки между данните.



Фигура 1: Основни характеристики и ползи от каталогизирането на данни

3. AWS Data Catalog и AWS Glue Crawler

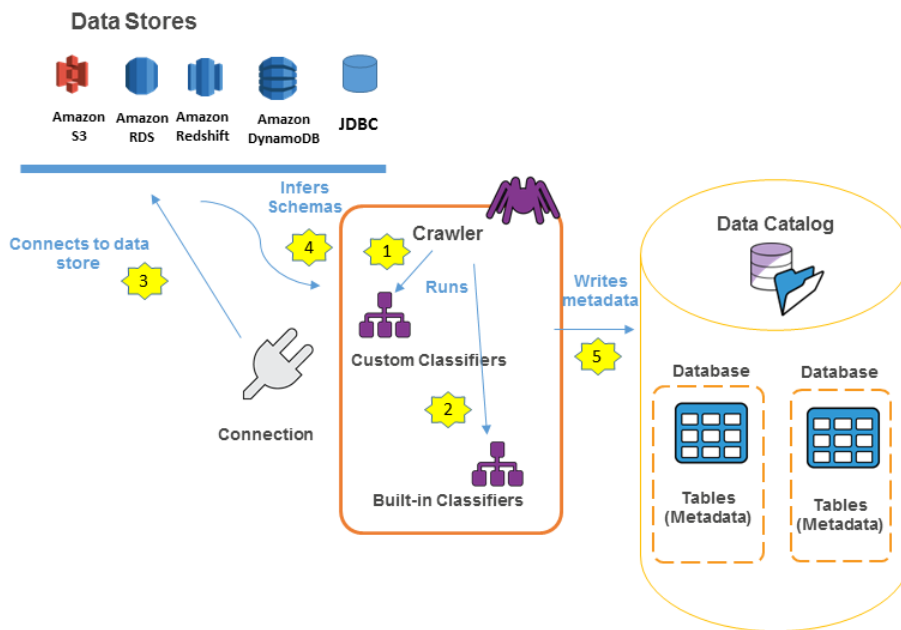
Като един от водещите доставчици на технологични услуги в сферата на Големите данни, AWS (Amazon Web Services) има свое решение за каталог на данни и автоматично каталогизиране на данни. AWS Data Catalog е част ETL инструмента Glue, но може да се използва и самостоятелно като отделна услуга.

Според [6] каталогът с данни на AWS Glue е напълно управлявано хранилище за постоянни метаданни, което позволява съхранението, аотирането и споделянето на метаданни. Той предоставя унифицирано хранилище на метаданни за различни източници на данни и формати на данни, като може да се интегрира с Amazon EMR, както и с Amazon RDS, Amazon Redshift и Redshift Spectrum, Amazon Athena, AWS Lake Formation и всяко приложение, съвместимо с metastore Apache Hive.

Типичното приложение на каталогът за данни на AWS е като част от процесите по извличане, трансформиране и зареждане на данни (ETL). Каталогът за данни на AWS Glue представлява индекс на местоположението, схемата и показателите за време на изпълнение относно данните.

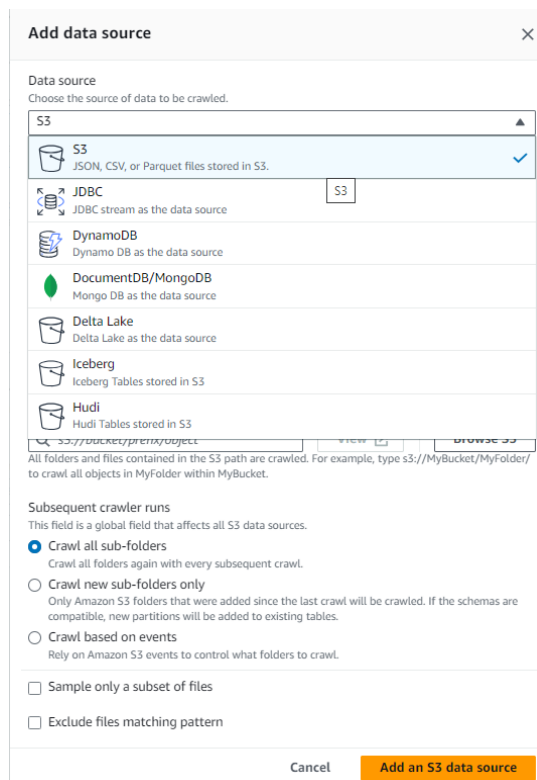
Информацията в каталога за данни се съхранява като таблици с метаданни, където всяка таблица посочва едно хранилище на данни. Обикновено автоматичното управление на тези таблици с метаданни става чрез използването на AWS Crawler.

Следната диаграма представя типичен работен процес, който показва как AWS Glue Crawler си взаимодейства с хетерогенните източници на данни за да ги каталогизира в каталога за данни.



Фигура 2: AWS Glue Crawler и Data Catalog

Една от основните функции на Glue Crawler е да намира данни, без значение дали те са във формати като JSON, CSV, Parquet, или дори бази данни като Amazon RDS или Amazon Redshift. Този автоматичен процес позволява да се сведе до минимум ръчното каталогизиране на данните и ускорява времето за подготовката на данните за последваща обработка. Освен това Glue Crawler предоставя възможности за достъп до разнообразни източници на данни(Фиг. 3).



Фигура 3: AWS Glue Crawler източници на данни

Основната услуга за формиране на Езеро от данни в AWS е S3. AWS Crawler-a е специално разработен да обхожда автоматично данни в S3 и да записва в Data Catalog-a наличните промени или нови метаданни за данните.

AWS Glue Crawler генерира богати метаданни за откритите данни, включително информация за структурата на таблиците и полетата, типовете данни и ключовите връзки. AWS Glue Crawler предоставя автоматизация на процеса по обновление на метаданни, като сканира периодично източниците на данни и актуализира информацията за тях. Също така, услугата е скалируема и може да се адаптира към растящия обем на данни и източници.

Общият работен процес за това как AWS Crawler-a попълва каталога за данни е следния:

1. Изпълняват се така наречените класификатори, които проверяват дали данните могат да бъдат прочетени в определен формат. Важно е да се отбележи, че има възможност за създаване на персонализирани класификатори.
2. Crawler-a се свързва с източника на данни, като е важно да се отбележи, че в зависимост от типа на източника на данни може да се изискват определени свойства за достъп до източника.
3. Изведената схема се създава за откритите данни.
4. Crawler-a записва метаданните в каталога за данни под формата на таблица. Дефиницията на таблица съдържа метаданни за данните. Таблицата се записва в база данни, която е контейнер от таблици в каталога за данни. Атрибутите на таблица включват класификация, която е етикет, създаден от класификатора, който е извел схемата на таблицата.

4. Заключение

В заключение може да кажем, че каталогът за данни е ключов инструмент за ефективното управление, изследване и използване на големи и сложни набори от данни. Той играе съществена роля в улесняването на откриването на данни, насърчаването на сътрудничество, осигуряването на управление на данните и подобряването на разбирането на данните в организацията. В епохата на Големите данни ефективната стратегия за каталогизиране на данни се явява крайъгълен камък за постигане на успех в осъществяването на решения, базирани на Големи данни. AWS Glue Catalog и AWS Glue Crawler успешно се справят с каталогизирането на различни източници и то автоматично

References

1. E. Zaidi, G. De Simoni, R. Edjlali, and A. D. Duncan, "Data Catalogs Are the New Black in Data Management and Analytics," Gartner, no. December, pp. 1–16, 2017
2. L. Ehrlinger, J. Schrott, M. Melichar, N. Kirchmayr, and W. WöB, "Data Catalogs: A Systematic Literature Review and Guidelines to Implementation," in Database and Expert Systems Applications - DEXA 2021 Workshops, 2021, vol. 2, pp. 148–158, doi: 10.1007/978-3-030-87101-7_15.
3. J. Riley, Understanding Metadata. Baltimore: National Information Standards Organization (NISO), 2017.
4. J. Nogueras-Iso, J. Lacasta, M. A. Urena-Camara, and F. J. Ariza-Lopez, "Quality of Metadata in Open Data Portals," IEEE Access, vol. 9, pp. 60364–60382, 2021, doi: 10.1109/ACCESS.2021.3073455.
5. J. Klímek, "Reflections on: DCAT-AP representation of Czech national open data catalog and its impact," CEUR Workshop Proc., vol. 2576, no. 19, pp. 1–9, 2019.
6. AWS Best Practices for Building a Data Lake on AWS for Games <https://docs.aws.amazon.com/whitepapers/latest/best-practices-building-data-lake-for-games/data-cataloging.html>

7. Delchev, D., Lazarova, V., Big Data Analysis Architecture, Economic Alternatives, 2021, Issue 2, pp. 315-328
8. Boyanov L., The Digital World - The Change, The global digital transformation - enriching or impoverishing humanity, ISBN 978-619-239-637-4, Avangard Prima Publ., Sofia 2021, 188 p.
9. M. Tsaneva, "A Practical Approach For Integrating Heterogeneous Systems," Business management, no. 2, p. 11, 2019.
10. V. Mihova, Common Architecture Design of a Business Information System for Performance Management of the Business Applications, in 3rd International conference on application of information and communication technology and statistics in economy and education ICAICTSEE–2013, Sofia, Bulgaria, 2013.
11. P. Milev, Technological Issues of Storing Dynamic Data in a Relational Database on Research Projects, Trakia Journal of Sciences, vol. 13, pp. 22-25, 2015
12. E. Karkalikova, A. Murdjeva, Organization of Data in Data Lake – Real-Life Practice, 11th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education ICAICTSEE– 2021, Sofia, Bulgaria.
13. P. Milev, Approach for Analysis and Comparison of Search Query Results in Web Publications, 11th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education ICAICTSEE– 2021, Sofia, Bulgaria
14. Marzovanova M., Building Multi-Touch User Interface, 4TH International Conference on Application of Information and Communication Technology and Statistics in Economy And Education (ICAICTSEE-2014), 2014, (icaictsee.unwe.bg/past-conferences/ICAICTSEE-2014.pdf).
15. Mihova V., Murdjeva A. Metadata for generating a specific data warehouse. International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2012), Sofia, Bulgaria, 2012. (icaictsee.unwe.bg/past-conferences/ICAICTSEE-2012.pdf)

ЕЗЕРА ОТ ДАННИ(DATA LAKES) КАТО ХРАНИЛИЩЕ ЗА ГОЛЕМИ ДАННИ. АНАЛИЗ НА AWS ПОДХОДА

Гено Стефанов

главен асистент, доктор, катедра Информационни технологии и комуникации, УНСС,
e-mail: genostefanov@unwe.bg

Резюме

Езерата от данни(Data Lakes) станаха неизменна част от Големите данни(Big Data) в ролята си на хранилище за хетерогенни и големи по обем данни. Езерата от данни предлагат всички необходими инструменти за съхранение на Големи данни. Тези инструменти могат да бъдат разнообразни в зависимост от софтуерния доставчик. Amazon Web Services(AWS) като един от големите доставчици на услуги за Големи данни е подходящ избор за анализиране на ползите и предизвикателствата пред съхранението на Големи данни в Езера от данни. Освен това ще се разгледа в детайли подхода на AWS.

Ключови думи: Големи Данни, Езера от данни

DATA LAKES AS DATA STORE FOR BIG DATA. ANALYZING AWS APPROACH.

Geno Stefanov

Abstract

Data Lakes have become an integral part of Big Data in their role as storage for heterogeneous and large-volume data. Data lakes offer all the necessary tools to store Big Data. These tools can vary depending on the software vendor. Amazon Web Services (AWS) as one of the major Big Data service providers is a suitable choice to analyze the benefits and challenges of storing Big Data in Data Lakes. In addition, the AWS approach will be discussed in detail.

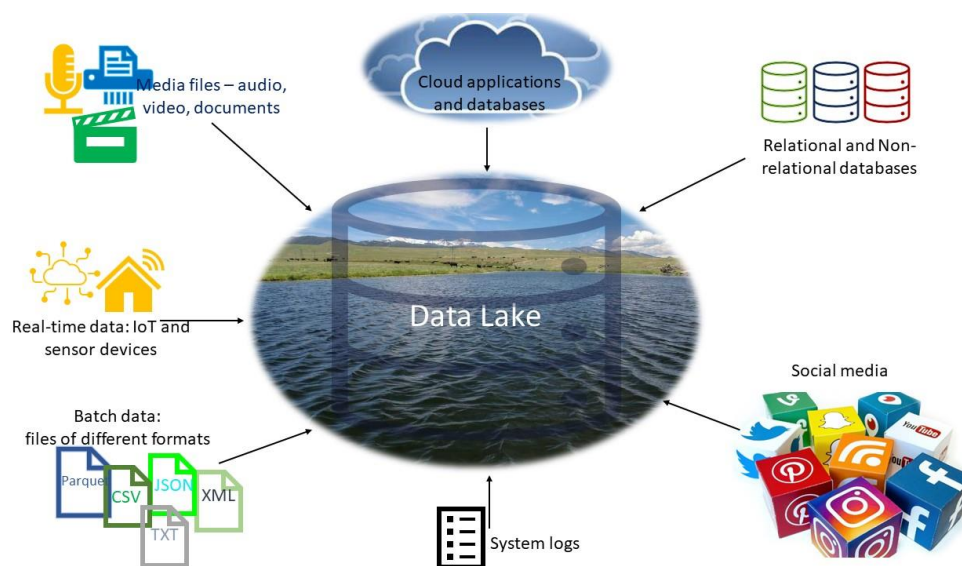
Key words: Big Data, Data Lakes

1. Въведение

В съвременния свят обемът на данни нараства експоненциално, предизвиквайки огромни предизвикателства при управлението, съхранението и анализа на тази информация. Този феномен е известен като "Големи данни" (Big Data). За да се справят с тези предизвикателства, Езерата от данни (Data Lakes) се появяват като нова и иновативна архитектура за съхранение на големи обеми, разнородни и неструктурирани данни. В този доклад ще се разгледа концепцията на езерата от данни и тяхната роля като хранилище за Големи данни.

2. Езера от данни като хранилища за Големи данни

Езеро от данни може да се дефинира като централизирано хранилище, което съхранява големи обеми от сурови и необработени данни в техния естествен формат, без необходимостта от предварително дефинирана схема. Това архитектурно решение позволява съхранението на данни от различни източници, включително структурирани, полуструктурирани и неструктурирани данни, по ефективен и мащабируем начин.



Фигура 1: Езеро от данни

Идеята за Езера от данни за първи път е иницирана от изпълнителния директор на Pentaho Джейм Диксън [3]. Ако складът от данни или базата данни ги представим като бутилка вода, чиста и готова за консумация, тогава Езерата от данни е цяло езеро от данни, което е почистено и готово за употреба. [4] добавя по-задълбочено определение за Езера от данни като „Езерата от данни съхраняват различна информация, като същевременно игнорира почти всичко“. Някои автори смятат, че е необходима нова архитектура на данните [5] в ерата на Големите данни, тъй като тази интензивна изчислителна ера изисква нови идеи и техники за съхраняване и обработка на обемни, разнообразни, променящи се и развиващи се данни.

Езерата от данни са проектирани да обработва Големи данни, което се отнася до големи и сложни масиви от данни, които надхвърлят възможностите на традиционните системи за съхранение и обработка на данни. Езерата от данни предлагат гъвкав и рентабилен подход за съхранение и анализ на данни, позволявайки съхраняването на данни от различни източници.

За разлика от традиционните хранилища за данни, при които структура на данните трябва да е известна [13], преди да се съхранят, езерата от данни поддържат данните в тяхната сурова форма, запазвайки оригиналната детайлност. Тази характеристика позволява на Езерата от данни да съхраняват данни от всякакъв тип и обем, което ги прави подходящи за обработка на непрекъснато нарастващите потоци от данни, генерирани от модерни технологии, като социални медии, IoT устройства, сензори и бизнес приложения [10].

Езерата от данни предлагат всички необходими инструменти за съхранение на Големи данни, за това може да се счита, че те могат да играят съществена роля като хранилище за Големи данни. Тук ще разгледаме предимствата и недостатъците на това Езерата от данни да се използват като хранилище за Големи данни.

Езерата от данни предлагат няколко предимства като хранилище за големи данни:

- **Мащабируемост:** Езерата от данни могат лесно да нарастват и се справят с нарастващия обем на данни, осигурявайки дългосрочно съхранение без компрометиране на производителността.

- **Гъвкавост:** Запазвайки суровия формат на данните, езерата от данни предоставят гъвкавост за изследване и анализ на данните, позволявайки на специалистите по данни да извличат информация с различни инструменти.
- **Икономичност:** Чрез използването на стандартно оборудване и облачни решения, езерата от данни могат да бъдат по-икономични от традиционните решения за съхранение на данни.
- **Възможност за обработка в реално време:** Езерата от данни позволяват обработка на данни в реално време или близко до реално време, което дава възможност на организациите да взимат бързи решения основани на данни.

Предизвикателства при използването на Езерата от данни. Въпреки своите предимства, езерата от данни представят и някои предизвикателства:

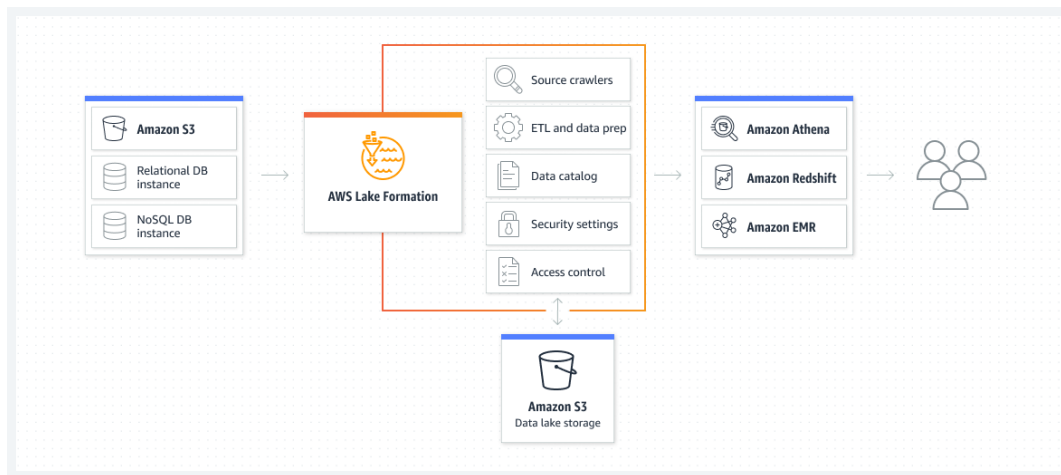
- **Управление на данните:** Управлението на качеството, сигурността и поверителността на данните в Езерата от данни може да бъде сложно, тъй като липсата на предварително дефинирана схема може да доведе до неизползваемост на данните и грешно тълкуване.
- **Управление на метаданните:** Управлението на метаданните е от съществено значение в езерата от данни, тъй като помага на потребителите да разберат структурата и контекста на съхранените данни, позволявайки ефективно откриване и анализ на данните.
- **Разделение на данните:** Ако не се управляват правилно, Езерата от данни могат да се превърнат в “блато” от данни, което ще затрудни или да направи невъзможно управлението на данните.

Най-добри практики при използването на Езерата от данни: За максимизиране на предимствата на Езерата от данни като хранилище за Големи данни, трябва да се следват следните най-добри практики:

- **Управление на метаданните:** Въвеждането на ефективна стратегия за управление на метаданните позволява каталогизиране на данните, проследяване на техния произход и откриване
- **Интеграция с инструменти за обработка на данни:** Езера от данни могат да се интегрират с инструменти за обработка на данни като Apache Spark или Hadoop, което ще подобри анализа на данните и тяхната трансформация.
- **Управление на сигурността –** Трябва да има въведен контрол на достъпа, криптиране и техники за маскиране на данните, за да са защитени чувствителните данни от неоторизиран достъп.

3.AWS решението за Езера от данни

Подходът на AWS за разработване на решения за Езера от данни се нарича AWS Lake Formation. През 2019 г., AWS обявява нова услуга - AWS Lake Formation. AWS Lake Formation е услуга на AWS, която улеснява процеса на създаване, осигуряване и управление на Езера от данни. Тази услуга предоставя комплексен набор от инструменти и функционалности, които позволяват на организациите да интегрират, съхраняват, каталогизират и анализират големи обеми данни от различни източници[15].



Фигура 2: AWS Lake Formation архитектура и функционалности

Основни характеристики на AWS Lake Formation включват:

- **Хетерогенни източници на данни:** AWS Lake Formation позволява обработката на данни от разнообразни източници, включително бази данни, хранилища за данни, стрийминг данни, лог файлове и други, като се поддържа както пакетно, така и в реално време на приемане.
- **Съхранение на данни:** С AWS Lake Formation организациите могат да съхраняват големи обеми от сурови, структурирани, полуструктурирани и неструктурирани данни в Amazon S3 (Simple Storage Service) - услугата за съхранение на AWS, която предлага мащабируемост и икономичност.
- **Каталог на данни:** AWS Lake Formation включва каталог на данни, който автоматично обхожда и каталогизира метаданни от различните източници на данни, съхранени в Езерото от данни. Този каталог позволява на потребителите да откриват, търсят и използват данни по-лесно. Автоматичното каталогизиране на данните става възможно с услугата Crawlers.
- **Сигурност и управление на данните:** AWS Lake Formation осигурява надеждни механизми за сигурност, включително подробен контрол на достъпа, криптиране и интеграция със системата за управление на идентичност и достъп на AWS (AWS IAM), за контрол на достъпа до ресурсите в Езерото от данни. Също така се поддържа и класификация на данните, което гарантира, че чувствителните данни са подходящо идентифицирани и защитени.
- **Трансформация на данните:** С помощта на AWS Lake Formation потребителите могат да дефинират задачи за трансформация на данните, използвайки AWS Glue - друга услуга на AWS, която позволява подготовка и трансформация на данните за анализ.
- **Интеграция с аналитични услуги:** AWS Lake Formation се интегрира безпроблемно с различни аналитични услуги и услуги за машинно обучение в екосистемата на AWS, като например Amazon Athena, Amazon Redshift, Amazon QuickSight и Amazon SageMaker, което позволява на потребителите да извличат ценни аналитични данни от данните, съхранени в Езерото от данни.

