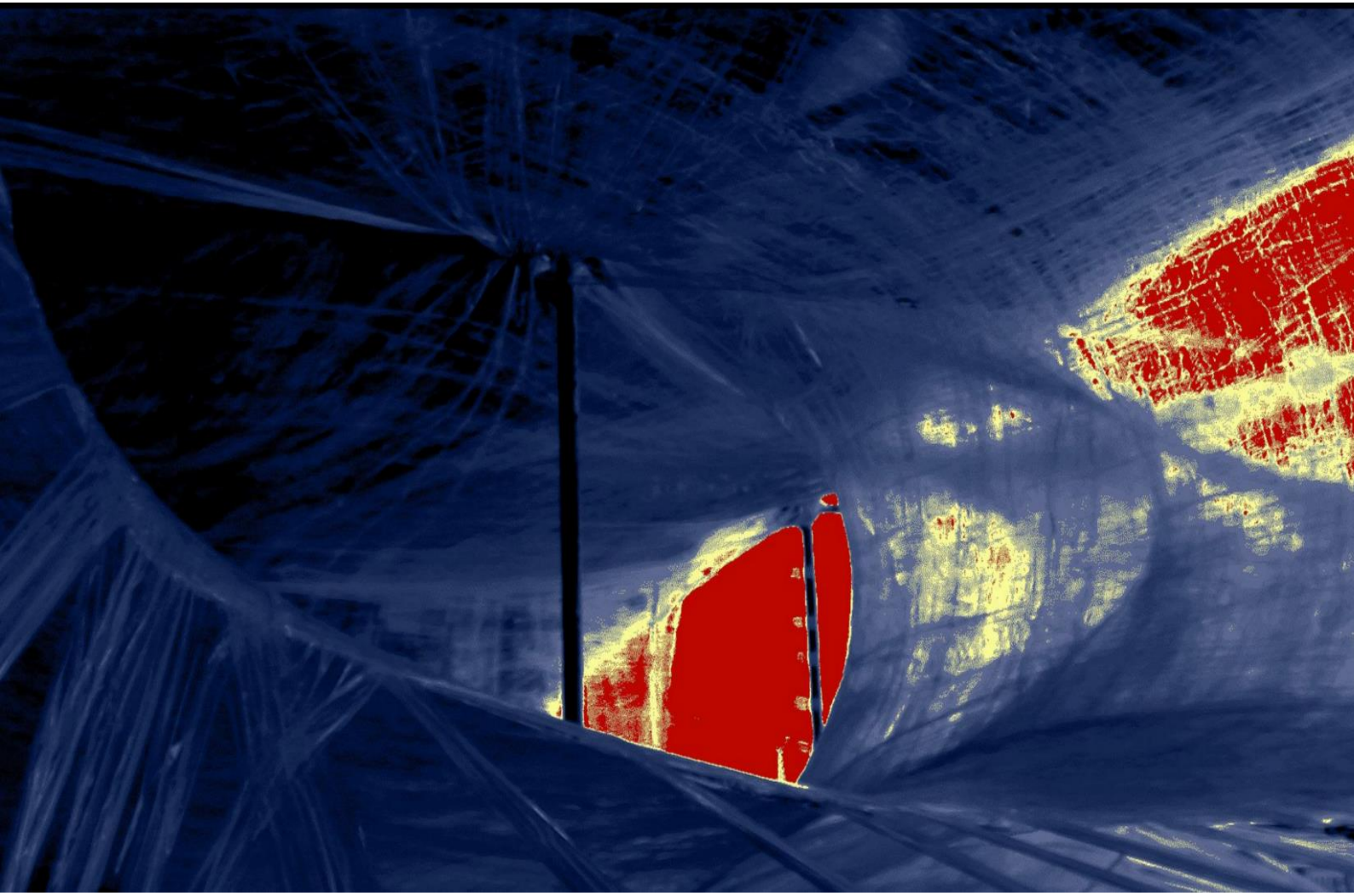


BIG DATA

Sztuczna inteligencja i analityka wielkich zbiorów danych w firmie i gospodarce – zarządzanie informacją i zastosowania praktyczne

Redakcja naukowa:

Aleksander Żołnierski, Dariusz Jaruga

Instytut Nauk Ekonomicznych
Polskiej Akademii Nauk

BIG DATA

Sztuczna inteligencja i analityka wielkich zbiorów danych w firmie i gospodarce – zarządzanie informacją i zastosowania praktyczne