AWS Lake Formation предлага редица предимства, включително лесно създаване и управление на Езера от данни, мащабируемост, икономичност и интеграция с широка гама от аналитични и услуги за машинно обучение на AWS. Улеснява процеса на обработка на Големите обеми данни и извличане на ценни знания за организациите, които се нуждаят от вземане на бързи и информирани решения.

4. Заключение

В заключение може да кажем, че Езерата от данни е мощно решение за управление на данни, което позволява на организациите ефективно да обработват Големи данни, което им позволява да получат ценна информация и да вземат ефективно решения, базирани на данни

References

1. Brian Stein, Alan Morrison, "The enterprise data lake: Better integration and deeper analytics, Technology Forecast: Rethinking integration", Issue 1, 2014,
2. Pwint Phyu Khine, Zhao Shun Wang, Data lake: a new ideology in big data era, ITM Web of Conferences, 2018, DOI:10.1051/itmconf/20181703025.
3. James Dixon, Pentaho, Hadoop and Data Lakes. <https://jamesdixon.wordpress.com/2010/10/14/pentaho-hadoop-and-data-lakes/>
4. Timothy King "The Emergence of Data Lake: Pros and Cons", March 3, 2016, <https://solutionsreview.com/data-integration/the-emergence-of-data-lake-pros-and-cons/>
5. Delchev, D., Lazarova, V., Big Data Analysis Architecture, Economic Alternatives, 2021, Issue 2, pp. 315-328
6. Boyanov L., The Digital World - The Change, The global digital transformation - enriching or impoverishing humanity, ISBN 978-619-239-637-4, Avangard Prima Publ., Sofia 2021, 188 p.
7. M. Tsaneva, "A Practical Approach For Integrating Heterogeneous Systems," Business management, no. 2, p. 11, 2019.
8. V. Mihova, Common Architecture Design of a Business Information System for Performance Management of the Business Applications, in 3rd International conference on application of information and communication technology and statistics in economy and education ICAICTSEE–2013, Sofia, Bulgaria, 2013.
9. P. Milev, Technological Issues of Storing Dynamic Data in a Relational Database on Research Projects, Trakia Journal of Sciences, vol. 13, pp. 22-25, 2015
10. E. Karkalikova, A. Murdjeva, Organization of Data in Data Lake – Real-Life Practice, 11th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education ICAICTSEE– 2021, Sofia, Bulgaria.
11. P. Milev, Approach for Analysis and Comparison of Search Query Results in Web Publications, 11th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education ICAICTSEE– 2021, Sofia, Bulgaria
12. Marzovanova M., Building Multi-Touch User Interface, 4TH International Conference on Application of Information and Communication Technology and Statistics in Economy And Education (ICAICTSEE-2014), 2014, (icaictsee.unwe.bg/past-conferences/ICAICTSEE-2014.pdf).
13. Mihova V., Murdjeva A. Metadata for generating a specific data warehouse. International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE-2012), Sofia, Bulgaria, 2012. (icaictsee.unwe.bg/past-conferences/ICAICTSEE-2012.pdf)
14. Dan Wood, Big data requires a big new architecture, Forbes, <https://www.forbes.com/sites/ciocentral/2011/07/21/big-data-requires-a-big-new-architecture/#66609cb61157>
15. AWS Lake Formation. <https://aws.amazon.com/lake-formation/>

SIGNIFICANCE AND COMPARABILITY OF UNSTRUCTURED AND SEMI-STRUCTURED DATA IN THE MODERN WEB

Yavor Tabov

assistant, PhD. Department of Information Technologies and Communications, University of National and World Economy, e-mail: jtabov@unwe.bg

Abstract

This document examines the characteristics and features of unstructured and semi-structured data in the modern web. Scales and trends for the growth of these types of data are presented. Popular approaches for analyzing unstructured and semi-structured data are discussed, with a comparison made regarding the different and common approaches. Finally, conclusions are drawn based on the prepared material.

Key words: unstructured data, semi-structured data, Internet, web scraping, machine learning

JEL: C88, L86.

1. Introduction of unstructured and semi-structured data and their place in the modern web

Technological progress has led to a significant increase in data volume in recent years. This growth in big data has brought about changes in the capacity and capabilities of computing systems due to the sheer volume, diversity (including structured, unstructured, and semi-structured data), and speed at which data is generated. A vast volume of data, whether it's unorganized or partially structured, is being produced continually, whether it's on a daily, minute-by-minute, or even second-by-second basis, stemming from a multitude of sources in our everyday lives. The count of internet users is experiencing rapid and substantial growth with each passing day [3]. Extracting valuable information from various data types has become a challenging endeavor, particularly given the vast volume and intricacy of unstructured data [1].

Unstructured data originates from both machine-generated and human-generated sources, and it is generally categorized into two main groups: non-textual and textual. Non-textual unstructured data includes multimedia content such as static images, videos, and audio files. Examples of textual unstructured data encompass items like email messages, collaborative software and instant messages, memos, word processing documents, and PowerPoint presentations. Various standards for handling unstructured data include open XML, SMTP, SMS, CSV [8].

Unstructured data refers to data in an unfamiliar format, which can be challenging and time-consuming to handle. It encompasses various types, including text files, videos, images, and more. Nowadays, organizations possess abundant data resources, but extracting value from unstructured data remains a significant challenge [3]. Early data extraction systems primarily focused on text, which remains the most extensively explored form of data by both commercial systems and the scientific community to this day [11].

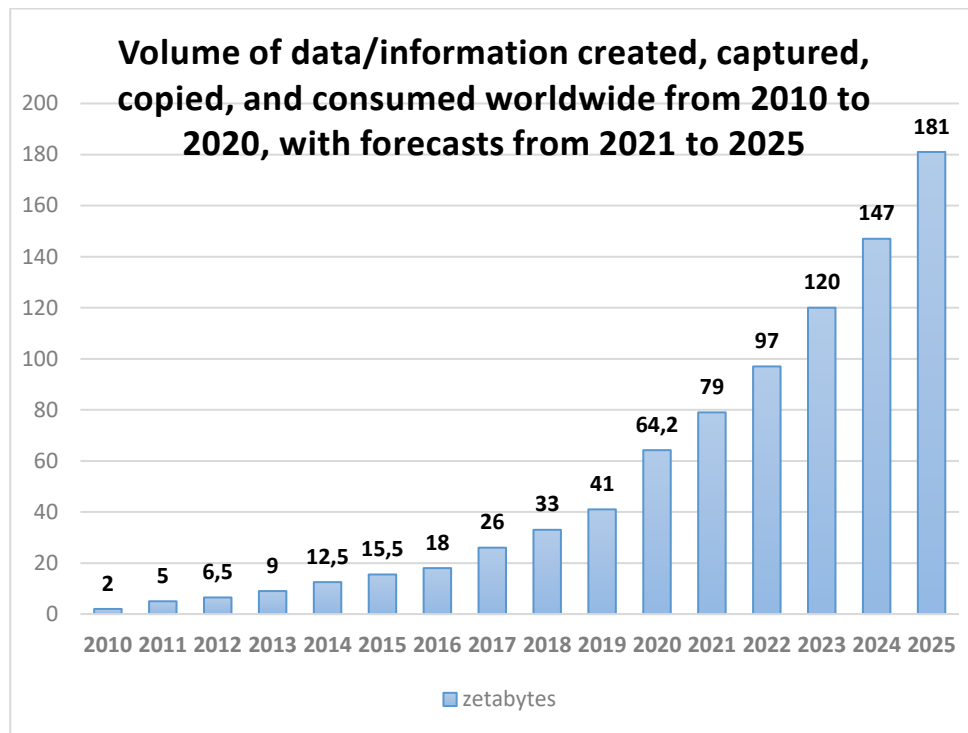
The significance of unstructured data is gaining greater recognition among both scholars and the public. Machine learning paradigms offer a computational advantage through their capacity to derive meaningful patterns from unstructured data for addressing real-world problems. However, it's widely acknowledged that one limitation of these paradigms is the often-inscrutable nature of the models they employ. To casual observers, the data modeling process is concealed by machine learning technologies, which sometimes leads to the perception of AI as a form of magic [7].

Semi-structured data, while considered a subset of structured data, lacks a strict data model or a formal structure. It doesn't require a predefined schema and often includes optional tags or markers

that help delineate semantic elements and establish hierarchies within the data. This type of data has become more prevalent, especially since the internet hosts various data types beyond just full-text documents and databases, and various applications require a means of information exchange. Managing semi-structured data typically involves using markup languages like XML or JavaScript Object Notation (JSON). [8].

Semi-structured data consists of both structured data and unstructured data [3]. It can be seen as a variant of structured data, sharing some characteristics with structured data while not adhering to traditional database models. An instance of this is represented by Common Separated Value (CSV) files [7]. Semi-structured data encompasses formats like emails, XML, and JSON. When dealing with semi-structured data, traditional relational databases are not suitable. Instead, this type of data is represented using concepts like edges, labels, and tree structures [17]. Due to the absence of a predetermined schema definition, semi-structured data offers significant scalability and flexibility [6].

Every day, countless structured, semi-structured, and unstructured documents are generated worldwide [22]. Digital data generated by various digital platforms and devices are growing at an astounding rate globally [15]. To illustrate this increase, one can point to the statistics from the German platform statista.com, specialized in data collection and visualization, which shows that the amount of produced information has grown ninefold from 2013 to 2023. It is anticipated that by 2025, this daily volume will surge to 181 zettabytes [20]. This data growth trend is illustrated in Fig. 1.



Source: statista.com

Figure 1. Data growth trend

2. Approaches for analyzing unstructured and semi-structured data.

Unstructured and semi-structured data pose challenges for professionals in the field of information technology because they often require more time than initially anticipated for structuring and subsequent preparation for analytical processing. These types of data on the Web come in very large volumes. This implies the application of specific approaches when working with them. This document will draw a comparison between unstructured and semi-structured data, examining the most popular approaches for processing these types of data.

In the context of processing unstructured data, various approaches have been formulated in literature. These approaches include, namely:

- **Speech Recognition.** Speech recognition is the procedure of converting spoken language into a sequence of words using computer programs or algorithms. This field is a fascinating aspect of signal processing [5];

- **Machine Translation.** Machine translation, often abbreviated as MT (not to be confused with computer-aided translation, machine-aided human translation, or interactive translation), is the automated translation of text from one natural language, like English, to another, such as Ibo, by computer software [16].

Analyzing semi-structured data demands tailored strategies due to their adaptable and variable nature. Popular approaches for analysis and processing of this type of data are:

- **XML parsing.** XML parsing involves the action of reading an XML document and offering a user application interface to access that document. An XML parser is a software tool that carries out this function. Furthermore, most XML parsers assess the document's proper structure, and many can even validate it against a Document Type Definition (DTD) or an XML schema. By utilizing the parsing interface, the user application can direct its attention to the core application logic without becoming entangled in the intricate specifics of XML [13].

- **NoSQL Databases.** Relational databases are databases that can be scaled vertically. To manage extensive quantities of semi-structured and unstructured data, NoSQL databases are used. NoSQL databases adhere to the BASE principle, which stands for "Basically, Available, Soft state, and Eventually consistent." Relational databases excel in data integrity, security, and reliable transactions. In contrast, NoSQL databases are suitable for handling large volumes of data in various formats. NoSQL databases manage big data with lower associated costs and minimal overhead. The scalability of NoSQL databases is achieved horizontally by simply adding new servers in a clustered environment. Commodity hardware is utilized for storing substantial data within the cluster [10].

NoSQL databases can be categorized based on the data model used. There are databases of the Document store type. They utilize formats like XML, JSON, BSON, and more [12]. The data within them is semi-structured and comprises pairs of attribute name-value. Retrieving data involves searching for both key values and attribute values. They are well-suited for storing text, XML documents, and other semi-structured data. Prominent examples of this category include MongoDB, Amazon DynamoDB, Couchbase, and CouchDB [18].

- **Web parsing.** Web parsing involves the automated retrieval of extensive publicly accessible data from websites and web-based information systems. It streamlines the process of collecting data and transforms the acquired information into diverse formats like HTML, CSV, Excel, JSON, and text. This procedure typically encompasses three main phases: parsing HTML pages, extracting data, and storing it [2]. However, some authors suggest expanding the phases, emphasizing data analysis within the web scrapping process [14]. In the realm of data analysis, proficient programming is crucial. Consequently, many businesses seek to employ developers well-versed in web scrapping techniques [2].

There are approaches that are applied to both types of data:

- **Natural Language Processing.** Natural Language Processing involves various techniques based on linguistic features. It analyzes a given text syntactically using formal grammar rules, and the resulting data is semantically interpreted to extract information from it. Natural Language Processing can be applied deeply, where each part of each sentence is analyzed, and attempts are made to interpret it, or it can be applied superficially by conducting limited semantic analysis on only some passage**s or phrases in sentences. Natural Language Processing allows for techniques to distinguish specific words or word parts. This approach is used by a significant portion of modern information systems designed for automatic text translation [9]. Natural language processing is used to transform semi-structured and unstructured data into structured data;

- Knowledge Discovery in Text. Text mining, also known as Knowledge Discovery from Text (KDT), involves extracting information and identifying patterns from unstructured data. It's particularly useful for analyzing semi-structured or unstructured datasets, such as emails, full-text documents, and HTML files. Text mining, also referred to as Knowledge Discovery from Text (KDT), entails uncovering valuable patterns within extensive text databases to gain knowledge. Text mining applies not only data mining's analytical techniques but also incorporates methods from natural language and information retrieval [21]. The approach of Knowledge Discovery in Text primarily utilizes analytical techniques related to information extraction and machine self-learning. The goal of this approach is to extract models for processing a large number of textual documents. This approach is typically used by contemporary search engines. The Knowledge Discovery in Text approach involves various activities, namely: automatic text classification according to a defined set of categories; grouping texts based on common features; automatic summarization; extracting topics from texts and analyzing thematic trends in text flows [9];

According to the literature in the field of information analysis, Natural Language Processing and Knowledge Discovery in Text are the two main approaches in the context of representing textual data in a structured form used in practice [4];

- Data visualization. Data visualization is a method used to illustrate intricate data through graphical means, making it easier to comprehend. When dealing with structured data, traditional graphical representations are straightforward. However, for unstructured or semi-structured data, achieving effective real-time visualization can be challenging due to their diverse nature [19].

The approaches for analyzing semi-structured and unstructured data are illustrated in Fig. 2.

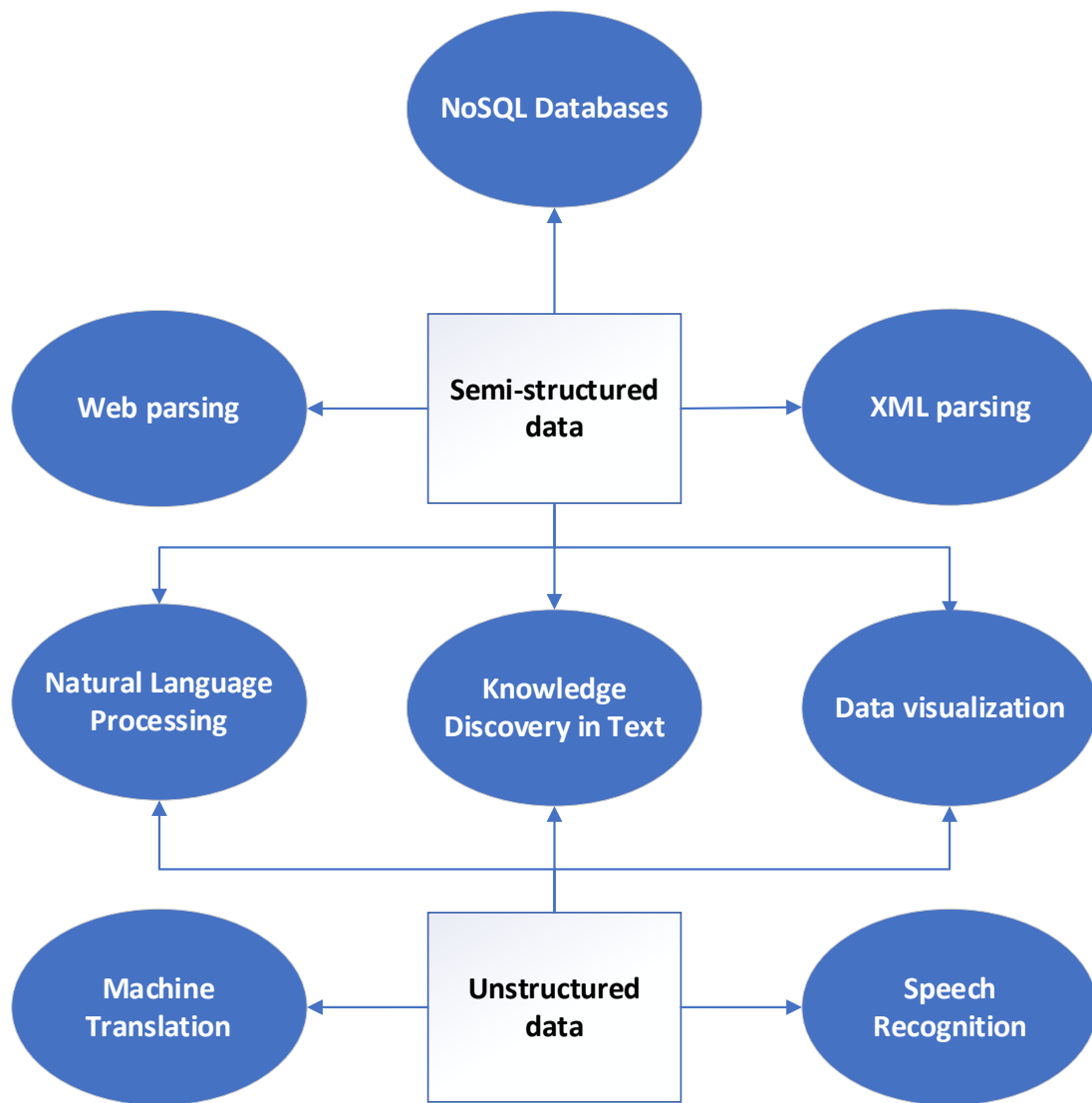


Figure 2. Approaches for analyzing semi-structured and unstructured data

Conclusion

In conclusion from the presented research of the significance and comparability of unstructured and semi-structured data in the modern web, we can summarize the following notes:

- Unstructured and semi-structured data are present in the modern web on a large scale and are growing rapidly.
- Unstructured data is divided into two groups - textual, consisting of text, and non-textual, representing images, video, and audio files.
- Semi-structured data consists of structured and unstructured data. Popular formats for semi-structured data are XML and JSON.
- Popular approaches for analyzing and processing unstructured and semi-structured data include Natural Language Processing, Knowledge Discovery in Text, and Data Visualization. Processing approaches characteristic of unstructured data include Machine Translation and Speech Recognition. Processing approaches characteristic of semi-structured data include Web Parsing, XML Parsing, and NoSQL Databases.

References

1. Adnan, K., Akbar, R. (2019). An analytical study of information extraction from unstructured and multidimensional big data. *Journal of Big Data* 6, Article number 91, 2019.
2. Britvin, A., Alrawashdeh, J., Tkachuck, R. (2022). Client-server system for parsing data from web pages. *Advances in cyber-physical systems*, Vol. 7, No. 1, 2022, ISSN: 2524-0382.
3. Chowdhury, A. (2020). *Computer & Information Systems into Next Frontier: The Emergence*. New Delhi Publishers, 2020.
4. Das, T., Kumar, P. (2013). BIG Data Analytics: A Framework for Unstructured Data Analysis. *International Journal of Engineering and Technology (IJET)*, Vol 5 No 1 Feb-Mar 2013, India, ISSN : 0975-4024.
5. Gaikwad, S, Gawali, B., Yannawar, P. (2010). A Review on Speech Recognition Technique. *International Journal of Computer Applications*, Volume 10, No. 3, November 2010, ISSN 0975-8887.
6. Halcu, I., Grigoriu, O., Neculoiu, G., Sandulescu, V., Marinescu, M., Marinescu, V. (2013). Converting unstructured and semi-structured data into knowledge. *Roedunet International Conference*, 2013, ISSN: 2068-1038.
7. Jiang, S., Nocera, A., Tatar, C., Yoder, M., Chao, J., Wiedemann, K., Finzer, W., Rosé, C. (2022). An empirical analysis of high school students' practices of modelling with unstructured data. *British Journal of Educational Technology*, Volume 53, issue 5, 2022, ISSN 1467-8535.
8. Kanimozhi, K. and Venkatesan, M. (2015). Unstructured Data Analysis-A Survey. *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 4, Issue 3, 2015.
9. Kao, A. and Poteet, S. (2005). Text Mining and Natural Language Processing – Introduction for the Special Issue. *SIGKDD Explorations*, Volume 7, Issue 1, 2005.
10. Khan, W., Ahmed, E., Shahzad, W. (2017). Predictive Performance Comparison Analysis of Relational & NoSQL Graph Databases. *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 5, 2017, ISSN 2156-5570.
11. Khder, M. (2021). Web Scraping or Web Crawling : State of Art, Techniques, Approaches and Application. *International journal of advances in soft computing and its applications*, Vol. 13, No. 3, 2021, ISSN 2074-8523.
12. Kovacheva, M. (2021). Classification, Comparison and Criteria for Choosing NoSQL Databases. *ICAICTSEE-2021*.
13. Li, C. (2009). XML Parsing, SAX/DOM. *Encyclopedia of Database Systems*, 2009, ISBN 978-0-387-39940-9
14. Milev, P. (2017). Conceptual Approach for Development of Web Scraping Application for Tracking Information. *Economic Alternatives*, (3), 475-485.
15. Nti, I., Quarcoo, J., Aning, J., Fosu, G. (2022). A Mini-Review of Machine Learning in Big Data Analytics: Applications, Challenges, and Prospects. *Big Data Mining and Analytics*, Volume 5, Number 2, June 2022, ISSN 2096-0654.
16. Okpor, M. (2014). Machine Translation Approaches: Issues and Challenges. *International Journal of Computer Science Issues*, Vol. 11, Issue 5, No 2, September 2014, ISSN 1694-0784.
17. Praveen, S., Chandra, U. (2017). Influence of Structured, Semi-Structured, Unstructured data on various data models. *International Journal of Scientific & Engineering Research* Volume 8, Issue 12, December 2017, ISSN 2229-5518.
18. Radoev, M. (2017). A Comparison between Characteristics of NoSQL Databases and Traditional Databases. *Computer Science and Information Technology*, November 2017, ISSN 2331-6071.

19. Sandhu, A. (2022). Big Data with Cloud Computing: Discussions and Challenges. *Big Data Mining and Analytics*, Volume 5, Number 1, March 2022, ISSN 2096-0654.
20. Statista Research Department. (2021). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025 [Infographic]. Statista. <https://www.statista.com/statistics/871513/worldwide-data-created/>
21. Yehia, A., Ibrahim, L., Abulkhair, M. (2016). Text Mining and Knowledge Discovery from Big Data: Challenges and Promise. *International Journal of Computer Science Issues*, Volume 13, Issue 3, May 2016, ISSN 1694-0784.
22. Zaman, G., Mahdin, H., Hussain, K., Rahman, j A. (2020). Information extraction from semi and unstructured data sources: a systematic literature review. *ICIC Express Letters*, Volume 14, Number 6, June 2020, ISSN 1881-803X.

DESIGN AND IMPLEMENTATION OF PROJECT MANAGEMENT APPLICATION IN REACT.JS

Daniel Nalbantov

Student in Information Technologies and Communications, University of National and World Economy; dnalbantov_20114013@unwe.bg

Victor Del Aguila Angulo, Flavio Gorini, Alessio Mancini, Nicola Rispoli

Erasmus Students in Information Technologies and Communications, University of National and World Economy

Abstract

In this paper, we're going to explore the field of project management, which is an organized method of effectively accomplishing our objectives that goes much beyond a list of chores. We think that efficient goal-achieving and positive impact-making are more important aspects of good project management than just finishing duties. Our application aims to enable teams and organizations to operate more productively; it's not simply about software. Our extensive functionality and user-friendly interface make project administration less complicated. We make it unnecessary to retain manual records, lower the possibility of mistakes, and allow for instantaneous decision-making.

Key words: project management, React.JS, design.

JEL: O3, Z0

1. Introduction

A wide range of duties are included in project management, such as task definition, task organization, and close task execution monitoring. Planning, resource management, budgetary control, time management, quality assurance, and upkeep of thorough project documentation are all part of it. It's about making sure we accomplish our objectives quickly and effectively.

Project management requires more than just spreadsheets and manual procedures in the fast-paced world of today. It calls for an advanced information system that can give thorough project and task descriptions.

Why does this matter? All projects, no matter how big or little, require a variety of professionals, resources, and moving components. Handling all of these details by hand can be difficult, error-prone, and inefficient most of the time. This is the role that our project plays.

Our React.js application aims to empower teams and organizations to operate more productively and efficiently—it's not just about code and design. By offering necessary features including project and task registration, contractor administration, and thorough reporting, we want to simplify project management.

Our goal in attending to these demands is to help initiatives and, eventually, the organizations that depend on them succeed. Our initiative is a response to actual project management difficulties, not merely a piece of software.

2. Main features and functionality

Our React.js application is made to provide you an extensive feature set that will revolutionize project management. Let's examine a few of the main features:

Registering Projects and Tasks:

The process of registering projects and the duties related to them is made easier by our system.

You may simply add project details, such as a unique project code, name, description, client information, start and finish dates, and project status, whether you're managing a single project or a portfolio of them. You can be sure that all your active tasks are clearly visible to you thanks to this feature.

Contractor and Expert Management:

The secret to a successful endeavor is teamwork. You may register contractors and assign them to particular project tasks using our application. In addition, specialists that are either inside or external to your company may be registered and managed. By ensuring that the appropriate individuals are assigned to the appropriate duties, this promotes efficient project execution.

Reporting and Insights:

Our application's capacity to provide informative reports is among its most useful features. You can keep an eye on how activities are being completed, assess how prepared they are, and see how project costs are being managed. You may use this reporting feature to make well-informed decisions, allocate resources effectively, and monitor project progress.

The importance of these features is found in their capacity to make project management seem less complicated. They make it unnecessary to preserve handwritten records, lower the possibility of mistakes, and facilitate decision-making in real time. Our goal is to enhance the efficiency and streamliness of project management by offering a user-friendly interface and strong functionality.

We'll go more deeply into the particular features, showing how our application satisfies the pre-established requirements. We go further into these features.

Let's examine the core of our product, which is how it makes project registration and related activities simpler.

Our system makes it simple for users to browse and search for projects, which simplifies the process of managing projects. We can help you with any kind of project, from specifics to an overview of all your active projects. A number of criteria, such as the project code, name, client, start and finish dates, and status, are available for searching through projects. This guarantees that you will be able to find the information you want fast.

Several key data elements are used to characterize each project in our system.

- To distinguish projects within the system, each project has its own code.
- The project name ensures clarity in project management by serving as a unique identifier.
- A project description enables thorough comprehension.
- The name of the project customer is registered; they can choose from the list of clients for projects that have already been registered in the system, or they can register as a new client.
- The project's start and finish dates determine how long it will take to complete.
- The state of a project can be indicated by its status, which might be "New," "In progress," "Terminated," "Frozen," or "Completed."
- An essential element is the hourly rate, which represents the amount paid to a participant for each hour of labor completed.

Usually, a project consists of several tasks, each of which requires careful attention. With the help of our program, tasks inside projects may be defined, providing an organized method of managing project activity.

- The names assigned to each assignment give clarification on what it entails.
- The desired outcome of each activity is outlined in the task description.
- There are start and finish dates to make sure every task is completed within the project's allotted time.
- The three priority levels—High, Medium, and Low—help you differentiate between important and non-important chores.
- There are several possible task statuses, including "Scheduled," "Pending approval," "Approved," "In progress," "Rejected," and "Completed." The state of newly created tasks is "Planned."
- A task's readiness is expressed as a percentage between 0 and 100. The readiness of new tasks begins at 0.
- Only when a job is in the "Completed" stage may hours worked on it be recorded.

We've made project and task management easier with these features, putting all the information you want at your fingertips.

The capacity to see and modify project data is essential for project management. That's exactly what our program does, giving you the resources, you need to efficiently manage your tasks.

An easy-to-use interface is offered by our system to view and amend project data. It's crucial to remember that editing data is only permitted in certain situations.

Editing project data is possible, however there are certain limitations. The only conditions under which you can change project data are those that do not include "Terminated," "Completed," or "Frozen." This restriction makes sure that data is only changed while the project is still active, avoiding unintentional modifications to finished or abandoned projects.

In order to enable effective project tracking, our program does more than just enter data; it also visualizes specific project information.

- You can quickly view how many days are left until each task's deadline or, in the event that it has already passed, how many days it is behind schedule. A precise schedule for work completion is provided by this feature.
- Furthermore, each project is given a thorough review by our system, complete with both summary and detailed data. The overall number of tasks, the total number of hours done, and the project's total cost determined by the number of hours worked are all included in this overview.

Our program makes sure you have a real-time grasp of the state of your projects and the advancement of specific tasks by offering these visualization options.

One of the most important aspects of project management is tracking project status. Our program makes this process easier by giving precise guidelines for altering the status of a project. Let's see how this operates.

Changing a project's status in our system complies with pre-established guidelines. These guidelines are intended to encourage effective project management and guarantee that status updates correspond with the project's advancement.

The following are instances of when a project may change from one state to another:

- Once a project is marked as "New," it may go on to "In progress," "Terminated," or even "Frozen." This makes it possible to manage new initiatives with flexibility.
- The project may be at the "In progress," "Completed," "Terminated," or "Frozen" stages at this point. The project's progress from the planning to the execution phases is represented by this shift.
- If the project status is "Terminated," it can be changed back to "In progress," which allows you to restart projects that were previously put on hold.
- Likewise, in case a project is designated as "Frozen," it has the ability to revert to "In progress," thus providing project managers flexibility.
- Projects remain in their final state after they are marked as "Completed" since they cannot be altered after that point.

Emphasizing the significance of user confirmation in the process of changing a status is crucial. Verification by users guarantees that the modifications correspond with the project's real advancement. This prevents unintentional status changes by adding an additional degree of responsibility and control.

Our tool streamlines project status management by imposing these guidelines and requesting user consent. This gives you a methodical and well-defined approach to efficiently monitor the advancement of your project.

Having the proper professionals on board is crucial in the field of project management. Our application makes it easier for experts—internal and external—to register, ensuring that the proper people are working on your projects.

For specialists, our system offers a simple and easy registration process. We've planned the procedure to be easy and effective, whether you're collaborating with professionals from outside the company or team members from within.

Let's examine the information related to experts in more detail:

- Name, Last Name, and Surname: An essential component of the expert's registration is their personal data. In order to make sure you can quickly recognize them in your project management system, this includes their whole name.
- Expert Type: distinct backgrounds can produce distinct experts, as our approach acknowledges. They may be external experts who work with your team on projects or internal experts who are currently employed by your company.
- Data Registration: Our application goes beyond simply registering professionals that work for your company. You have the option to register the pertinent expert's data if they are not yet included in the system. This guarantees that you may include outside specialists into your project management procedures with ease.

Our application facilitates the process of selecting the best team members for your projects by streamlining the registration procedure for specialists and offering comprehensive information about them. Our system makes sure you always have access to the necessary knowledge, whether you're working with in-house resources or outside specialists.

Viewing and editing experts' data is crucial after they've been registered in your project management system. With the help of our program, which offers a thorough platform for efficiently managing expert knowledge, you may accomplish exactly that.

Our solution is intended to make accessing and modifying expert data easy and effective for users. This guarantees you the freedom to easily lead your group of professionals.

We go one step further with our application, which lets you see expert data and their involvement in projects. You can see how your specialists are contributing to your projects using this graphic.

- You may examine comprehensive details about each expert, such as their kind (internal or external), name, and more. This makes it possible for you to comprehend the knowledge that your company possesses.
- The system also provides details on the projects that each expert is working on. You may easily keep track of the expert's contributions to several projects with the help of this graphic, which highlights their involvement.

This feature facilitates efficient cooperation among your project teams while also making the administration of expert data easier. It guarantees you have an understandable picture of your specialists and their responsibilities, which is essential for a project's effective completion.

Effective cooperation and data protection are based on secure access to the project management system. It's simple to utilize our program, which makes sure that only authorized users may access the system.

Our user login method, which requires a username and password, establishes access to our system. This is how it operates:

- Username: Every user in the system is uniquely identified by their username. In the registration procedure, this username is chosen or entered.
- Password: Users must provide a password linked to their username in order to maintain security. Only the user knows this password, which is kept private.
- Authentication: The system checks the credentials to make sure the user is allowed to access the system after the username and password have been entered.
- Access Granted: The user is granted access to the project management system following a successful authentication process. With this access, they can operate in accordance with their permissions and function inside the system.

By requiring a login, you can safeguard your project data and uphold the integrity of your project management procedures by limiting system access to only those who are allowed.

We'll close to review the main ideas we covered and emphasize how our application successfully satisfies project management requirements.

- We began by outlining the meaning and idea of project management. It's about efficiently achieving goals, not just a list of duties.
- We emphasized the necessity of an information system for efficient project management, particularly in the fast-paced world of today.

- A number of essential features are included in our React.js-built application, such as the ability to register tasks, projects, specialists, and contracts.
- We underlined the significance of reporting as it enables you to monitor job preparedness and efficiently control project expenses.
- We investigated project and task registration, comprehending the information related to every project and the specifics of task administration.
- To ensure that you can manage your projects efficiently, we spoke about accessing and modifying project data.
- The topic of changing a project's status was discussed, emphasizing the guidelines for doing so and the value of user approval.
- We made sure you have the correct personnel on your team by going through the expert registration procedure.
- We discussed how to see and amend expert data as well as how the system shows the experts' involvement in projects.
- Finally, we went over how users may log in to the system securely.

Our application aims to enable teams and organizations to operate more productively; it's not simply about software. Our extensive functionality and user-friendly interface make project administration less complicated. We make it unnecessary to retain manual records, lower the possibility of mistakes, and allow for instantaneous decision-making.

3. Conclusion

To sum up, our application offers a thorough response to the problems associated with project management. It is intended to streamline and improve the efficiency of project management while giving you access to all the resources you require for success.

We think that efficient goal-achieving and positive impact-making are more important aspects of good project management than just finishing duties.

Literature

- React.JS. Learn React. <https://react.dev/learn>
- Krug, S. Don't Make Me Think, Revisited. A Common Sense Approach to Web Usability. 2014, New Riders
- Weichbroth P., Sikorski M. User Interface Prototyping. Techniques, Methods and Tools. Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach. 2015.

OPERATING AND ANALYSING OF BIG DATA OVER PENTAHO BUSINESS ANALYTICS, TABLEAU AND QLIK

Mariana Kovacheva

асистент, доктор, катедра „Информационни технологии и комуникации“, УНСС
e-mail: mkovacheva@unwe.bg

Abstract

Technologies and digitalization are two of the biggest driving forces for generating, gathering, working, and analysing Big Data. This paper examines three analytical tools – Pentaho Business Analytics, Tableau and Qlik, their characteristics, advantages, and disadvantages for working with Big Data – structured, semi-structured and unstructured.

Keywords: Big Data, structured data, semi-structured, unstructured data

Introduction

The digitalization and digital transformation are two of the most important processes happening in the last decade and will continue to be so in the near future. Both are a reason for generating big amounts of data, known also as Big Data. All the data that is gathered can be from one of the three types – structured, semi-structured and unstructured, and shouldn't be only gathered but also analyzed and used for taking decisions.

Tools for analyzing data

In today's world with the vast number of different types of digital devices – from smartphones, tablets, laptops, smartwatches, other types of smart devices that we are using in our everyday life, sensors, etc. which are gathering data every second of their existence, the problem does not end with the data being gathered, but only starts with it. All the types of data – structured, semi-structured and unstructured – after being gathered, need to be stored properly, processed then analysed and after that conclusion to be made, based on it. Big Data presents many advantages for the people and organizations using it, but also many challenges as working with different sources of data, which from side need new means and methods for processing the data, but also new skills and vision for working with it. [1]

Nowadays, there is large selection of different tools for analysing data. Ranging from simple spreadsheets software (like Excel, Google Spreadsheets, etc.) to very advanced statistical or machine learning platforms (like Apache Spark, SAS, etc.). The choice of a tool for analysing data depends highly on many requirements and questions that a company should ask itself before choosing. [2]

Part of the things that need to be considered when choosing an analytical tool are the shown in Table 1.

Table 1 Requirements for choosing analytical tool

Questions	Variations
What is the type of data that is gathered?	There is a possibility for the company to work with one of the following types structured, semi-structured or unstructured, or even with more than one.
What is the size of the data?	Is the company generating small files of MBs or maximum GBs, or is generating Big Data – PB, ZB, etc.
Who will use the analytical tool?	Experts like data analysts or data scientists, or marketers and salespeople.
What types of visualizations are needed?	Some analytical tools offer visualizations that another cannot make and vice versa. For this

	reason, is important for the company to clear what type it will need.
What is the budget?	Another important question is what the price of the tool be. There are different options from free, open source, paid tools. This choice varies depending on the previous questions – data types that are going to be gathered and analysed, who will use the tool, size of the data, are specific visualizations needed.

Based on the questions in the table above and their answers, it becomes easier for a company to make a proper choice for analyzing tool for their needs.

Comparison of tools for analyzing data

This paper examines Tableau, Pentaho Business Analytics and Qlik as analytical tools, which according to Gartner are in the Top 20 of their classification for Analytics and Business Intelligence Platforms Reviews and Ratings. [3]

Tableau

Tableau is a data visualization tool and business intelligence software that allows the users to connect, visualize and share data interactively. The instrument is used to show data in interactive and understandable way. [4]

Tableau is used for [5]:

- 1) **Data visualization** – transforming the data into easily understandable and readable charts, maps, and graphs.
- 2) **Business intelligence** – insights from data for the business to make informed decisions.
- 3) **Data analysis** – for exploration and analysing big sets of data.
- 4) **Dashboard creation** – for creating interactive real-time dashboards.
- 5) **Operational reporting** – for creation of reports.

Advantages and Disadvantages of Tableau

Tableau is one of the most popular business intelligence tools on the market. It has user friendly interface which allows the user to make the needed dashboards and reports with minimum effort. With the fact that Tableau supports different data sources like spreadsheets, databases, big data platforms make it easy to connect and analyse data. The tool offers great variety of interactive visualizations, which enable the user to customize its needs and demands easy. Tableau is scalable and can work with small and big data sets, to meet the needs of every type of business. Finally of its advantages, shown on Figure 1, is the fact that Tableau can be integrated with different data sources and with other business intelligence tools. [6][7]

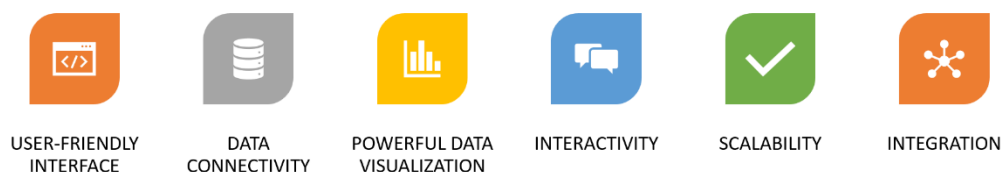


Figure 1 Advantages of Tableau

As every tool on the market, Tableau has not only positive sides, but also negative which are shown on Figure 2. The BI tool can be expensive for larger organizations and the cost may include licenses, support, training, etc. Even that Tableau is scalable and can work with different sources, when it's working with extremely large datasets, some issues on the performance may arise. On the other side, some of the complex features of Tableau to be used, may require additional learning for the user to maximize the potential of the software. One of the other downsides of Tableau is that the data needs to be pre-processed – cleaned and transformed in other tools before coming to the BI tool. Also, Tableau doesn't support statistical analysis and some features for advanced analytics and interactivity are limited or unavailable when Tableau is offline. [6] [7]



Figure 2 Disadvantages of Tableau

Pentaho Business Analytics

Pentaho Business Analytics is a business intelligence platform which is an open source and provides data integration, reporting, data visualization and analytics.

Pentaho Business Analytics is used for [9]:

- 1) **Data Integration** – extracting data from various sources, transforming it into usable format and loading the data into a system that will work with it. This is also known as ETL or Extract, Transform and Load. This process is very important for organizations that want to gather data from different data sources.
- 2) **Business Intelligence Reporting** – this tool allows creation of reports based on the data loaded.
- 3) **Data Visualization and Dashboards** – the BI tool provides instrument for creating dashboards and visualizations which are interactive. The users can design their dashboards and communicate the insights within the organization.
- 4) **Predictive Analytics** – allows to discover trends and patterns inside the data.
- 5) **Aggregation and Performance Optimization** – designs and manages aggregate tables, optimizes queries performance in the data warehouse, which is very important when talking about large datasets.
- 6) **Ad hoc Reporting and Analysis** – the tool allows creation of reports on the go, which is important for companies that need fast visualization and analysing of data.
- 7) **Data Warehousing** – can be used for easier ETL processing.
- 8) **Open-source customization** – as the platform is open-source can be customized by the specific needs and demands of the organization, which is very useful for companies with specific data formats.
- 9) **Integration with other systems** – Pentaho Business Analytics integrates good with different databases, sources, tools, and different infrastructure.

Advantages and Disadvantages of Pentaho Business Analytics

Pentaho Business Analytics has many advantages shown on Figure 3. One of the biggest pros of the BI tool is that is open source, which means that because the source code is open, it can be modified according to the needs of the company and can be fitted in budget constraints. The tool is designed to be scalable, which means that with the growing of a small company to big one, Pentaho can still be the software that is used, because can also work good with large datasets. In addition to that, Pentaho has good integration capabilities and work with different data sources, formats, databases, etc. [10] [11]



Figure 3 Advantages of Pentaho Business Analytics

As every tool on the market, Pentaho Business Analytics also has some disadvantages, which depending on the company and the case, may become crucial for the selection of the tool – Figure 4. The analytics capabilities of Pentaho are not as advanced as of the specialized tools and especially in the advanced predictive analytics or machine learning. The BI tool has limited visualization capabilities, even that provides reporting and dashboards, the quality and advancement is not as good as this of the specialized tools. Pentaho is built on Java and if the organization is using as a tool, there might be need of Java knowledge. Some companies may prefer other technologies. Pentaho requires bigger amounts of resources as computing power and memory; this also highly depends on the task and on the size of the dataset. Some of the features of the tool are not available in the open-source edition and if the company needs some advanced features, which are only available in the commercial version, this will add additional cost to the organization and directly becomes disadvantage. [10] [11]



Figure 4 Disadvantages of Pentaho Business Analytics

Qlik

Qlik is a business intelligence tool and data visualization platform. It is known for its associative data model which enables the users to explore the relationships between the data. According to Gartner Magic Quadrant, Qlik is the leader in the sphere in data integration, quality and analytics solutions which are offered by the tool [12].