Warszawa 2024

Instytut Nauk Ekonomicznych
Polskiej Akademii Nauk

INSTYTUT NAUK EKONOMICZNYCH POLSKIEJ AKADEMII NAUK

BIG DATA

Sztuczna inteligencja i analityka wielkich zbiorów danych w firmie i gospodarce – zarządzanie informacją i zastosowania praktyczne

Redakcja naukowa:

Aleksander Żołnierski, Dariusz Jaruga

Warszawa 2024

Instytut Nauk Ekonomicznych
Polskiej Akademii Nauk

Recenzenci:

dr hab. Leszek Bohdanowicz

dr hab. Zbigniew Matyjas

Wydawca:

Instytut Nauk Ekonomicznych Polskiej Akademii Nauk

Warszawa 2024

Wydanie I

Aleksander Żołnierski – Instytut Nauk Ekonomicznych Polskiej Akademii Nauk

Dariusz Jaruga – Wydział Nauk Politycznych i Studiów Międzynarodowych Uniwersytetu Warszawskiego

Objętość: 8,22 ark. wyd.

ISBN: 978-83-68039-08-5

Instytut Nauk Ekonomicznych PAN

00–330 Warszawa, ul. Nowy Świat 72

www.inepan.pl

e-mail: inepan@inepan.pl

Spis treści

Wprowadzenie - wszyscy tworzymy przyszłość

Gdy w 1999 Darcy DiNucci użyła po raz pierwszy terminu Web 2.0 określając nim proces współtworzenia treści przez użytkowników internetu, zapewne nie miała na myśli potencjału, jaki jest z nim związany. Ponad dwie dekady po artykule Fragmented Future, burzliwy rozwój sztucznej inteligencji i aplikacji opartych na analizie Big Data pozwala na coraz szersze wykorzystanie niestrukturyzowanych zasobów sieci do predykcji przyszłości – ciągle bardziej tej bliskiej, niż dalszej w tak wielu dziedzinach, że praktycznym wyzwaniem jest znalezienie obszaru, który nie poddaje się temu zjawisku.

Od gospodarki, poprzez zarządzanie, medycynę, nauki o bezpieczeństwie aż do filozofii, lingwistyki czy literaturoznawstwa, wszędzie narzędzia do analiz Big Data i sztuczna inteligencja idą w sukurs badaczom, praktykom i strategom. Im więcej z nas korzysta z narzędzi online, smartfonów, smartwatch'y i innych urządzeń przenośnych, komputerów, laptopów, kamer monitoringu, usług medycznych, dowozu posiłków, tym większe są szanse na to, że nasze zachowania przyczyniają się do tworzenia innowacyjnych, niespotykanych dotąd, rozwiązań. Im większa możliwość interakcji z siecią i integracji z urządzeniami IoT pracującymi nie tylko w chmurze, ale także w rozwiązaniach fog-computing'u, tym większa ilość generowanych danych, które ktoś, gdzieś, w jakimś celu może analizować. Każdy z nas dostarcza tych danych do specjalistycznych baz danych – w większym lub mniejszym stopniu (choć coraz częściej w większym). Danymi „karmimy" technologie AI i BD, ale także sprawiamy, że wiedza i możliwości sprawnego analizowania danych kreuje fundament władzy. Władzy realnej, niezależnej od wpływów politycznych, ekonomicznych, czy woli suwerena. Dziś już wiemy,

że decyzja i wybór nie są zależne wyłącznie od woli wyborcy, bo tę można w pewnym zakresie kształtować narzędziami wykorzystującymi AI. O potencjale takich narzędzi oficjalnie dowiedzieliśmy się po wyborach prezydenckich w USA w 2016 roku, zaś o mechanizmach wpływających na wyniki wyborów odbywających się rok wcześniej w Polsce pewnie wkrótce będziemy mogli wiedzieć o wiele więcej.

Wielu z nas, mówiąc, pisząc, czy myśląc o sztucznej inteligencji personifikuje algorytmy zgodnie z kulturowym przesłaniem popkultury lat ’50 XX wieku. Wydawało się, że sztuczna inteligencja uczyni nas zdrowszymi i bardziej zamożnymi. Teraz, zaczynamy dostrzegać zagrożenia z nią związane, ale także gigantyczny pozytywny potencjał rozwiązań technologicznych niemożliwych do osiągnięcia wcześniej. Narrację na temat zagrożeń zdają się kształtować przedstawiciele największych koncernów w branży – od OpenAI po Google DeepMind.