Qlik is used for [13]:

- 1) **Business Intelligence** – this tool is used for analysing and interpreting the organization's data. Qlik has associative data model which allows to explore the data relationships and to make more informed decisions for the business.
- 2) **Data Visualization** – interactive and visually appealing dashboards and reports can be created. Through the creation of different types of dashboards, graphs, charts, etc. users can make more informed decisions. Qlik allows the customization of dashboards and reports as in this way the KPI's are well communicated through the organization.
- 3) **Data Integration** – allows connection and integration of data different sources, which allows companies to view their data connected at one place.
- 4) **Collaborative Analytics** – users can collaborate in between each other as they can share dashboards and visualizations.
- 5) **Predictive Analytics** – Qlik has abilities for predictive analytics, even it is not the tool main aim, but the users are able to use it for forecasting trends and to make predictions.

Advantages and Disadvantages of Qlik

As every tool, Qlik has its own advantages and disadvantages, which will be examined in the following paragraphs. The first advantage is that Qlik is scalable, which means that its power can grow depending on the needs of the company, which usually means bigger datasets to work with. The data visualizations of the BI instrument are very powerful, which enables the users to create easily interactive and understandable dashboards. The associative model of Qlik makes it easy and flexible for the users to look and explore the relationships in the dataset. Qlik supports integration of data from various sources, which makes it easier for companies to work with unified data sources – databases, files and many more repositories. It is very user-friendly and easy to use, only with drag and drop the user can create dashboard. [14]

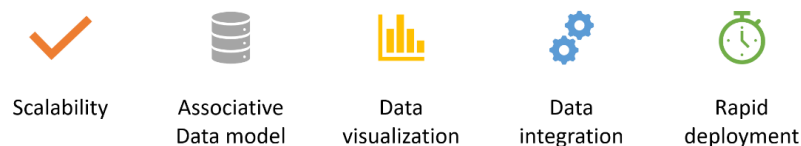


Figure 5 Advantages of Qlik

Qlik has limited predictive analytics and organizations which need advanced predictive modelling will need specific tools for this aim. Another disadvantage is the price of Qlik, because of licensing and another costs, especially for bigger companies the price can be very high. Another thing that can make organizations choose a different tool is that Qlik is very resource intensive, which depends highly on how big the data sets are and how complex should the visualizations be. The training of

personnel who is using Qlik may be needed to fully use the tools capabilities. Finally, Qlik highly depends on the data quality. If the data is incomplete, inaccurate, or misleading in some way, the results and analysis won't be accurate. [15]



Figure 6 Disadvantages of Qlik

Comparison of Tableau, Pentaho Business Analytics and Qlik

Based on the previous points and the detailed information that we looked in this paper we can make a comparison table of the three analytical tools in Table 2 [16] [17].

Table 2 Comparison of Tableau, Pentaho and Qlik

	Tableau	Pentaho	Qlik
Data Integration	More focused on visualizations.	Focus on ETL processes.	Data integration
Ease of Use	Most user-friendly	Less user-friendly than the three	User-friendly
Cost	More expensive with licenses and everything needed for its use.	Free open-source version, the enterprise one is expensive.	More expensive with licenses and everything needed for its use.
Scalability	Less scalable than Qlik	Less scalable than Qlik	Most scalable of the three, because of its in-memory processing and associative data model.
Support	Largest and active community.	The support may be dependent on enterprise edition.	Big and active community.
Customization	The tool provides customization in terms of visualizations and dashboards.	It's customizable, but Pentaho requires more technical knowledge.	The tool provides customization in terms of visualizations and dashboards.

Conclusion

Digitalization and digital transformations are two of the most important driving forces nowadays. All the data that is generated needs to be not only gathered, but to be properly processed and analysed for the companies to be able to use it for their business problems. In this paper we examined three of the most popular analytical tools, which are used not only for visualizations and dashboards, but for ETL processes, data integration and customizations for the need of the business to be met. A comparison of the most important features of the three tools was made.

References

- [1] S. Yordanova and K. Stefanova, 'Big Data Challenges - Definition, Characteristics and Technologies', no. 1, pp. 13–31, 2019.
- [2] J. Zhang, '10 Data Analysis Tools for Beginners and Experts', Medium. [Online]. Available: <https://towardsdatascience.com/10-data-analysis-tools-for-beginners-and-experts-2d083203b06e>
- [3] 'Analytics and Business Intelligence Platforms Reviews and Ratings'. [Online]. Available: <https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms>
- [4] 'What is Tableau?', Tableau. [Online]. Available: <https://www.tableau.com/why-tableau/what-is-tableau>
- [5] 'What is Tableau? | Beginner's Guide to Key Features', DATADRIVE. [Online]. Available: <https://godatadrive.com/blog/what-is-tableau>
- [6] 'Advantages and Disadvantages of Tableau', JavaTpoint. [Online]. Available: <https://www.javatpoint.com/advantages-and-disadvantages-of-tableau>
- [7] 'What is Tableau? A Guide to Tableau Features and Advantages', BeyondKey. [Online]. Available: <https://www.beyondkey.com/blog/what-is-tableau/>
- [8] D. Tarnaveanu, 'Pentaho Business Analytics: a Business Intelligence Open Source Alternative', vol. III, pp. 23–34.
- [9] 'Pentaho Platform: Pentaho Business Analytics - Datasheet', Hitachi Vantara. [Online]. Available: <https://www.hitachivantara.com/en-us/web/datasheet/lumada-dataops-pentaho-business-analytics-datasheet.html>
- [10] 'Pros and Cons of Pentaho', Trust Radius. [Online]. Available: <https://www.trustradius.com/products/pentaho/reviews?qs=pros-and-cons#overview>
- [11] N. Singh, 'Pentaho and Big Data'. [Online]. Available: <https://www.codingninjas.com/studio/library/pentaho-and-big-data>
- [12] 'What is Qlik?', Qlik. [Online]. Available: <https://www.qlik.com/us/>
- [13] K. Schlegel *et al.*, 'Magic Quadrant for Analytics and Business Intelligence Platforms', Apr. 2023. [Online]. Available: <https://www.gartner.com/doc/reprints?id=1-2D773G95&ct=230411&st=sb>
- [14] L. Katerova, 'All You Need to Know About Qlik Cloud. Moving From On-Premise to Cloud.' [Online]. Available: <https://b-eye.com/blog/all-you-need-to-know-about-qlik-cloud/>
- [15] 'QlikView Advantages and Disadvantages | QlikView Benefits', Data Flair. [Online]. Available: <https://data-flair.training/blogs/qlikview-advantages-and-disadvantages/>
- [16] R. Kaur, 'Pentaho vs Tableau: Which Business Intelligence Software is the Winner?' [Online]. Available: <https://www.selecthub.com/business-intelligence/pentaho-vs-tableau/#:~:text=Tableau%20supports%20only%20row%2Dlevel,tool%20overall%20for%20this%20category.>
- [17] 'Qlik Vs Pentaho: In-Depth Comparison'. [Online]. Available: <https://6sense.com/tech/data-visualization/qlik-vs-pentaho>

ТЕХНИЧЕСКИ РЕСУРСИ ЗА ПОДПОМАГАНЕ НА ХОРАТА СЪС ЗРИТЕЛНИ УВРЕЖДЕНИЯ

Елена Филипова

Докторант, катедра „Информационни технологии и комуникации“, УНСС, efilipova@unwe.bg

Резюме

В настоящия доклад се прави опит да се изследва как техническите ресурси и новите информационни технологии и системи ще подпомогнат професионалната реализация на хората със зрителни увреждания. Представена е обобщен списък на техническите средства и информационни технологии, посредством което се прави опит да се улесни връзката работодател – човек със зрителни увреждания с цел професионална реализация.

TECHNICAL RESOURCES TO HELP PEOPLE WITH VISUAL IMPAIRMENTS

Elena Filipova

Abstract

This report attempts to explore how technical resources and new information technologies and systems will support the professional realization of people with visual impairments. A summarized list of technical means and information technologies is presented, by means of which an attempt is made to facilitate the relationship between an employer and a visually impaired person with the aim of professional realization.

Въведение

Развитието на техниката и технологиите, както и това на икономическите отношения в обществото водят до това, че някои професии, които са били традиционни занимания на слепите днес вече изчезват, заменят се с други или включват се в рамките на някакви други конфигурации. От друга страна същите тези процеси пораждаат и нови професии и нови възможности за тяхното усвояване от страна на незрящите.

Съвременното информационно общество има достатъчно натрупан потенциал за използване на информационните и комуникационни технологии (ИКТ) в помощ на хората със зрителни, слухови и физически увреждания. ИКТ могат да играят ключова роля в осигуряването на равен достъп до образование и работа за тези лица, ето защо е важно да се популяризират и дискутират съществуващите съвременни ИКТ решения, които могат да им бъдат полезни в този процес.

Днешните информационни технологии дават големи възможности да се подпомогнат хората със зрителни увреждания в тяхната социализация.

Държавата губи от неползването на този икономически ресурс. Числеността на този клас хора не е никак малък те биха могли значително да подпомогнат БВП. Ако си позволим да използваме икономическа терминология те са икономически и финансов лост които се използва от държавата в твърде нисък процент спрямо тяхната численост. В тази разработка ще се опитаме да изследваме как техническите ресурси и новите информационни технологии и системи ще подпомогнат професионалната реализация на хората със зрителни увреждания. Представен е

обобщен списък на техническите средства и информационни технологии, посредством които може да се улесни връзката работодател – човек със зрителни увреждания с цел професионална реализация.

Технически средства и информационни технологии използвани за подпомагане на хората със зрителни увреждания

При обзора на научните изследвания, свързани с използването на ИКТ за работа с хора с увреждания веднага може да се установи, че в глобален план ролята на ИКТ в процеса на приобщаване на тези лица не е задълбочено и мащабно изследвана.

Съществуват частични решения на проблема с готови вече ИТ приложения за отделни задачи като например Брайлова азбука, ОЦР, четци на текст и др., но те не са обединени и съгласувани в единна информационно-педагогическа система.

Тези хора имат нужда от допълнително внимание и подготовка, за да бъдат включени пълноценно в обществото. За тях трябва да се разработват специализирани програми за обучение и работа, които да са подходящи за възприемане и адаптирани към специфичните проблеми на различните типове увреждания. Същото важи и за учебните ресурси, разработвани с помощта на ИКТ.

За тази цел са необходими както методически знания за спецификата на техните затруднения, породени от вида и степента на тяхното увреждане, така и знания за технологичните параметри на различните специализирани хардуерни и софтуерни приложения.

По тази причина по-нататък може да се направи опит за систематизиране съществуващите и използваните информационните технологии за интеграция на хората с зрителни увреждания, в зависимост от техните увреждания и категорията в която попадат. Целта е да се покаже какви информационни технологии биха били най-полезни за обучението, интеграцията и трудовата реализация на тези хора.

Технологичните решения, които се използват за хора със зрителни увреждания най-общо могат да се разделят на високотехнологични и ниско технологични помощни средства. Към ниско технологичните решения спадат различни помощни средства от вида на уплътнители за химикалки, адаптирани средства за писане, изрязване, четене и други. Ние ще се спрем по-подробно на високотехнологичните средства.

При лицата със сензорни увреждания отношението към ИКТ е по-различно – за тези хора ИКТ са инструмент и помощно средство, което играе ключова роля в тяхното приобщаване и те гледат на тях не като на развлечение или награда, а като на съществена необходимост.

За да се впишат успешно в информационното общество и да се възползват от основното му благо – информацията в цифров формат, незрящите хора се нуждаят от допълнителни хардуерни и софтуерни решения, които да трансформират и адаптират цифровите ресурси в подходящ за възприемане вид. Има много програми, синтезиращи човешка реч, но не всички са предназначени изцяло за хора без зрение.

Тук правим едно важно уточнение: изложения по долу подбор на технологични решения за интеграция и трудова реализация на незрящи хората – не е напълно изчерпателен, тъй като съществува многообразие от приложения, софтуерни продукти и информационни технологии и не бихме могли да обхванем всички тях. Акцент е поставен върху най-приложимите, общо известни и функционални технологични решения. Избора на конкретно ИТ решения изцяло трябва да се базира върху вида, спецификата и степента на заболяването на конкретния човек със зрително увреждане. Ето защо е от първостепенно значение при интеграцията, социалната и

трудова реализация на тези хора те да посочат: какво би им било най-полезно, с какво са свикнали да работят до момента и какво би ги предразположило, мотивирало и подпомогнало в ежедневната и трудовата им дейност.

Не може да говорим за интеграция и реализация без социалния контакт с тези хора, т.е. контакта с тях е от първостепенно значение.

Ако се разгледа едно работно място, като съвкупност от технологични средства – хардуерни и софтуерни, то за да е в равностойна ситуация човекът със частична или пълна загуба на зрение, неговото работно място трябва да осигурява възможности за пълноценна адаптация и трудова реализация.

Най-общо информационни и комуникационни технологии в помощ на хората със зрителни увреждания (като се фокусираме върху технологии използвани на работното място) могат да бъдат разделени на: софтуерни, хардуерни, IT средства и помощните технологии със специфичен дизайн в помощ на хората с увредено зрение.

Софтуерни технологии и средства

Софтуерните технологични средства, в по-голямата си част са свободно достъпни за хората със зрителни и други увреждания. Най-популярните сред тях са:

- Браузъри за достъп до Интернет ресурси – WebbIE, свързан с екранен четец като JAWS, WindowEyes, Thunder, NVDA, Narrator. Като пример за четец, който интерпретира български език може да бъде посочен SpeechLab 2.0.;
- Програми за оптично разпознаване на текст, с които може да се трансформира печатен текст в дигитален и след това да бъде възпроизведен чрез екранен четец;
- Софтуер за превод от текст на български език към Брайлова азбука;
- Софтуер за конвертиране на електронен текст в Брайлов код;
- Програми за екранни четци и за увеличение на екранното изображение;
- Софтуерен екранен четец съчетани със синтетична реч;
- Програми, синтезиращи човешка реч, като най-популярната и използваната в България програма ГЕРГАНА, която синтезира реч на български, т.е. има вграден кирилизатор на речта;
- Специализиран софтуер за оптично разпознаване на текст - софтуерът използва брайлов код, който е система за декодиране и кодиране на информация за улеснение на незрящите хора;
- Приложения за мобилни телефони, с помощта на които може да снима и разпознава печатен текст, който се превежда и изговаря на съответния език, ако е необходимо;
- Мобилни приложения, специално разработени за търсене на работа от хора със специфични потребности;
- Софтуерна лупа за лица с остатъчно зрение, възможно е тя да бъде комбинирана с брайлов дисплей;
- Брайлови асистент: Брайловият асистент е с инсталиран софтуер за четене на екрана.
- Софтуер за Брайлов превод на текстови документи - позволява превръщането на документи под формат HTML, XML, RTF, TEXT и Microsoft Word в брайлова и синтетична реч.
- Брайлов четец: Брайловият четец е предназначен за хора със зрителни увреждания, които не могат да четат по Брайловата тактилна система.

- Концепция за естествен потребителски интерфейс или физически интерфейс – система за взаимодействие между човек и компютър като потребителят извършва интуитивни действия, свързани с естественото му поведение, и чрез тях контролира компютъра.
- Говорещи програмни средства, които подобряват уменията за четене и произнасяне на думи и изрази;
- Уеб базирани среди за електронно обучение на хора с увреждане на слуха.

Хардуерни средства

Докато софтуерните средства в по-голямата си част са свободно достъпни, хардуерни средства са със сравнително висока цена и достъпът до тях изисква настройка и зареждане. Сред най-популярните хардуерни средства са:

- Брайловата мишка, която осигурява възможност на хората с различни увреждания на зрението да четат текст от компютърен екран;
- Скенери и програми за оптично разпознаване на текст, с които може да се трансформира печатен текст в дигитален;
- Брайлов телефон, който е наличен на пазара от 2014 г. Корпусът на телефона се изработва с помощта на 3D принтер.
- Текстов телефон с визуален и брайлов дисплей за глухи и за сляпо-глухи лица;
- Брайлов дисплей и брайлов принтер за сляпо-глухи с пълна глухота и значително зрително увреждане, и за напълно сляпо-глухи;
- Екранен четец - е софтуерно приложение, което се интегрира и идентифицира това, което се показва на екрана.
- Преносим компютър SIAFU, който е проектиран специално за хора със зрителни увреждания и цели да направи работата с компютри по-интуитивна.
- Таблетът за хора със зрителни увреждания – използва цилиндърчета от специален материал, които динамично ще се показват и скриват и така върху четеца ще се образува релефното Брайлово писмо.
- Брайлов асистент е съвременно устройство, което притежава предимства, изразяващи се в наличие на достъпен сензорен екран, Брайлов дисплей и бележник. Брайловият асистент включва: 12-клетъчен обновяем брайлов дисплей с тактилен сензор и технология за автоматична маршрутизация на курсора; 6 входящи брайлови и 3 функционални клавиша за писане; джойстик и 2 навигационни клавиша.
- Тактилни хардуерни устройства – очила за управление с поглед с тактилен преобразувател - оптикон.

IT средства и помощни технологии със специфичен дизайн в помощ на хората с увредено зрение

- Монокояр: така се нарича увеличителен уред мини-скоп, който може да се държи в ръка или в джоб и действа на принципа на телескопа.
- Говорещ калкулатор: осигурява звукова обратна връзка на направените изчисления.
- Уголемяващ печатащ софтуер, който уголемява текста на компютърния екран.
- Компютърна клавиатура с уголемени главни букви,; черни на бял фон, клавиши отпечатващи букви в голям размер.

- Компютърни и технологични аксесоари: Увеличителен екран и четящ софтуер, софтуер записващ глас, клавиатурен водач на Брайл (списък и описание на всички команди на клавиатурата и нейните функции), оцветени клавиатури.
- Силно отбележима писалка: това е писалка със специфичен вид течност, с помощта на която като се маркира, дава възможност за осезаемо отбелязване и контрол на обекти и съоръжения.
- Свободен увеличител: той се използва при увеличаването на печатни материали, извършване на ръчна работа с фини детайли и др.
- Подръчен увеличител: този увеличител изпълнява същите функции както и свободния увеличител, само че той е портативен и има възможност към него да бъде включена и допълнителна светлина.
- Говорещ будилник/Скали/Рулетка.
- Индикатор показващ нивото на течност: този индикатор възпроизвежда пронизателен звук, когато някаква течност достигне горната граница на определен съд или контейнер.
- Ориентировъчни инструкции и инструкции за предвижване: те могат да бъдат професионално подготвени с персонала на компанията, така че да дават възможност човекът с увредено зрение да се предвижва свободно и независимо.
- Ограничител на точки: Това са осезателни маркери, които се поставят/залепват на предмети и обекти, като на контролния пулт на автоматична пералня, където могат да се маркират копчетата за съответните функции.
- Едро отпечатани телефонни номера: за да могат да бъдат видени и запазени всички изходни телефонни номера.
- Телефон с говорещо устройство, за това кой ни търси ID (идентификационен номер).

Клетъчни телефонни принадлежности:

- Клетъчен телефонен увеличител;
- Нормировъчна облицовка;
- Свободни ръце (хендсфри);
- Звуков усилвател;
- Клетъчни телефонни увеличителни лещи (те уголемяват дисплея и копчетата от Клавиатурата на телефона, така сякаш виждаме образа през монокъл.

Четящи средства:

- Четящи машини, говорещ рекордер и плейър, светлинни книги, светлинен увеличител за книга;
- Поставка за книги (поставка на която може да бъде нагласена и поставена книгата заедно със съответен увеличител);
- Говорещи електронни речници.
- Аудио информация: от говореща книжна библиотека – включваща книги, новели и списания.
- Гласови активни системи: те позволяват на потребителите да командват компютрите си с помощта на своя глас.
- Гласов синтезатор на работното поле на компютъра: издава звуков сигнал за обратна връзка за това какво се случва на вашия компютърен екран.

- CCTV - Closed Circuit Television (Затворен Телевизионен Кръг): Той уголемява напечатания материал, за четене на печатно издание или писане на бележки.
- Паричен тест за монети и бележки: това е средство за идентификация на пари.
- Говорещ личен органайзер
- Брайлова щампа/печат: това е вид хардуер, който отпечатва документи и копия на Брайл. Брайлво превеждащ софтуер осъществява превод на текст от компютъра на Брайл.
- Брайлов софтуер (Брайлов преводач): това е софтуер за превод на текстови книги, писма, ADA-компиляция на песни и др.
- Брайлова маркировка на продукти.
- Брайлови етикети.
- Брайлови часовници, таймери.
- Печатни скенери: сканират материала и го възпроизвеждат като аудио прочит на текст.
- Голям монитор: По-голям от обикновено използвания компютърен екран.
- Слухов сигнал за тревога/информационна система: те се използват при лифтовете и другите обществени съоръжения.
- Telebraille(Телебраил): това е вид дисплей който обединява ГТУ и брайлов дисплей.

Обзорът на тези технически средства за подпомагане на хората със зрителни увреждания не е самоцел. Дори само ако се проследи техният брой става ясно, че пазарът е наситен с технически средства, но това не е достатъчно. Те трябва да достигнат до тяхното местоназначение. Тук е ролята на държавата, да подпомага хората със зрителни увреждания, тъй като голяма част тези хора не могат да си позволят да закупят такива средства. Тук не се прави преглед на цената им, но тя никак не е малка. Голяма част от тези средства не се произвеждат в големи серии, количествата им са ограничени, което води до доста висока цена.

Заклучение

Държавата ни абдикира на практика от задълженията си към тази социална група. Проблемите на незрящите станаха в много по-голяма степен техни собствени проблеми и те трябва да търсят тяхното решение със собствени сили, с помощта на различни организации на и за незрящи хора, на други институции на гражданското общество, чрез които да оказват натиск или по-точно влияние върху държавната политика и по този начин стъпка по стъпка да се стремят да подобряват социалното положение на хората от тази социална група включително и тяхното положение на пазара на труда.

Използвана литература

1. Das, S., Chen, M. Yahoo! for Amazon: Extracting market sentiment from stock message boards, Proceedings of the Asia Pacific Finance Association Annual Conference (APFA), Vol. 35/2001.
2. Сълова, С. Методи за извличане на мнения и чувства от текст. В: Предизвикателства пред информационните технологии в контекста на "Хоризонт 2020", СА- Свищов, 2016.
3. French, Richard Slayton (23 March 1919). The Education of the Blind, a Critical and Historical Survey with Special Reference to the United States of America (Thesis). University of California.
4. Siu Yue-Ting, Ike Presley, Access Technology for Blind and Low Vision Accessibility Illustrated Edition, Guide to Assessment, 2008

5. "The Land of Blind". Animator.ru. - "The Country of the Blind".
6. Ike Presley, Frances Mary D'Andrea, Assistive Technology For Students Who are Blind or Visually Impaired: A Guide to Assessment. Illustrated Edition, 2012.
7. Roberto Manduchi, Sri Kurniawan, Assistive Technology for Blindness and Low Vision, 2017.

ПРОДУКТОВИ КАТЕГОРИИ ЗА ХОРА СЪС ЗРИТЕЛНИ УВРЕЖДЕНИЯ

Елена Филипова

Докторант, катедра „Информационни технологии и комуникации“, УНСС, efilipova@unwe.bg

Резюме

В настоящата разработка ще представим различните продуктови категории за хората със зрителни увреждания. Целта е да извлечем софтуерните, хардуерните средства, адаптивни съоръжения и помощни средства с помощта на които незрящите хора биха се справили с ежедневните предизвикателства и трудности. Ще представим и наша авторска класификация на тези средства, адаптирана към българските условия и реалности.

PRODUCT CATEGORIES FOR THE VISUALLY IMPAIRED PEOPLE

Elena Filipova

Abstract

In this paper, we will present the different product categories for the visually impaired. The aim is to derive the software, hardware, adaptive equipment and assistive devices with which blind people would cope with everyday challenges and difficulties. We will also present our author's classification of these funds, adapted to Bulgarian conditions and realities.

Продуктова база данни и продуктови категории за хора със зрителни увреждания според American Foundation for the Blind Product

Ще се спрем на софтуерните, хардуерните информационни технологии и адаптивни съоръжения и помощни средства в помощ на незрящите хора. Като база ни служи категоризацията на продуктите в помощ на хората с увредено зрение от American Foundation for the Blind Product Search. Избрали сме тази фондация, тъй като тя е едно от най-популярните структури за популяризиране на продуктите за хора със зрителни увреждания.

По-долу на кратко се дава пояснение на основните категории продукти. Разбира се тук не може да говорим за изчерпателен и пълен списък на съществуващите ИТ технологии, адаптивни съоръжения, технически средства и мобилни приложения които могат да се използват от хората със увредено зрение. С бързото развитие и бума на технологиите този списък се обогатява постоянно.

Но дори и така, с този ограничен списък, ясно се вижда от представеното от Американския съюз на слепите, че продуктите за подпомагане на хората с увредено зрение са разнообразни и разработени с идеята да покриват голяма част от нуждите им. В настоящата разработка не изследваме цени на продуктите, но трябва да е ясно, че тъй като това са специално разработени продукти, които не са масово производство, цената им е доста висока.

Преглед на продуктовата категория на American Foundation for the Blind Product Search:

Продукти по категория		
1	Accessible Mobile Apps	Достъпни мобилни приложения
2	Accessible Software	Достъпен софтуер
3	Audible and Tactile Signs and Warning Surfaces	Звукови и тактилни знаци и предупредителни повърхности
4	Braille Printers	Брайлови принтери
5	Braille Translators	Брайъл преводачи
6	CCTVs/Video Magnifiers	Видеонаблюдение / Видео лупата
7	Deaf-Blind/Multiple Disabilities	Глухо-слепи / множество увреждания
8	Digital Talking Book Players (Hardware)	Играчи с цифрова тетрадка (хардуер)
9	Digital Talking Book Players (Software)	Плейъри за цифрова книга (софтуер)
10	Educational Technology	Образователна технология
11	Electronic Notetakers (Braille)	Електронни бележници (Брайлова азбука)
12	Electronic Notetakers (Speech)	Електронни бележници (реч)
13	Games and Activities	Игри и дейности
14	GPS (Hardware)	GPS (хардуер)
15	GPS (Software)	GPS (софтуер)
16	Household, Personal and Other Independent Living Products	Домакински, лични и други независими продукти за живот
17	Lighting	осветление
18	Low Vision Optical Devices	Оптични устройства с ниска визия
19	Miscellaneous Speech Products	Разни продукти на речта
20	Optical Character Recognition Systems	Оптични системи за разпознаване на символи
21	Refreshable Braille Displays	Брайлови дисплеи с възможност за обновяване
22	Screen Magnification Systems	Системи за увеличаване на екрана
23	Screen Readers	Четец на екрани
24	Speech Synthesizers	Реч синтезатори
25	Windows-Based Tutorials	Windows-базирани уроци

На базата на информацията за продуктите и средствата в помощ на хората със зрителни увреждания, направена от American Foundation for the Blind Product Search, ще правим опит за

извеждане на структура и подредба на тези технологии и технологични продукти, подходяща за нашите, български условия.

Защо е необходимо такава преподреждане на продуктите категории, специфично за нашите условия? Причините са няколко:

1. Българските организации, занимаващите с проблемите на хората с увредено зрение имат известни традиции и възможности за работа с част от продуктите. Това е необходимо, защото хората трябва да се обучат в използването на новите технологии.

2. Технологичните средства и продукти за хора с увредено зрение трябва да са съвместими с тези, които се използват в икономиката на страната ни, за да има безпрепятствен трансфер на данни между системите.

3. Не на последно място значение има и цената на предлаганите продукти – тя трябва да бъде съобразена с нашия стандарт на живот и възможностите на българските граждани. Дори и държавата да подпомага закупуването на технологични средства за хората със зрителни увреждания, трябва да съществува разумно съотношение между цена – качество – необходимост от продукта. Поддържането на баланс между тези три характеристики е важно, за да могат да се удовлетворят нуждите на повече хора от тази обществена група.

Нашата авторска класификация на продукти и технически средства за хора със зрителни увреждания включва:

Софтуер	
1,1	Достъпен софтуер
1,2	Образователна технология
1,3	GPS (софтуер)
1,4	Брайлови преводачи
1,5	Видеонаблюдение / Видео лупата
1,6	Системи за увеличаване на екрана
1,7	Четец на екрани
1,8	Реч синтезатори
1,9	Плейъри за цифрова книга (софтуер)
Хардуер	
2,1	Брайлови принтери
2,2	Електронни бележници (Брайлова азбука)
2,3	Електронни бележници (реч)
2,4	GPS (хардуер)
2,5	Играчи с цифрова тетрадка (хардуер)
2,6	Оптични системи за разпознаване на символи
2,7	Брайлови дисплеи с възможност за обновяване

Адаптивни Съоръжения	
1	Адаптивни Съоръжения
2	Звукови и тактилни знаци и предупредителни повърхности
3	Домакински, лични и други независими продукти за живот
4	Осветление
5	Оптични устройства с ниска визия
6	Продукти на речта
7	Глухо-слепи съоръжения / множество увреждания

Заклучение

Трябва да се отбележи фактът, че днес радикално се променят както отношението на обществото към проблемите и възможностите на хората със зрителни увреждания, така и отношението на лицата от тази група към самите себе си, към своите собствени възможности, към тяхното място в обществото и в техния собствен живот и в решаването на техните собствени проблеми. Те вече се чувстват като пълноправни граждани и изискват държавата да им обръща подобаващо внимание. Това внимание трябва да се изразява в грижа за техните потребности и максимална подкрепа за трудова реализация и интеграция в обществото.

Всеки човек с нарушено зрение има своя уникална характеристика от професионални качества. Онова, което е възможно за един, може да не е възможно за другите и обратно. Трудовата реализация е въпрос на лична мотивация, професионална подготовка, способност за интеграция в обществото и справяне с ежедневните предизвикателства.

Използвана литература

1. American Foundation for the Blind. <https://www.afb.org/>
2. Съюз на слепите в България. <https://ssb-bg.net/>
3. Лазарова, В., Филипова, Е. (2023). Big Data в помощ на реализацията на хора със зрителни увреждания. В Дигитализация , големи данни, изкуствен интелект. Сборник с доклади от научен семинар в рамките на проект „Дигитализация на икономиката в среда на големи данни" (с. 24–32). Варна: Издателство „Наука и икономика“.
4. Lazarova, V., & Filipova, E. (2020). Assimilating the skills of people with specific needs for the IT labor market by utilizing the information technologies for the workplace. In 8th [Eight] International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE – 2018), October 18-20th, 2018, UNWE, Sofia, Bulgaria (285-287). Sofia: Publishing Complex - UNWE. ISSN (online) 2367-7643. Retrieved from <http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2018.pdf>
5. Siu Yue-Ting, Ike Presley, Access Technology for Blind and Low Vision Accessibility Illustrated Edition, Guide to Assessment, 2008

6. Ike Presley, Frances Mary D'Andrea, *Assistive Technology For Students Who are Blind or Visually Impaired: A Guide to Assessment. Illustrated Edition*, 2012.
7. Roberto Manduchi, Sri Kurniawan, *Assistive Technology for Blindness and Low Vision*, 2017.

DATA PREPARATION TECHNIQUES AND PLATFORMS IN THE CONTEXT OF MACHINE LEARNING

Genka Miteva

PhD Student, Department of Information Technologies and Communications, University of National and World Economy, e-mail: gmiteva@unwe.bg

Aleksandrina Murdzheva

Assoc. Prof., Department of Information Technologies and Communications, University of National and World Economy, e-mail: amurdjeva@unwe.bg

Abstract

As data is becoming crucial for the efficient functioning of any organization, properly preparing it for processing is also getting increasingly important. This article presents an outline of different data preparation techniques, which can be defined in the context of machine learning. Overview of the techniques in combination of algorithms and their specific requirements for the data they can work with can serve as basis of the interesting research task of finding automated solutions that not only enable the use of software solutions, but also build complete solutions that support the detection of the potential of the data and application of these techniques that can be automatically identified.

Automating data preparation might be one of the steps which ensures that the machine learning process becomes quicker and more accessible. Using data preparation tools is a way to ensure more reliable and accurate data. This article aims to create an overview of existing data preparation tools and platforms. Different aspects of these platforms are considered, including data source compatibility, the data preparation techniques it includes, possibilities for integration, data security etc.

Key words: AI, data preparation, data collection, data cleaning, data transformation, feature engineering, data labeling, data preparation platforms, automation

1.Data preparation

Data preparation is the process of transforming data in a way that is suitable for machine learning and other processing purposes. This is one of the main prerequisites for achieving precise results when dealing with machine learning and AI. Incredible amounts of data are being collected constantly, but a lot of it contains inaccuracies, missing values, and outliers. Furthermore, nowadays a lot of the data comes from many different sources. Finding a way to transform the data into a coherent and usable set is crucial for the machine learning process.

Some of the steps of the process usually include collection, integration, cleaning, transformation, etc. It may also include data labeling, validation, visualization, data enrichment, feature engineering.

Data collection

As previously mentioned, data is crucial to achieving precise results. Some important features of data should be taken into consideration when choosing a data set to work with:

- **Size.** The size of the dataset is one of the main characteristics, and in general it is considered that a simpler model which is used on a larger data set might do better than using a more complex machine learning method on a small dataset.
- **Reliability.** Making predictions which can later be used is impossible without data that is reliable. Unreliable data might include label errors, noise, or data that is not entirely relevant to the project at hand. Other possible problems are duplicates, omitted values etc.
- **Feature representation.** Some values in the data set might need to be normalized. Handling outliers might also be needed.

Possible issues that might occur with data collection also include different biases. Reporting bias, for example occurs when a data set only covers a fraction of the actual data. It can include citation bias, language bias, location bias, outcome reporting bias etc. Selection bias is especially important when talking about data collection – it occurs when proper randomization is not achieved during the data collection phase.

Data might come from many different sources. It might relate to certain events, or it might be just a snapshot of information. In general, data can be derived from sensors that collect information, tracking social media, using surveys and forms, focus groups, transactions etc.

Data integration

Data integration is the process of combining data collected through different means. It usually includes ingestion and transformation. The aim of data integration is to store the transformed data in a data warehouse, data lake, or data lakehouse.

According to Qlik, there are five approaches to data integration. The following illustration shows where they sit in the data management process:

The first part of the illustration shows that the data management process consists of data sources, data processing, data storage and analytics/apps. Data processing includes:

- **ETL** – converts raw data through three steps – extracting, transforming, and loading. The data is transformed in a staging area and then loaded, usually in a data warehouse.
- **ELT** – with the ELT process, data is first loaded and then transformed when it is already in the target system.
- **Data streaming** – the data streaming process consists of constantly moving data in real-time from source to target system.
- **Application integration** – application integration allows different applications to sync data between each other.
- **Data virtualization** – similar to streaming, in the sense that it also delivers data in real time. The difference is that data is delivered only when requested by a user or application.

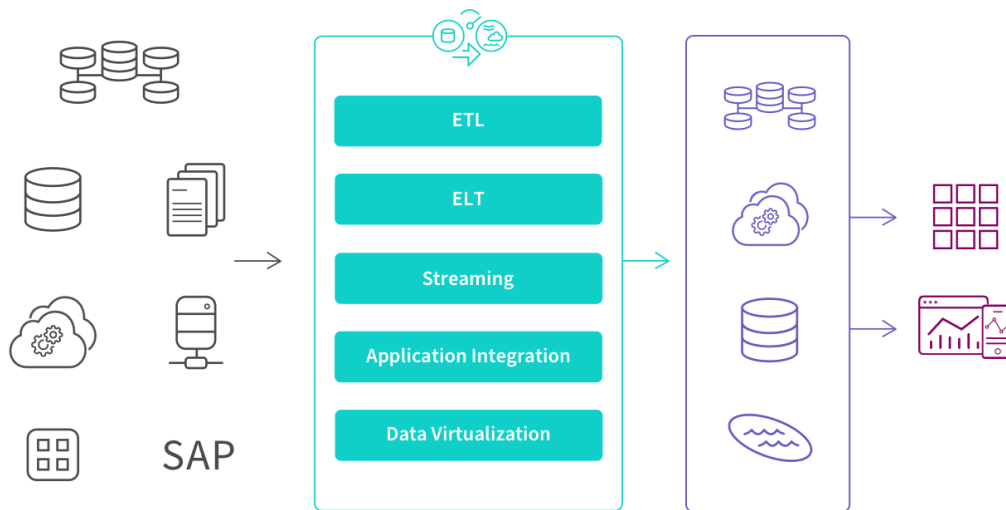


Fig. 1

Data cleaning

Data cleaning is the process of removing incorrect, duplicate or incomplete data from a data set. As mentioned before, there is no exact way of cleaning data – it depends on the needs of the organization and the specifics of the data. Nevertheless, there are a few main steps which are usually involved in the data cleaning process:

- Removing duplicate data

Removing duplicate data can begin at data collection and transformation. When combining data from different sources, duplicate data can be hard to avoid. Removing repeated values is crucial to receiving accurate results.

- Removing outliers

Removing data that does not fit within the analyzed data when it's the result of improper data entry for example, this will improve the results from the analysis. However, the existence of an outlier does not mean it is incorrect data that needs to be removed.

- Handling missing data

Many algorithms do not accept missing data, so dealing with it is crucial. Some options of handling missing data are inputting missing data based on other observations or dropping observations with missing values.

- Validation

Validating data includes making sure the data makes sense and confirming that it follows the appropriate rules for its field.

Data transformation

Data transformation is a crucial step in the data preparation process for AI and machine learning applications. It involves converting raw data into a format that is suitable for analysis, modeling, and training machine learning algorithms. The goal of data transformation is to improve the quality and relevance of the data for the specific tasks at hand. Here are some key aspects of data transformation in the context of AI:

Scaling

Scaling in AI data preparation refers to the process of adjusting the scale or range of numerical features in a dataset to make them more suitable for machine learning algorithms. Many machine learning models are sensitive to the scale of input features, and scaling is applied to ensure that all features contribute equally to the model training process. Scaling helps in improving the convergence speed and stability of certain optimization algorithms used in machine learning.

Normalization and Standardization

Normalization and Standardization is an approach to bring numerical features to a common scale. This ensures that features with different units or scales do not disproportionately influence the model. Often normalization and standardization are referred as main method for achieving data scaling.

Encoding Categorical Variables

Encoding Categorical Variables is an approach of converting categorical variables into a numerical format that can be used by machine learning algorithms. This technique is used because machine learning algorithms typically work with numerical data. In other words, encoding stands for converting categorical data into a numerical format. Common techniques include one-hot encoding, label encoding, or ordinal encoding.

Handling Outliers

An outlier is an observation in a data set that is distant from other observations. Outliers can significantly impact the performance of machine learning models. Identification of such values is very important. Various methods like trimming, winsorizing, or transformation can be applied to handle outliers.

Discretization

Discretization is an approach to preparing data for machine learning that allows transforming continuous variables, such as time, temperature, or weight, into discrete ones. Consider a dataset that contains information about people's height. The height of each person can be measured as a continuous variable in feet or centimeters. However, for certain ML algorithms, it might be necessary to discretize this data into categories, say, "short", "medium", and "tall". This is exactly what discretization does. It helps simplify the training dataset and reduce the complexity of the problem. Common approaches to discretization span clustering-based and decision-tree-based discretization.

Dimensionality reduction

Dimensionality reduction stands for limiting the number of features or variables in a dataset and only preserving the information relevant for solving the problem. This can be particularly useful when dealing with high-dimensional data. Techniques like principal component analysis (PCA), linear discriminant analysis (LDA) or t-SNE (t-distributed stochastic neighbor embedding) can be used.

Log transformation

Another way of preparing data for machine learning, log transformation, refers to applying a logarithmic function to the values of a variable in a dataset. It is often used when the training data is highly skewed or has a large range of values. Applying a logarithmic function can help make the distribution of data more symmetric.

Data Imbalance

Data imbalance refers to a situation in a classification problem where the distribution of classes in the training dataset is not equal. In other words, one or more classes have significantly fewer instances compared to other classes. The approach of handling imbalanced classes in classification tasks is very important. Techniques such as oversampling, undersampling, or using different sampling strategies can be applied to address class imbalance.

Handling Skewed Distributions

A skewed distribution, also known as asymmetry, is a statistical term that describes the shape of a probability distribution. In a skewed distribution, the data points cluster more toward one side of the distribution than the other, creating a tail that extends in one direction. This approach refers application of techniques logarithmic or square root transformations to help make the data more symmetric and improve model performance.

Temporal Data Processing

For time-series data, features like date and time may be decomposed into components such as year, month, day, or hour. Lag features or rolling statistics can also be created to capture temporal patterns.

Feature engineering

The scope of data transformation also included feature engineering. Feature engineering is more than a transformation technique or a step in the process of preparing data for machine learning. Feature engineering involves a combination of statistical, mathematical, and computational techniques, including the use of ML models, to create features that capture the most relevant information in the data.

Model features are the inputs used by machine learning during the training process and they are significant to the accuracy of the process. Feature engineering is a key part of preparing data for machine learning. It consists of adding and constructing variables to a dataset to better understand the data and achieve a better performance of the machine learning model. It is important to have knowledge of the business problem, as well as the data source, to achieve more accurate results.

Feature engineering includes a few main processes:

- Feature creation – creating new variables.
- Transformations – transforming features from one representation to another.
- Feature extraction – selecting features from a dataset which will help find more meaningful information.
- Exploratory data analysis (EDA)– exploring the properties of data to create new hypotheses, find patterns, etc.

Feature engineering, if done correctly, can optimize a dataset to contain more important factors which affect the business model.

Data labeling

In the context of machine learning, data labeling is the process of identifying raw data and adding meaningful labels to it, in order for a machine learning algorithm to be able to learn from it. For example, labels might indicate what a photo contains, what words are used in a recording etc. Data labeling usually consists of users making judgements about unlabeled data. The algorithm then uses the labels provided by humans to learn in a process called training.

2.Data preparation platforms

According to Gartner, data preparation tools help accelerate the delivery of curated data, allowing the users to find anomalies in their data. Some popular data preparation platforms include:

Alteryx

Alteryx is a data analytics platform that provides automated data preparation and analytics with the help of machine learning and AI. The platform allows the user to integrate different data sources, and then build data pipelines to extract and load different sources into targets such as cloud data warehouses or cloud data lakes. Alteryx provides the user with the ability to automate, document, share and scale the data preparation process with low-code and no-code tools.

Talend

Talend is described as an end-to-end platform which combines data integration, data quality, and data governance in a low-code tool. It can be deployed on premises, as well as in the cloud, multi-cloud, or hybrid cloud. According to Gartner, Talend is focused on improved data quality, with the platform allowing the user to measure the validity and completeness of datasets. The platform provides a browser based self service tool which helps shorten the time it takes to clean data. It also lets the users automate data preparation and turn it into a reusable process, using data from different sources.

Power BI

Power BI provides self-service data preparation capabilities. It uses dataflows to ingest, cleanse, transform, integrate and enrich data from different sources – on-premise ones, as well as cloud based ones, which includes Dynamics 365. These dataflows store data in Azure Data Lakes, which means users can utilize Azure Machine Learning, Azure Databricks, and Azure SQL Datawarehouse for advanced analytics. Nevertheless, Power BI seems to be mainly a data visualization tool, with some users suggesting it is better to prepare data as much as possible prior to loading it into Power BI.

RapidMiner

RapidMiner is an end-to-end data science platform which allows users to create machine learning pipelines. RapidMiner allows for scaling from on-premises servers to cloud implementations. The platform allows users to extract and transform data from many sources, which include PDFs, spreadsheets and text files. RapidMiner works with both structured and unstructured data, with the ability to automate data transformation tasks.

DataRobot

DataRobot AI Platform is a full AI lifecycle platform used for predictive and generative AI. It interfaces with DataRobot Data Prep to assist with data preparation. Data Prep allows the user to explore the data, with the ability to clean, combine and shape it in a proper format for machine learning. It offers a user interface which visualizes the data in a spreadsheet style, requiring no coding.

Informatica Data Prep

Informatica Data Prep provides cloud data preparation capabilities with low-code/no-code on cloud data warehouses. It provides a data catalog with metadata which can be useful in understanding data. It also provides data profiling capabilities to find anomalies, outliers etc. The platform is also useful for data transformation and enforcing data governance and compliance policies.

3. Conclusion

Undoubtedly, the preparation of data for use in AI and Machine Learning is of utmost importance. It demonstrates the serious number of techniques that are defined and the coverage of different aspects of working with data. Clear definitions and desired end results are a good basis for seeking automation of data preparation techniques, which in the context of large volumes of data is already a serious necessity. The technological world offers a significant set of interesting solutions for automating individual data preparation tasks or of the entire processes. The wide variety of algorithms and their specific requirements for the data they can work with pose the interesting research task of finding automated solutions that not only enable the use of software solutions, but also build complete solutions that support the detection of the potential of the data and application of these techniques that can be automatically identified.

References

1. Stefanov, G. (2019). Analysis of Cloud based ETL in the Era of IoT and Big Data. In Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education. 198-202. ICAICTSEE.
2. E. Karkalikova, A. Murdjeva, Organization of Data in Data Lake – Real-Life Practice, 11th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education ICAICTSEE– 2021, Sofia, Bulgaria.
3. Genka Miteva, Alexandrina Murdjeva CHARACTERISTICS AND PREPARATION OF DATASETS FOR MACHINE LEARNING ALGORITHMS, 11th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education ICAICTSEE– 2021, Sofia, Bulgaria.
4. Geno Stefanov, ANALYSIS OF SERVERLESS CLOUD DATA WAREHOUSE SOLUTIONS, 11th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education ICAICTSEE– 2021, Sofia, Bulgaria.
5. Delchev, D., Lazarova, V., Big Data Analysis Architecture, Economic Alternatives, 2021, Issue 2, pp. 315-328
6. Verdonck, T., Baesens, B., Óskarsdóttir, M. et al. Special issue on feature engineering editorial. Mach Learn (2021). <https://doi.org/10.1007/s10994-021-06042-2>
7. Edward Baumann, Charles Hsu, Hayley Buba, Taylor Cox, An Introductory Approach to Time-Series Data Preparation and Analysis, Annual Conference of the PHM Society, October 2023, DOI: 10.36001/phmconf.2023.v15i1.3561
8. Mohamad Fariq Rahmat, Zed Zulkafli, Asnor Juraiza Ishak and others, Supervised feature selection using principal component analysis, Knowledge and Information Systems, November 2023, DOI: 10.1007/s10115-023-01993-5
9. Lamia AbedNoor Muhammed, Role of data normalization in k-means algorithm results, • AL-KADHUM 2ND INTERNATIONAL CONFERENCE ON MODERN APPLICATIONS OF INFORMATION AND COMMUNICATION TECHNOLOGY, March 2023, DOI: 10.1063/5.0119267
10. What is Data Preparation? Accessed: Oct. 29, 2023. [Online]. Available: <https://aws.amazon.com/what-is/data-preparation/>

11. The Size and Quality of a Data Set; Google [Online]. Available: <https://developers.google.com/machine-learning/data-prep/construct/collect/data-size-quality>
12. What is Data Integration? Talend, [Online], Available: <https://www.talend.com/resources/what-is-data-integration/>
13. Data Integration, What it is, why it matters, and best practices. This guide provides definitions, examples and practical advice to help you understand the topic of data integration, QLIK, [Online], Available <https://www.qlik.com/us/data-integration>
14. <https://www.tableau.com/learn/articles/what-is-data-cleaning>
15. What is Data Labeling?; Amazon, [Online]. Available: <https://aws.amazon.com/sagemaker/data-labeling/what-is-data-labeling/>
16. What Is Data Preparation? Cut through the noise with data ready for analytics, ML and more; Informatica, [Online]. Available: <https://www.informatica.com/resources/articles/what-is-data-preparation.html#4>
17. What is Feature Engineering?; Amazon, [Online]. Available <https://aws.amazon.com/what-is/feature-engineering/>
18. Feature Engineering; What is Feature Engineering for Machine Learning, Data Robot, [Online]. Available :<https://www.datarobot.com/wiki/feature-engineering/>
19. What is Feature Engineering — Importance, Tools and Techniques for Machine Learning; Harshil Patel, Towards Data Science, Aug 30, 2021, <https://towardsdatascience.com/what-is-feature-engineering-importance-tools-and-techniques-for-machine-learning-2080b0269f10>

РАЗШИРЯВАНЕ ВЪЗМОЖНОСТИТЕ НА ИНФОРМАЦИОННА СИСТЕМА ЗА РАБОТА С ГЛАС

Веска Михова

Гл. ас. д-р в катедра „Информационни технологии и комуникации“, УНСС
e-mail: vmihova@unwe.bg

Резюме

Този научен доклад разглежда възможностите и перспективите за разширяване на функционалността на информационни системи, които използват гласов интерфейс. Гласовите технологии са доказали своята значимост в съвременния свят, предоставяйки удобен начин за комуникация между хората и машините. Докладът обръща внимание на иновативни решения и технологии, които могат да разширят приложенията и възможностите на такива системи.

Abstract

This scientific report examines the possibilities and prospects for expanding the functionality of information systems that use a voice interface. Voice technologies have proven their relevance in the modern world, providing a convenient way for humans and machines to communicate. The report highlights innovative solutions and technologies that can extend the applications and capabilities of such systems.

Ключови думи: гласови технологии, гласово разпознаване, гласов синтез

JEL: O30, C88

Въведение

Разширяването на възможностите на информационна система за работа с глас е важна стъпка в напредъка на технологичния свят и подобряването на взаимодействието между хората и компютрите. Включването на гласови интерфейси и обработка на гласова информация може да увеличи удобството и ефективността на множество приложения и системи.

Системите за работа с глас са внушителен пример за интелигентни информационни системи, които се основават на гласови команди и разпознаване на гласова информация. Те не само улесняват ежедневието на потребителите, но също така имат голям потенциал в разнообразни области като медицината, образованието, бизнеса и социалната интеграция на хора с увреждания. Този доклад изследва настоящите и бъдещи възможности за разширяване на функционалността на тези системи.

Предизвикателство в тази област, което е извън обхвата на доклада, е важната тема свързана със сигурност и поверителност на информацията. Трябва да се обърне особено внимание на сигурността на гласовата информация и личните данни на потребителите, особено при съхранението и обработката на чувствителни гласови записи.

Гласово разпознаване и синтез

Гласово разпознаване

Гласовото разпознаване е технология, която позволява компютърните системи да разпознават и интерпретират човешка реч. Това представлява процес на анализиране и преобразуване на гласовите сигнали в текст или друг вид машинно обработваем изход.

Използването на технологии за разпознаване на реч може да позволи на информационната система да преобразува гласови команди или гласови записи в текст, който след това може да бъде обработван и анализиран от информационни системи ([1], [2]).

Изследванията в областта на машинното обучение са позволили на системите за разпознаване на реч да стават по-точни и адаптивни. Интегрирането на напреднали алгоритми и модели може да увеличи точността при разпознаването на глас. Разширяването на способностите за работа с различни езици може да повиши достъпността и употребата на гласовите системи по света.

Гласовото разпознаване е напреднала технология със значителен потенциал за решаване на различни предизвикателства и за предоставяне на редица ползи в различни области. Някои от основните цели и приложения на гласовото разпознаване включват:

- Подобряване на потребителското изживяване;
- Автоматизиране на процеси и контрол на устройства чрез гласови команди;
- Облекчаване на функционални ограничения на хора със зрителни, моторни или други увреждания;
- Улесняване транскрипцията на медицински записи и подпомагане здравеопазването.
- Подобряване мултиезичната комуникация и взаимодействие с технологията.
- Използване за биометрична идентификация и аутентикация на потребители.
- Използване в образователни приложения, които улесняват ученето и комуникацията.
- Гласовото разпознаване стимулира разработката на нови услуги и приложения, които подобряват ежедневието и бизнес процесите.

Освен всичко това, гласовото разпознаване продължава да е фокус в иновациите в областта на изкуствения интелект и машинното обучение, като целта е постигане на по-голяма точност и ефективност в разпознаването на различни езици и диалекти.

Гласов синтез

Гласовият синтез е технология, която се използва за генериране на изкуствено създадени гласови сигнали или реч посредством компютърни програми и системи. Този процес включва преобразуване на текст или друга форма на символна информация в аудио формат, който след това може да бъде изговорен или чул от хора чрез високоговорители или слушалки.

Гласовият синтез имитира човешкия глас и може да се използва за множество цели, включително:

- Предоставяне на аудио книги и текстове в говорима форма за лица със зрителни увреждания или с дислексия.
- Разработка на гласови асистенти и виртуални говорители, които да предоставят информация и да извършват задачи посредством комуникация с потребителите.
- Гласови обявления и реклами.
- Интерактивни гласови системи за обучение и образование.
- Аудио-синтез на данни и информация за потребление на медицински съвети, новинарски приложения и др.

Съвременните системи за гласов синтез се базират на компютърни алгоритми и изкуствени невронни мрежи, които могат да имитират различни интонации, тонове и емоции в човешкия глас, правейки ги по-натурални и приятни за слушателя.

Преобразуването на текстова информация в аудио в една система може да бъде реализирано посредством гласов синтезатор с API за гласов синтез. Интегрирането на API за гласов синтез може да предостави информация на потребителите на една система чрез глас. Синтезът на реч може да бъде използван, за да позволи на системата да отговаря на потребителите с гласови отговори. Този процес включва преобразуване на текст в артикулирана реч.

Вграждането на емоционален израз в гласовия синтез може да направи комуникацията по-човешка и емоционално адекватна. Комуникацията между потребителите и системите може да бъде улеснена и чрез възможността на потребителите да избират различни гласове за системата, както и персонализация на гласовите настройки, тон и скорост на речта, които отговарят на техните предпочитания ([3], [4])

Инструменти и библиотеки за гласово разпознаване

Гласовото разпознаване е процес на трансформиране на изречение, произнесено на човешки глас, в текст. Това е ключова технология, използвана в гласови асистенти, системи за навигация, софтуер за транскрипция и други приложения.

Водещите технологии и инструменти за гласово разпознаване притежават характеристики като:

- Висока точност при разпознаването на речта, както и поддръжка на различни езици, акценти, а също така и на диалекти.
- Позволяват обработка на аудио данни в реално време и поддържат паралелна обработка за бърз отговор.
- Разполагат с вградени професионални модели за различни области като медицина, финанси и други.
- Позволяват създаването на персонализирани модели, които се обучават да разпознават специфични термини и изрази.
- Различни ценови планове, включително безплатен план с ограничен брой безплатни заявки в месец и тарифи за по-големи обеми.

Водещи инструменти и библиотеки за гласово разпознаване

Google Cloud Speech-to-Text:

- Платформа за гласово разпознаване, предоставяща висока точност и поддръжка на различни езици. Интегрира се с облачните услуги на Google.
- Предлага безплатен тестов период с кредит за потребление на услугите на Google Cloud. Може да се използва безплатен период, след което се налага заплащане според тарифите на Google Cloud.

Microsoft Azure Speech Services:

- Облачна услуга за гласово разпознаване от Microsoft. Предлага възможности за разпознаване на реч, конверсия на реч в текст и обратно, както и аудио анализ.
- Лесно може да се интегрира с други услуги в облачната платформа на Azure, като Azure Cognitive Services и Azure Logic Apps.
- Azure Speech SDK предоставя безплатен план с ограничен брой заявки в месец. Има и допълнителни тарифи за по-големи обеми.

IBM Watson Speech to Text:

- Система за гласово разпознаване, използваща технологии на изкуствения интелект. Предоставя API за интеграция в различни приложения.
- IBM Watson предлага безплатен план с ограничение на броя на заявките в месеца. Необходима е регистрация и използване на ключове за достъп.

Amazon Transcribe:

- Услуга за автоматично гласово разпознаване от Amazon Web Services. Поддържа различни езици и предоставя възможности за транскрипция на аудиофайлове.

Dragon NaturallySpeaking:

- Продукт на Nuance Communications, предлагащ точно и бързо гласово разпознаване. Използва се в медицински, правни и други професионални среди.

Някои компании предоставят безплатни API услуги за гласово разпознаване с определени ограничения. Важно е да се отбележи, че безплатните планове обикновено имат ограничен брой заявки в месеца или предоставят базови функционалности. Няколко такива безплатни API услуги:

Wit.ai:

- Wit.ai е платформа с отворен код, управлявана от Facebook, за обработка на естествен език (NLP), която предоставя API за гласово разпознаване и обработка на текст.
- Предоставя безплатен план с ограничение на броя на заявките и обема на обучение.

PocketSphinx:

- PocketSphinx е библиотека с отворен код за гласово разпознаване, предназначена за вградени системи и устройства с ограничени ресурси. Това е част от проекта CMU Sphinx (Carnegie Mellon University Sphinx), който се занимава с разработването на технологии за автоматично разпознаване на реч.
- Проектиран да бъде лек и ефективен, подходящ за вграждане в устройства с ограничени ресурси, като например мобилни устройства и вградени системи. Подходящ за вграждане в умни устройства, които използват гласови команди за управление.
- Позволява офлайн работа, като не изисква постоянна връзка с интернет.

CMU Sphinx (Sphinx-4):

- Система за гласово разпознаване, разработена от Карнеги-Мелън Университет. Sphinx-4 е с отворен код и предоставя гъвкавост, но разработчиците трябва да имат познания в Java за оптимално използване. Предоставя инструменти за изследователи и разработчици.

SpeechRecognition (Python библиотека):

- Проста и лесна за използване библиотека за гласово разпознаване в програмен език Python.
- Тази библиотека използва различни back-end двигатели за гласово разпознаване, като Google Web Speech API, Sphinx, Microsoft Bing Voice Recognition и други.
- SpeechRecognition се предоставя под лиценза Apache 2.0, който позволява свободно използване и промяна на кода.

Kaldi:

- Отворена платформа за гласово разпознаване, разработена от университета в Джонс Хопкинс. с акцент върху иновации и ефективност в обучението на модели.

Mozilla DeepSpeech:

- DeepSpeech на Mozilla е проект с отворен код, насочен към разработването на система за автоматично разпознаване на реч (ASR).
- Използва DeepSpeech модел, базиран на рекурентни невронни мрежи (RNN), което позволява на системата да улавя сложни зависимости в речта.
- Тази система е разработена от Mozilla и използва DeepSpeech архитектурата, която е основана на рекурентни невронни мрежи (RNN) и е направена с отворен код с цел предоставяне на свободно налични средства за гласово разпознаване.
- Може да се използва с обучен модел или да се обучи собствен модел.

Тези технологии и инструменти предоставят широк спектър от възможности за разработка на приложения, които използват гласово разпознаване, независимо дали става въпрос за създаване на гласов асистент, транскрипция на аудио, или други приложения ([5], [6]).

Инструменти и библиотеки за гласов синтез

Гласовият синтез е процесът на генериране на изкуствени гласове или реч чрез компютърни програми и системи ([7], [8]). Тази технология има различни приложения, включително гласови асистенти, аудио книги, системи за навигация и други. Някои от водещите технологии и инструменти за гласов синтез са:

Google Text-to-Speech:

- Облачна услуга от Google, която преобразува текст в гласов сигнал. Предлага различни гласове и поддържа многоезичие.

Amazon Polly:

- Услуга за гласов синтез от Amazon Web Services. Предоставя голям набор от гласове и опции за персонализация.

Microsoft Azure Text to Speech:

- Облачна услуга, предоставена от Microsoft Azure, която преобразува текст в гласов сигнал. Поддържа няколко езика и гласове.

IBM Watson Text to Speech:

- Система, използваща технологии на изкуствения интелект, която преобразува текст в реч. Предоставя възможности за персонализация на гласовете.

Nuance Communications:

- Компания, която предлага различни продукти за гласов синтез, включително Dragon NaturallySpeaking и други инструменти за гласова комуникация.

Festival Speech Synthesis System:

- Отворен проект, който предоставя гласов синтез в реално време. Често се използва в научни и образователни приложения.

CereProc:

- Компания, специализирана в гласов синтез, предлагаща персонализирани гласове и услуги за клиенти.

Neospeech:

- Компания, която предлага системи за гласов синтез със синтетични гласове, които са близки до човешките.

ResponsiveVoice:

- Уеб базиран инструмент за гласов синтез, който може да бъде лесно интегриран в уеб сайтове и приложения.

eSpeak:

- Компактен и отворен инструмент за гласов синтез, който се фокусира върху простота и ефективност.

Гласовият синтез се развива бързо, и тези технологии предоставят широк спектър от възможности за приложения в различни области, включително виртуални асистенти, образование, развлечения и много други ([9]).

Заклучение

Разширяването на възможностите на информационна система за работа с глас може да подобри потребителския опит, да увеличи ефективността и да създаде нови възможности за бизнеса.

Технологичните иновации, обсъдени в настоящия доклад, предоставят основа за бъдещето на гласовите системи и тяхната роля в съвременния свят.

Този доклад е само начало на разглеждането на темата за системи за работа с глас и може да послужи като отправна точка за допълнителни изследвания и иновации в тази област.

Литературни източници

1. Neha Jain, Somya Rastogi, 2019, Speech recognition systems - a comprehensive study of concepts and mechanism, Acta Informatica Malaysia (AIM), <https://actainformaticamalaysia.com/archives/AIM/1aim2019/1aim2019-01-03.pdf>
2. Dong Yu , Li Deng, 2015, Automatic Speech Recognition A Deep Learning Approach, Springer, <https://link.springer.com/book/10.1007/978-1-4471-5779-3>
3. Sabato Marco Siniscalchi, Torbjørn Svendsen, Chin-Hui Lee, 2014, An artificial neural network approach to automatic speech processing, Neurocomputing Volume 140, 22 September 2014, Pages 326-338.
4. Andreas M. Klein, Kristina Kölln, Jana Deutschländer & Maria Rauschenberger, 2023, Design and Evaluation of Voice User Interfaces: What Should One Consider?
5. International Conference on Human-Computer Interaction, HCII 2023: Design, Operation and Evaluation of Mobile Communications pp 167–190.
6. Plamen Milev, "Approach for Analysis and Comparison of Search Query Results in Web Publications, 11th International conference on application of information and communication technology and statistics in economy and education (ICAICTSEE – 2021), November 25-26th, 2021, UNWE, Sofia, Bulgaria.
7. Митко Радоев, 2021, Съвременни тенденции в развитието на базите от данни, Икономически алтернативи, <https://www.unwe.bg/doi/alternativi/2021.1/ISA.2021.1.01.pdf>
8. Monika Tsaneva, 2019, A practical approach for integrating heterogeneous systems - Бизнес управление.
9. Geno Stefanov, Maria Marzovanova, Building IoT Solution for Better University Using IBM Watson IoT Platform 2017, ICAICTSEE - 2017, <http://icaictsee.unwe.bg/past-conferences/ICAICTSEE-2017.pdf>
10. E. Karkalikova, A. Murdjeva, Organization of Data in Data Lake – Real-Life Practice, 11th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education ICAICTSEE– 2021, Sofia, Bulgaria.

АРХИТЕКТУРА НА ИНФОРМАЦИОННА СИСТЕМА ЗА РЕАЛНИ И СИМУЛИРАНИ СТОЙНОСТИ НА ИОТ УСТРОЙСТВА

Александър Танев

Хон.ас в катедра „Информационни технологии и комуникации“, УНСС
e-mail: atanev@unwe.bg

Резюме

Този доклад представя изследване на популярни облачни платформи за обработка на данни от IoT устройства и проект на архитектура за информационна система, предназначена за управление на реални и симулирани стойности, генерирани от IoT устройства. IoT играе ключова роля в съвременните технологии, като сензорите и устройствата събират големи обеми данни. Важно е да се създаде ефективна информационна система, която да обработва, управлява и анализира този поток от данни. Представят се някои от предизвикателствата пред подобни системи, както и възможните подходи за решаването им.

Ключови думи: информационна система, IoT устройства

JEL: L86

Въведение

Информационните системи с IoT (Интернет на нещата) устройства са много актуални и все повече навлизат в нашето ежедневие. Причината за толкова бързото им разпространение е, че хората обръщат все по-голямо внимание на качеството на живот, а съвременните технологии осигуряват големи възможности за неговото подобряване.

Ще бъдат посочени само няколко примера, при които системите, свързани с вградени IoT устройства повдигат качеството на живот на много високо ниво.

- На първо място е човешкото здраве и живот. IoT устройства прикрепени към човек и свързани със съответната система, могат да наблюдават и поддържат човешкия тонус и здраве, дори да управляват такива заболявания като диабет, високо кръвно налягане, тиреотоксикоза и др.
- На следващо място сред примерите, при които системите, свързани с вградени IoT устройства повдигат на много високо ниво качеството на живот може би е домът. Хората могат да инсталират в домовете си, автоматизирани печки, хладилници и пр. домашни уреди, както и системи за сигурност или виртуални асистенти.
- В сферата на обслужването. IoT устройства могат да се инсталират в МОЛ, магазини, банки, ресторанти, за да улеснят плащането, като се даде възможност на хората сами да извършват плащане или други дейности, без задължителното наличие на посредник за плащане (касиер).

- Офиси. IoT устройства в офисите могат да спомогнат за управление съхранението на енергия или защита на сградите.
- Производството. IoT устройства в различни производствени фирми и организации могат да способстват за оптимизиране използването на оборудването и инвентара, да облекчат човешкия труд чрез включване на агрегати и пр.
- Превозни средства и градска среда. IoT устройствата могат да помогнат с поддръжка, базирана на състоянието, проектиране, базирано на употребата на автомобили и камиони, кораби, самолети и влакове, железопътни линии, автономни превозни средства или навигация за полети. IoT приложенията могат да включват маршрутизиране в реално време, свързана навигация или проследяване на пратки.

В този доклад се разглеждат популярни облачни услуги за обработка на данни от IoT устройства, ключови аспекти на архитектура на информационна система за реални и симулирани стойности на IoT устройства, необходимост от комуникационни протоколи, сигурност, управление на данни и възможности за разширение. Предложената архитектура на информационна система представлява иновативен и цялостен подход за управление на данни от IoT устройства.

Облачни услуги за обработка на данни от IoT устройства

Облачните услуги за обработка на данни от IoT устройства предоставят платформи за съхранение, анализ и управление на големи обеми от данни, генерирани от различни IoT устройства. Тези услуги често се използват за обработка на данни в реално време, мащабируемост и централизиран достъп до информацията.

Облачните платформи предоставят възможности за бърза и ефективна обработка на данни в реално време, което е от съществено значение за приложения, които изискват бърз отговор, като например интелигентни системи за управление на трафика, мониторинг на здравното състояние и др..

Облачните услуги позволяват лесно мащабиране на ресурсите в зависимост от обема на данните от IoT устройствата. Това е важно, тъй като броят на устройствата и обемът на данните от тях могат значително да се увеличат с времето.

Този тип платформи предоставят централизиран достъп и управление на данните, което улеснява администрирането и мониторинга на цялостната система. Този подход осигурява по-лесна интеграция на нови устройства и приложения. Тези услуги обикновено предлагат инструменти за анализ и визуализация на данни, които могат да помогнат в извличането на ценна информация от големите обеми данни, събирани от IoT устройствата.

Такива платформи обикновено включват вградени мерки за сигурност, като криптиране на данните, контрол на достъпа и системи за мониторинг на събитията, което е от съществено значение за защита на чувствителната информация. Позволяват лесно внедряване на нови функции и обновления, което е важно в динамичната област на IoT, където технологиите и изискванията постоянно се променят.

Популярни облачни платформи за обработка на данни от IoT устройства са Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform и други. Изборът на конкретна платформа зависи от нуждите, бизнес целите и предпочитанията на организацията.[1]

AWS

Amazon Web Services (AWS) предоставя обширен набор от услуги, които се използват за обработка на данни от устройства за Интернет на нещата. Някои от ключовите услуги на AWS, които се използват за този тип обработка са AWS IoT Core, Amazon Kinesis, Amazon S3 (Simple Storage Service), AWS Lambda, Amazon Redshift, AWS Glue и Amazon SageMaker.

AWS IoT Core е услуга, която позволява свързването и управлението на милиони IoT устройства в облака. Предоставя сигурност, мащабируемост и възможности за обмен на данни между устройствата и облака.

Amazon Kinesis е семейство от услуги за обработка на поточни данни. Включва Kinesis Data Streams за събиране и обработка на поточни данни, Kinesis Data Firehose за пренос на данни към различни AWS услуги, и Kinesis Data Analytics за анализ на данни в реално време.

Amazon S3 представлява обектно съхранение, което предоставя възможности за съхранение на данни от IoT устройства. Потребителите могат да използват различни класове на съхранение в зависимост от изискванията на тяхната архитектура.

AWS Lambda е сървърна услуга за изпълнение на код в реакция на събития. Може да се използва за обработка на събития от IoT устройства, без да се налага поддържане на инфраструктура.

Amazon Redshift е услуга за управление на аналитични данни, която е подходяща за анализ на големи данни от IoT устройства.

Услугата за извличане, трансформиране и зареждане (ETL) на данни, която предлагат Amazon и AWS Glue. Може да се използва за подготовка и обработка на данни от различни източници, включително IoT устройства.

Amazon SageMaker пък е услуга за машинно обучение, която предоставя инфраструктура за обучение и развитие на модели, които могат да се използват за анализ на данни от IoT устройства.

Azure

Microsoft Azure също предоставя обширен набор от услуги за обработка на данни от устройства за Интернет на нещата. Някои от ключовите услуги на Azure, които се използват в този контекст са Azure IoT Hub, Azure Stream Analytics, Azure Blob Storage, Azure Functions, Azure Time Series Insights, Azure Machine Learning, Azure Synapse Analytics.

Azure IoT Hub е услуга, която позволява свързването и управлението на IoT устройства. Предоставя сигурно предаване на данни и възможности за обмен на съобщения между устройства и облака.

Azure Stream Analytics служи за анализ на поточни данни, като поддържа обработка на данни в реално време. Използва се за извличане на знания от поточни данни от IoT устройства.

Чрез Azure Blob Storage е организирано обектно съхранение, подходящо също и за съхранение на данни от IoT устройства. Потребителите могат да използват различни така наречени „температурни“ класове за оптимизация на разходите.

Azure Functions е сървърна слуга, която позволява изпълнението на код в реакция на събития. Може да се използва за обработка на събития от IoT устройства без нужда от поддържане на инфраструктура.

Azure Time Series Insights е услуга за съхранение и анализ на времеви серии от данни. Тази услуга е подходяща за обработка на данни от IoT устройства, които генерират времево зависими данни.

Azure Machine Learning предоставя инструменти и услуги за създаване и развитие на модели за машинно обучение, които могат да бъдат използвани за анализ на данни от IoT устройства.

Azure Synapse Analytics е услуга за аналитични данни, която предоставя мащабируема инфраструктура за анализ на големи данни.

Google Cloud Platform (GCP)

Google Cloud Platform предоставя обширен набор от услуги, които могат да бъдат използвани за обработка на данни от устройства за Интернет на нещата (IoT). Обработката на данни от IoT устройства обикновено включва събиране, предаване, съхранение, анализ и визуализация на информацията от сензори и други IoT компоненти. Някои ключови услуги на Google, които могат да бъдат полезни в контекста на IoT са Cloud IoT Core, Cloud Pub/Sub, Cloud Dataflow, Cloud Storage, BigQuery, Cloud Machine Learning Engine, Cloud Vision API и Cloud Speech-to-Text API.

Cloud IoT Core е управляема услуга за свързване, управление и сигурно предаване на данни от IoT устройства към облака. Поддържа протоколи като MQTT и HTTP и предоставя възможности за сигурност и мащабируемост.

Cloud Pub/Sub е съобщителна услуга, която позволява на различни части от IoT системата да обменят данни в реално време. С данни, пратени към Pub/Sub, може да се свърже редица от услуги за обработка и анализ.

Cloud Dataflow е услуга за обработка на поточни и паралелни данни. Може да се използва за обработка и трансформация на данни от IoT устройства, както и за изграждане на поточни анализи.

Чрез Cloud Storage е организирано обектно съхранение, което може да се използва за съхранение на данни от IoT устройства в облака. Поддържа различни класове на съхранение, включително стандартно, архивно и горещо съхранение.

BigQuery е услуга за анализ на големи данни, която позволява бързо изпълнение на SQL-подобни заявки. Подходяща е за анализ на данни от IoT устройства и създаване на доклади.

Cloud Machine Learning Engine е услуга, с която може да се приложи машинно обучение към данни от IoT устройства, тази услуга предоставя инфраструктура за обучение и предсказване.

Cloud Vision API и Cloud Speech-to-Text API се използват при съдържание на изображения или аудио данни. Тези API предоставят възможности за анализ на визуалното и звуково съдържание.

Като цяло, всички от изброените доставчици на облачни услуги предоставят цялостен пакет от инструменти за обработка на данни от IoT устройства в облака, от началното събиране и предаване на данни до обработка, анализ и визуализация. [2][3]

Други облачни платформи

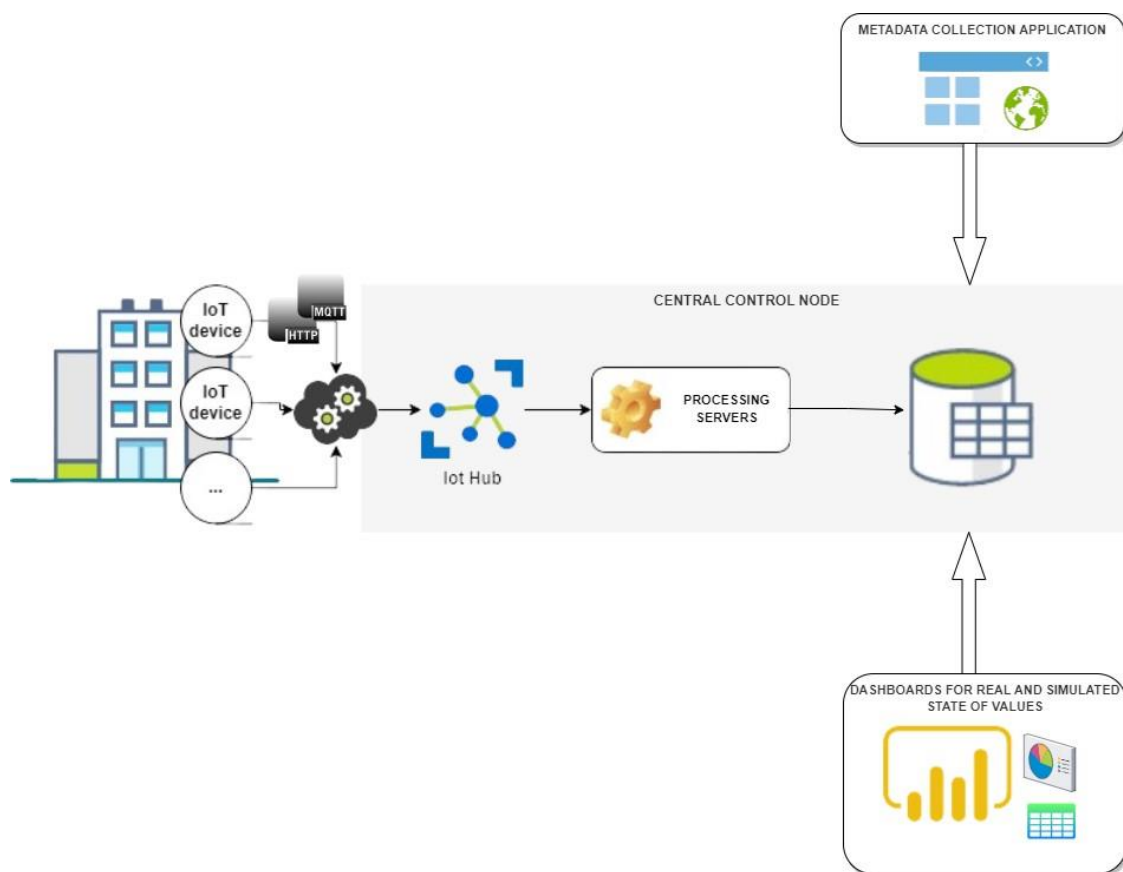
Съществуват и други облачни доставчици като IBM Cloud, Oracle Cloud и др., които предоставят собствени решения за IoT обработка на данни.

Архитектура на информационна система за реални и симулирани стойности на IoT устройства

Една архитектура на система за мониторинг на текущи и симулирани стойности от IoT устройства може да бъде построена с различни компоненти, които обработват, съхраняват и визуализират данните.

Архитектурата на предложената информационна система включва централен контролен възел, който управлява връзката с IoT устройствата и събира данни от тях. Системата поддържа възможността за връзка с реални устройства, както и със симулирани устройства, което улеснява разработката и тестването на приложения без реални хардуерни устройства.[4]

В резултат на анализа на данните системата предоставя възможност за извличане на ценна информация. Структурата ѝ е разработена с оглед на гъвкавост и мащабируемост, което позволява интеграция с различни видове IoT устройства и приложения. [5]



Фигура 1: Обща архитектура на информационна система за реални и симулирани стойности на IoT устройства.

- ✓ IoT Устройства - сензори и устройства, които измерват и събират данни като температура, влажност, налягане, осветеност и други.
- ✓ IoT Протоколи и Комуникация – протоколи като MQTT, CoAP или HTTP могат да се използват за комуникация между централния контролен възел и IoT устройствата.
- ✓ Централен контролен възел:
 - Основният елемент на архитектурата, който служи за координиране и управление на връзката с IoT устройствата.

- Брокер за съобщения - посредник, който приема данни от IoT устройствата и ги препраща към съответните компоненти в облака.
- База от данни - система за съхранение на данни, където се записват текущите и исторически стойности от IoT устройствата.
- Обработващи сървъри - сървъри, които извършват обработка на данните, прилагат правила за предупреждения или агрегират информация.
- ✓ Визуализация и интерфейс:
 - Уеб Интерфейс - графичен интерфейс за потребителите за въвеждане на зависимости и външни фактори - метаданни данни за информационна система за реални и симулирани стойности на IoT устройствата.
 - Мобилен достъп: Възможност за достъп и управление през мобилни устройства.
- ✓ Симулиране на данни:
 - Модули за Симулация - специализирани модули или инструменти за създаване на симулирани данни за тестване и разработка.
- ✓ Интеграция с аналитични инструменти:
 - Анализ и машинно обучение - интеграция с инструменти за анализ на данни и машинно обучение за предвиждане на бъдещи стойности или аномалии.

Заклучение

В този доклад беше представена накратко архитектура на информационна система, предназначена за управление на реални и симулирани стойности от IoT устройства. Основната цел на системата е ефективното събиране, обработка и управление на данните от разнообразни сензори и устройства. Архитектурата е изградена с оглед на гъвкавост и мащабируемост, което я прави приложима в различни области, където IoT технологиите играят ключова роля.

В заключение, предложената архитектура представлява цялостен и иновативен подход за управление на информацията от IoT устройства, като е основа за редица изследвания в тази посока. Предложената архитектура отговаря на текущите нужди за събиране на данни, гарантирайки готовност за бъдещите предизвикателства в областта на Интернет на нещата.

Цитирана литература

1. Aina'u Shehu Muhammed; Derya Ucuz; 2020; 8th International Symposium on Digital Forensics and Security (ISDFS); Comparison of the IoT Platform Vendors, Microsoft Azure, Amazon Web Services, and Google Cloud, from Users' Perspectives.
2. Stefan Forsström, Ulf Jennehag, 2017, Global Internet of Things Summit (GIoTS), A performance and cost evaluation of combining OPC-UA and Microsoft Azure IoT Hub into an industrial Internet-of-Things system.
3. Analysis of Serverless Cloud Data Warehouse Solutions, Geno Stefanov, 11th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE – 2021), November 25-26th, 2021, UNWE, Sofia, Bulgaria".

4. Plamen Milev, 2019, Opportunities for Presentation of Tag Cloud in Public Information Systems, 9th ICAICTSEE 2019 UNWE, Sofia, Bulgaria, October 24 – 26th.
5. Geno Stefanov, 2019, Analysis of Cloud based ETL in the Era of IoT and Big Data, 9th ICAICTSEE 2019 UNWE, Sofia, Bulgaria, October 24 – 26th.
6. S. Kouzmanov, M. Tsaneva. Impact of Business Problem Characteristics on the Architecture and Specification of Integration. Computer Science and Information Technology 4(1): 9-14, 2016, ISSN 2331-6063

УПРАВЛЕНИЕ НА ДИГИТАЛНАТА ТРАНСФОРМАЦИЯ – НОВИ ПОДХОДИ И РОЛЯ НА КОНСУЛТАНСКИТЕ КОМПАНИИ

Цветомир Цанов

Главен асистент, д-р. Катедра „Международни икономически отношения и бизнес“, Факултет „Международна икономика и политика“, УНСС, e-mail: tzanov@unwe.bg

Резюме

Ролята на консултантските компании за осъществяване на дигитална трансформация от различни организации привлича внимание не само в практико-приложен аспект. Същевременно навигирането на съответни промени, които касаят реконфигуриране на бизнес модела чрез използване на цифрови технологии, отнасящо се и до фирмената инфраструктура, работни процеси и хора, е обект на засилващ се интерес и в академичната литература. Целта на настоящия доклад е да бъде анализирано адаптирането на нови подходи за управление на дигиталната трансформация на фона на разрастващия се през последните години глобален пазар за предоставяне на такива консултантски услуги. Установява се префокусиране към интегрирането на новите технологии в дейността, за сметка на доскоро смятани за приоритетни стратегически решения.

Ключови думи: дигитална трансформация, управление на дигиталната трансформация, глобален пазар на консултантски услуги в областта на дигиталната трансформация

MANAGING DIGITAL TRANSFORMATION – NEW APPROACHES AND THE ROLE OF CONSULTING FIRMS

Tzvetomir Tzanov

Abstract

The role of consulting firms for implementing digital transformation by various organizations has been attracting attention not only in a practical-applied framework. At the same time, navigating relevant changes through the use of digital technologies, that are affecting the business model reconfiguration, along with the firm's infrastructure, processes and people, is the subject of growing interest in the academic literature as well. The aim of this paper is to analyze the adaptation of new approaches related to managing digital transformation in line with the growing market of providing such consultancy services worldwide. A major finding of this study is that consulting firms are refocusing their approaches towards the integration of new technologies instead of prioritizing related strategic issues.

Key words: digital transformation, managing digital transformation, global digital transformation consulting market

JEL: M15, L20

Увод

Реализирането на трансформационния потенциал от интегрирането на новите цифрови технологии в дейността е резултат от различни управленски решения, които касаят фирмени функции и структури, работни процеси и хора, но и външни интеракции с контрагенти, обуславящо генериране на повече стойност към клиента. Използването на технологиите в тази връзка следва да осигури радикална промяна в „обхвата на бизнеса“ на предприятията (*reach of enterprises*) и съществено подобряване на резултатите от дейността – по смисъла на една от основополагащите дефиницииⁱ за дигитална трансформация (DT),ⁱⁱ популяризирана в съвместна публикация на Центъра за дигитален бизнес в Масачузетския технологичен институт (MIT) и консултантската компания Capgemini (Westerman et al., 2011). Оттогава насам са познати широк набор определения, в които се изтъква значението на изграждането на капацитет за организационна промяна (Boston Consulting Group, 2018), ангажирането на таланти и преоткриването на нови бизнес модели (Deloitte, 2018), сливането на технологии и интегрирането на физически и дигитални системи (ЕК, 2019),ⁱⁱⁱ силното лидерство за провеждане на промени и формиране на дигитални способности във фирмите (Harbert, 2021). Описаният контекст на управление на дигиталната трансформация предполага ангажираност на мениджмънт, работници и служители, но и привличане на други заинтересовани страни, отчитайки възможностите за персонализиране на интеракции и трансакции в реално време, с фокус върху отделния потребител в онлайн среда. Изследователи (Li, 2020) изтъкват, че големи технологични компании (като Amazon, Alibaba, Alphabet, Uber) съумяват да установят успешен подход към DT чрез постоянно внедрявани иновации, но и рекалибриране на процеси, в отговор на новите промени в информационните и комуникационни технологии (ИКТ), налагащи внедряване на нови дигитални инструменти. Това от своя страна поражда допълнителни рискове пред повечето компании, осъществяващи DT.

Въпреки ограничения брой емпирични изследвания по отношение на управлението на дигиталната трансформация (Ross et al. 2019), в свързани публикации (Reeves et al., 2018)^{iv} се установява, че в над 70% от случаите не се постигат желаните и предварително планирани резултати. В този смисъл не разработването на нови стратегии, модели или нов организационен дизайн се оказват от ключово значение, а управлението на прехода от един етап на внедряване на нови ИКТ към друг, както и постоянното приспособяване към промени, отразяващи се върху осъществяването на дейността. Необходимо е и балансиране на стратегически и оперативни решения, възприемането на добри практики, съобразно сектора или отрасъла, в които фирмата оперира. В това отношение своята роля имат и специализирани консултантски компании.

ⁱⁱ Следва да се има предвид, че концептуализиране на дигитална трансформация е познато отпреди (Patel & McCarthy, 2000), макар понятието да набира популярност от 2011 – 2014 г. насам (Van Veldhoven & Vanthienen, 2022, p. 629).

ⁱⁱⁱ Съгласно възприетата практика понятието да се съкращава с абривиатурите “DX” или “DT”, последната е използвана и в настоящия текст.

ⁱⁱⁱ Цитирани определения, систематизирани след проучване на текстове в бази данни като EBSCO, Google Scholar (Verina & Titko, 2019, p. 721).

^{iv} Цитираните автори анализират оповестявания в мениджмънт отчетите на 70 000 публично листвани компании в САЩ по ключови думи, които касаят разходи по реструктуриране, анонсирани важни корпоративни промени, в т.ч. във връзка с внедрявани технологии.

Целта на настоящия доклад е да бъдат систематизирани предлагани нови подходи за управление на дигиталната трансформация, с оглед на разрастващия се глобален консултантски пазар за предоставяне на такива услуги. В тази връзка е представен синтез на препоръчвани управленски решения и насоки за работа, съгласно обхвата на предлагани услуги от фирмите в Голямата четворка (Big Four),ⁱ отчитайки вертикалите на техния бизнес (в т.ч. в области като стратегическо планиране, корпоративни финанси, фирмено реструктуриране, управление на човешки ресурси и др.). Реферирано е към практиката и на други водещи на консултантския пазар за управление на DT дружества, като е взето под внимание и навлизането на технологични компании, което в редица случаи поставя под въпрос адекватната експертиза на фирмите в Big Four, в частност по повод внедряване и използване на нови цифрови технологии. В допълнение на избраната за изследване извадка компании е възприето ограничение и по отношение на времевия обхват. Разгледани са подходи, прилагани в годините преди и след избухването на пандемията от COVID-19, както и адаптирането на нови решения в контекста на последвало постпандемично възстановяване и широко навлизане на технологии като изкуствения интелект (AI), очертаващ се като ключов елемент за осъществяване на DT понастоящем.

Въз основа на проучената вторична информация се установява, че съобразно предлаганите нови подходи за управление на дигиталната трансформация, приоритетно е интегрирането на нови технологии в дейността, за сметка на реконфигурирането на фирмените стратегии.

Подходи за управление на DT, предлагани от водещи консултантски фирми

Водещи в световен мащаб консултантски компании предлагат различни подходи за управление на дигиталната трансформация, адаптирани към нуждите на клиентите си, но обичайно насочени към генериране на стойност чрез подобрени резултати от дейността и растеж на основата на дигиталните технологии.

От McKinsey (Rosenfield et al., 2018) разглеждат настъпващите с внедряването на ИКТ всеобхватни промени преди всичко в стратегическа перспектива, на фона на възникването на дигитални екосистемиⁱⁱ и с оглед съпътстващите трансформацията разрушителни иновации. Това може да доведе до формиране на няколко глобални екосистеми на свързани сектори и отрасли.ⁱⁱⁱ В такъв контекст успешната DT предполага изграждане на фирмени способности чрез използване на една или няколко дигитални платформи, които са в основата на формиране на дигитални екосистеми, постоянно адаптиране на бизнес модела, разглеждане на опции за неорганичен растеж (чрез сливания и придобивания), което да осигури бърз достъп до съответна кадрова и технологична обезпеченост. Сходни препоръки имат и други консултантски компании преди и по време на пандемията от COVID-19.

Таблица 1: Подходи за управление на DT, предлагани от Голямата четворка към 2020 г.

ⁱ “Big Four” е общоприето в световен мащаб означение на четирите най-големи одиторски компании, които с времето се специализират и в предоставяне на широка палитра от консултантски услуги.

ⁱⁱ Характеризирани с това, че възникват като мрежи от свързани услуги, интегрирайки цифрови данни от различни отрасли. От своя страна потребителят от единна точка за достъп достига до разнообразие от съответно оферирани продукти или услуги, без да напуска екосистемата (Rosenfield et al., 2018, p. 6).

ⁱⁱⁱ Такива примери, според консултантите на McKinsey, са сферата на финансовото посредничество, интегрирането на технологични, медийни и телекомуникационни отрасли (ТМТ), и др.

Насоки/ Компания от Big Four	E&Y	Deloitte	KPMG	PWC
Целеполагане/ стратегически решения	✓	✓		✓
Адаптиране на бизнес модела	✓		✓	✓
Преформатиране на оперативни процеси	✓	✓	✓	✓
Развиване на фирмени способности чрез интегриране на ИКТ		✓		✓
Ускоряване на интеракциите с клиенти и контрагенти	✓		✓	✓
Лидерство в организациите при управление на промените		✓		
Сътрудничество в екосистеми/ с широк кръг заинтересовани страни	✓	✓	✓	

Източник: Систематизация на автора.

От Ernst & Young предлагат следване на холистичен подход на трансформация в широка рамка: от определяне на целите пред компанията, адаптиране на бизнес модела, на процеси на оперативно равнище. Допускат се и опции за стратегическо коопериране, формиращо стойност във веригите за доставки, споделяне на услуги, определяне на портфейл от инициативи на основата на ИКТ, съобразени с пазарните възможности по отношение на предлагане на продукти, позициониране в подходящи сегменти и др. В тази връзка към 2020 г. от E&Y изтъкват как ИКТ допринасят за ускоряване и „индустриализиране“ на трансформацията в стратегически аспекти на бизнеса – имайки предвид възможностите за бързо приспособяване към промени в обкръжението, но и за взаимодействие със заинтересованите страни, което подкрепя развитието на компанията в по-дългосрочен времеви хоризонт и предполага устойчива рентабилност.ⁱ

В годините преди пандемията от COVID-19 Deloitte Touche Tohmatsu открояват потенциала за трансформация на бизнеса на основата на интегриран подход, използвайки технологиите – на равнище работни процеси, развивайки фирмените способности,ⁱⁱ но и чрез аналитично обосноваване на стратегическите решения. В съвместна публикация с MIT Sloan School of Management се изтъква значението на лидерството при управление на промените, съпътстващи реализацията на фирмени дигитални стратегии, възприемани като определящи за осъществяване на дигитална трансформация, а не конкретни технологии (Kane et al., 2015). В този контекст консултантите на Deloitte в САЩ предлагат концепцията за достигане на дигитална зрялост (*digital maturity*) на основа на няколко опори (*digital pivots*),ⁱⁱⁱ сред които: адаптивна и надеждна технологична инфраструктура, умело използване на цифрови данни и тяхното монетизиране (в т.ч. чрез вграждането им в конвенционални продукти и услуги, което да осигури

ⁱ Логично, с оглед профила на E&Y, е налице допълване на услуги по повод на стратегическо консултиране, реструктуриране и др.: https://www.ey.com/en_gl/digital/transformation (Accessed: 5 May, 2020).

ⁱⁱ В т.ч. умения и квалификация на служителите по повод интегриране на новите технологии в дейността.

ⁱⁱⁱ Тези своеобразни опорни оси (Gurumurthy & Schatsky, 2019) са общо седем. Освен изброените, се реферира и към изграждане на мрежи от екипи с подходящи цифрови компетенции, въвеждане на „интелигентни работни процеси“ – съчетавайки способностите на служителите с възможностите на технологиите, унифицирано потребителско изживяване – в осъществяваните интеракции онлайн и във физическа среда, адаптивност на бизнес модела спрямо променящите се пазарни условия.

достъп до нови клиенти и ръст в продажбите), ангажираност за сътрудничество в екосистеми с външни бизнес партньори (в т.ч. стартъп компании, бизнес инкубатори, развойни организации).

Сходни са насоките в публикации на KPMG (2017), отчитайки, че ключови отлики на дигиталната трансформация са възможностите за иновации, мащаба и гъвкавостта на решенията на основата на ИКТ. В този смисъл, за да бъдат успешни в дигиталния свят, бизнесите следва да са в състояние да отговорят на очакванията на потребителите, но и на широк кръг заинтересовани страни (в т.ч. такива в рамките на собствените структури – в лицето на служителите), стремейки се да предвиждат, но и да влияят върху техните действия и начин на работа – чрез бързо адаптиране на ключови оперативни процеси и при необходимост на бизнес модела.

От PricewaterhouseCoopers изтъкват необходимостта компаниите да интегрират напълно в своя бизнес модел дигиталните и конвенционалните продукти и услуги, което да се отрази и върху работата с клиенти и контрагенти в каналите на реализация. Консултантските услуги на PwC са фокусирани върху развиване на дигитални стратегии чрез рационализиране на процеси, подпомагане на генерирането на стойност чрез ускорени интеракции с клиентите. В тази връзка към 2020 г. се изтъква значението на съответни мобилни и уеб-базираны решения и използването на cloud инфраструктура.ⁱ

Припокриващ се елемент в подходите на разгледаните компании е доброто познаване и насърчаване на изпреварващо идентифициране на потребителските очаквания чрез акумулиране, обработка и анализиране на цифрови масиви от данни (в т.ч. големи бази данни). Съществена препоръка е развиването на фирмени способности чрез внедряване на нови базирани на ИКТ инструменти, в контекста на фирмената стратегия, като не се подценява и ролята на лидерите в компаниитеⁱⁱ (Kane et al., 2015), в частност за управление и насочване на настъпващите промени, и адаптиране на бизнес модела.

Адаптиране на нови подходи след пандемията от COVID-19

По време на пандемията от COVID-19 множество фирми бяха изправени пред необходимостта скоростно да дигитализират и реконфигурират работни процеси, не само въвеждайки дистанционен режим на работа, но и преминавайки към нови начини на взаимодействие с клиенти и контрагенти. Това от своя страна породило необходимост от поредно надграждане на фирмените способности на основата на ИКТ. Последващият период на постпандемично възстановяване е съпътстван от все по-широкото навлизане на изкуствения интелект (AI) и на самообучаващите се машини, което е предпоставка за нови фундаментални промени в бизнес модела на организации от различни сектори и отрасли. Всичко това способства за генериране на трилионни разходи в световен мащаб, предназначени за осъществяване на дигитална трансформация. Оценките са, че общо изразходваните средства в тази връзка през 2022 г. възлизат на 1,6 трлн. щ. дол. (Page, 2023), докато според прогнозите на International Data Corporation (IDC) през 2026 г. се очаква те да възлизат на 3,4 трлн. щ. дол.ⁱⁱⁱ

ⁱ Посочените решения са релевантни по преценка на съответното консултантско звено в PwC (Digital Services) към м. май на 2020 г.: <https://digital.pwc.com/en.html> (Accessed: 7 May, 2020).

ⁱⁱ В тази връзка лидерството в дигиталната среда не задължително се свързва с високи технологични умения, а с точното разбиране какво трябва да бъде извършено и накъде да бъдат насочени усилията в пресечните точки между бизнес и технологии.

ⁱⁱⁱ Очакванията са основна част от тези разходи да са насочени за трансформация на процеси, свързани с

Консултантските компании реагират като адаптират предлаганите за клиентите си подходи за управление на DT. Така например от McKinsey все повече акцентират върху развитието на технологичните способности на фирмите като източник на стратегическо конкурентно предимство в постпандемичната среда на ускорено внедряване на нови технологични решения. За онагледяване се посочва как трансформационни промени в банковия сектор на основата на AI адресират ключови индикатори за изпълнение (KPI, касаещи обслужване на клиенти, дял на операции извършвани в дигитални канали и др.), което в крайна сметка осигурява и подобряване на рентабилността (Lamarre et al., 2023b).

От страна на компаниите в Голямата четворка със спецификата на предлаганите решения се отличават от E&Y, които изтъкват необходимостта корпоративните им клиенти да са гъвкави и адаптивни по отношение на интегриране на новите технологии (възприемайки типичен подход за стартъп), с готовност да инвестират и развиват бизнеса си с цел осигуряване на ускорени темпове на растеж (подход на венчър фонд) и с ясното разбиране, че трансформацията изисква постоянни и повтарящи се усилия.ⁱ От KPMG също изтъкват значението на внедряваните технологични нововъведения, отразяващи се върху бизнес стратегията и управлението на промяната. Консултантите на PwC открояват интегрирането на данни, използването на cloud услуги, на специфични за конкретния бизнес приложения, без да се подценява киберсигурността. Съобразно различните вертикали на предоставяните услуги, решенията са относими и към стратегически аспекти, свързани с осъществяване на промените, но и оперативни насоки по повод оптимизация на разходи, управление във веригите за доставки и др.ⁱⁱ Що се отнася до новия подход, предлаган от Deloitte, той най-общо се определя от това, че дигиталните възможности следва да определят стратегията.ⁱⁱⁱ

В обобщение, открояващ се елемент в адаптирането на нови подходи за управление на DT от посочените консултантски компании е отдаването на по-голямо значение на внедряването на новите технологии пред прекрояването на фирмената стратегия. Това рефлектира и в навлизането на пазара на консултантски услуги за управление на DT на технологични компании (като IBM, Accenture и др.), чиято дългогодишна експертиза по отношение на системна интеграция, но и внедряване на нови решения, свързани с AI, облачен компютинг, блокчейн, им дава определени предимства при споделяне на опит и добри практики. Очакванията в средносрочна перспектива са прогнозните приходи от предоставянето на такива услуги в глобален мащаб да надхвърлят над два пъти сумарно отчетените равнища по време на пандемията от COVID-19, достигайки 11 млрд. щ. дол.^{iv}

управлението на веригите за доставки, инженерингови дейности, дизайн на продукти, развойни дейности, операции в производството.

В световен мащаб най-голям дял от разходите за DT (35%) се предвижда да се извършват в САЩ:
<https://www.idc.com/getdoc.jsp?containerId=prUS49797222> (Accessed: 10 October, 2023).

ⁱ Вж. https://www.ey.com/en_gl/digital/transformation (Accessed: 14 October, 2023).

ⁱⁱ Вж. <https://www.pwc.com/gx/en/services/consulting/technology.html> (Accessed: 14 October, 2023).

ⁱⁱⁱ Вж. <https://www2.deloitte.com/us/en/insights/topics/digital-transformation.html> (Accessed: 8 October, 2023).

^{iv} По оценки на компанията за пазарни проучвания Facts & Factors Research размерът на глобалния пазар на консултантски услуги за управление на DT през 2020 г. се оценява общо на 5 млрд. щ. дол.:
<https://www.fnfresearch.com/digital-transformation-consulting-market> (Accessed: 10 October, 2023).

Заклучение и дискусия

Ролята на консултантските компании по отношение на избираните подходи за управление на дигиталната трансформация през последните години е двойка. От средата на миналото десетилетие те отдават по-голямо значение на фирмената стратегия и лидирането на промените в организациите, дори в тази роля да попадат мениджъри, които не се открояват с високи технологични умения. Ускореното навлизане на нови цифрови технологии от последните години измества обратно фокуса на първо място към внедряване на решения, за които ключов фактор е технологичната експертиза в различните области на нейното приложение (като киберсигурност, големи бази данни, интернет на нещата, автономни роботи, AI и др.).

Във връзка с горното, своите основания може да има дискусията отпреди няколко години доколко уместно е придаването на ключова роля по отношение на управлението на DT на консултантски компании като тези съставляващи Голямата четворка, имайки предвид, че предоставяните от тях услуги не допускат серия от провали – често наблюдавани при внедряване на нови технологии, преди достигане до вярното решение (Tadros & Smith, 2016).ⁱ

Независимо от очертаните критики, водещи консултантски компании се стремят да адаптират предлаганите от тях подходи за управление на DT. В допълнение на тяхната експертиза в корпоративния сектор, те работят и по проекти с публични институции, университети, изследователски центрове. Такова сътрудничество логично допринася и за по-добра оценка на ефикасността на подходи, модели или специфичен инструментариум, но и за тяхното преосмисляне – по повод настъпващи нови промени в обкръжението или технологиите, отразяващо се и на осъществяването на DT. За онагледяване може да се посочи прилагането на модела 7S на McKinsey в процеса на дигитална трансформация в предприятия разработчици на софтуер (Demir & Kocaoglu, 2019), за които се потвърждава, че в годините преди пандемията от COVID-19 стратегическите решения имат приоритетно значение.ⁱⁱ От друга страна в скорошни публикации на консултантската компания (Lamarque et al., 2023a) се установява, че в извадка от 40 изследвани корпоративни структури едва 1/3 съумяват да подобрят приходите си от продажби или рентабилността,ⁱⁱⁱ поради недобро използване на нови решения от типа на AI. Именно това налага ново префокусиране върху технологиите и тяхното интегриране в дейността.

С оглед на ускореното внедряване на нови технологични решения и на фона на като цяло немногочислените изследвани казуси досега, резонни са очакванията прилагането на нови подходи за управление на DT да бъде обект на по-задълбочени проучвания през следващите години. В този контекст своето място и роля продължават да имат и компании с дългогодишен опит в предоставяне на широк набор от консултантски услуги. В потвърждение на това е и участието им в разрастващия се глобален пазар на услуги в подкрепа на управлението на DT, независимо от навлизащите на него технологични компании.

ⁱ Вместо това се изтъква, че консултантските компании изготвят стратегии, доклади и предлагат решения на предварително идентифицирани проблеми. Оценка на B. Easton от Accenture и на S. Forsterling, Head of Implementation & Account Management в Uber: <https://www.linkedin.com/pulse/why-big-four-right-digital-transformation-sarah-forsterling/> (Accessed: 9 October 2023).

ⁱⁱ Макар моделът 7S да е разработен още през 1980-те години, той има своето приложение в контекста на DT, доколкото обхваща основни фактори, рамкиращи дейността на фирмите, в т.ч. стратегия, структура, системи, споделени ценности, управленски стил, умения (skills), персонал (staff).

ⁱⁱⁱ Консултантите на McKinsey разглеждат показателя възвръщаемост на собствения капитал (ROE).

Исползвана литература

Demir, E. and Kocaoglu, B. (2019), The use of McKinsey's 7S framework as a strategic planning and economic assessment tool in the process of digital transformation. *PressAcademia Procedia (PAP)*, V.9, pp. 114–119.

Gurumurthy, R. and Schatsky, D. (2019), Pivoting to digital maturity, Seven capabilities central to digital transformation. *Deloitte Insights*, 13 March, 2019: <https://www2.deloitte.com/us/en/insights/focus/digital-maturity/digital-maturity-pivot-model.html> (Accessed: 1 October, 2023).

Kane, G. C., Palmer, D., Philips, A. N., Kiron, D. and Buckley, N. (2015), Strategy, not technology, drives digital transformation. *MIT Sloan Management Review* and Deloitte University Press.

Lamarre, E., Smaje, K. and Zimmel, R. (2023a), Rewired to compete. *McKinsey Quarterly*, June 2023.

Lamarre, E., Chheda, S., Riba, M. Genest, V. and Nizam, A. (2023b), The value of digital transformation. *Harvard Business Review*, July 31: <https://hbr.org/2023/07/the-value-of-digital-transformation> (Accessed: 8 October, 2023).

Li, F. (2020), Leading Digital Transformation: Three emerging approaches for managing the transition. *International Journal of Operations and Production Management*, 40(6), pp. 809–817.

Page, M. (2023), The evolution of digital transformation: from pre-internet to post-pandemic. *HatchWorks*, February 3, 2023, Updated: February 15, 2023: <https://hatchworks.com/history-digital-transformation/> (Accessed: 10 October, 2023).

Patel, K. and McCarthy, M. P. (2000), *Digital Transformation: The Essentials of E-business Leadership*. McGraw-Hill.

Reeves, M, L. Fæste, K. Whitaker and F. Hassan (2018), The truth about corporate transformation. *MIT Sloan Management Review*, January 31, 2018: <https://sloanreview.mit.edu/article/the-truth-about-corporate-transformation/> (Accessed: 10 October, 2023).

Ross, J. W., Beath, C. M. and Mocker, M. (2019), *Designed for Digital – How to Architect Your Business for Sustained Success*. Cambridge: MIT Press.

Rosenfield, J., Borruso, M.T. and Simcock, V. (Eds.), (2018), *Winning in digital ecosystems*, Digital McKinsey Insights.

Tadros, E. and Smith, P. (2016), New Accenture boss Bob Easton throws down gauntlet on big four on digital. *Financial Review*, September 6, 2016: <https://www.afr.com/companies/professional-services/new-accenture-boss-bob-easton-lays-down-gauntlet-to-big-four-on-digital-20160829-gr3huj>

Van Veldhoven, Z. and Vanthienen, J. (2022), Digital transformation as an interaction-driven perspective between business, society, and technology. *Electronic Markets* (32), pp. 629–644.

Verina, N. and Titko, J. (2019), Digital Transformation: Conceptual Framework, International Scientific Conference “Contemporary Issues in Business, Management and Economics Engineering”, Vilnius Gediminas Technical University, pp. 719–727: <https://doi.org/10.3846/cibmee.2019.073>

Wade, M. (2015), *Digital Business Transformation. A Conceptual Framework*, Global Center for Digital Business Transformation, June 2015.

Westerman, G., Calm ejane, C., Bonnet, D., Ferraris, P. and McAfee, A. (2011), Digital transformation: A roadmap for billion-dollar organizations. MIT Center for Digital Business and Capgemini Consulting.

Harbert, T. (2021), Digital transformation has evolved. Here's what's new. *Digital Economy*, MIT Sloan School of Management, May 18, 2021: <https://mitsloan.mit.edu/ideas-made-to-matter/digital-transformation-has-evolved-heres-whats-new> (Accessed: 10 October 2023).

KPMG (2017), Destination (un)known, Key steps to guide your digital transformation journey, KPMG International Cooperative, Switzerland.

**ИНОВАТИВНИ ИНФОРМАЦИОННИ ТЕХНОЛОГИИ
ЗА ДИГИТАЛИЗАЦИЯ НА ИКОНОМИКАТА**

Сборник с доклади

Колектив

Дизайн на корицата: Емилия Лозанова

Даден за печат: 18.03.2024 г.

Формат 8/60/84; ПК 21,75

ISSN 3033-0432 (print)

ISSN 3033-0467 (online)

ИЗДАТЕЛСКИ КОМПЛЕКС – УНСС