Próbując skutecznie korzystać zarówno z AI, jak i BD należy pamiętać stwierdzenie Darona Acemoglu; "Istnieją dwie sytuacje, w jakich społeczeństwo może czerpać korzyści z dowolnej technologii: albo być jej właścicielem, albo wówczas, gdy technologia rozwija się w sposób pozwalający na generowanie wysokich płac i zatrudnienia".

Nasza monografia podejmuje problematykę związaną z tymi obszarami, w których wykorzystanie sztucznej inteligencji i technik analiz Big Data pozwala na generowanie dobrobytu, podnoszenia poziomu bezpieczeństwa i dobrostanu społeczeństw, jednocześnie nie naruszając wolności i swobód jednostki. W książce poruszamy tematykę technologii sensu stricto od kwestii hardware’u i software’u, baz danych, technologii blockchain, po obszary bezpieczeństwa, identyfikacji kompetencji organizacyjnych, problematyki efektu dyspersji, percepcji informacji, kwestii języka i zaburzeń rozwojowych człowieka czy realizmu krytycznego, jako podejścia badawczego.

Dariusz Jaruga, w tekście na temat bezpieczeństwa i podatności oprogramowania komputerowego na ataki, porusza kwestie edukacji programistów w zakresie bezpieczeństwa, ale wskazuje także na administratorów i ich rolę w zachowaniu bezpieczeństwa konfiguracji sieci. Przedstawia znaczenie procesu identyfikacji nowych możliwości zabezpieczenia i kwestie odporności oprogramowania na błędy popełniane przez programistów. Wskazuje na nowe możliwości analiz repozytoriów danych w zakresie identyfikacji zagrożeń i badań podatności programów komputerowych na ataki.

Bartłomiej Moszoro koncentruje się na problematyce cyberbezpieczeństwa z punktu widzenia podmiotu gospodarczego. Autor podkreśla potrzebę akceptacji konkluzji, że nikt nie jest bezpieczny w cyberprzestrzeni, a największym cyberzagrożeniem dla polskich firm są wycieki danych. Pomimo faktu, że wzrost bezpieczeństwa przetwarzanych informacji stanowi główny czynnik skłaniający do inwestycji w bezpieczeństwo, wiele firm, które się na to decyduje myśli przede wszystkim o spełnieniu zgodności z regulacjami (compliance) a cyberbezpieczeństwo nadal jest uważane za kwestię wyłącznie techniczną.

W tekście na temat projektowania nowego kierunku studiów Wioletta Matosek i Agnieszka Heba opisują wykorzystanie badań nad problematyką bezpieczeństwa w cyberprzestrzeni. Znaczenie cyberbezpieczeństwa we współczesnym świecie nieustannie wzrasta. Z tego względu autorki przedstawiają koncepcję i cele kształcenia nowego kierunku studiów, który zaprojektowały na Wydziale Nauk Politycznych i Studiów Międzynarodowych Uniwersytetu Warszawskiego. Przedstawiają wyniki badań i analizy ofert podobnych kierunków studiów prowadzonych na krajowych i zagranicznych uczelniach.

Nataliya Poplavska porusza kwestie potrzeby stosowania krytycznej analizy przez odbiorców informacji i kształtowania mechanizmów zapewniających ich odporność informacyjną w warunkach wojennych. Istotnym narzędziem ograniczania wpływu dezinformacji jest edukacja medialna, która łączy w sobie trzy wzajemnie powiązane zestawy kompetencji: kompetencje informacyjne, kompetencje medialne oraz umiejętności technologiczno-cyfrowe. Autorka rozważa możliwości rozszerzenia edukacji medialnej na odbiorców w różnym wieku oraz określa główne formy i metody organizacji edukacji w tym zakresie.

Tomasz Gruszka, w swoim tekście na temat wykorzystania technologii Blockchain dla ograniczenia dezinformacji i redefinicji modelu biznesowego przedsiębiorstw medialnych, przedstawia przykład projektu Pix.T. Cyfryzacja w branży mediów przyniosła wiele pozytywnych zmian, ale i zmaterializowanych zagrożeń. Projekt Pix.T dąży do rozwiązania problemu utraty kontroli nad dystrybucją zdjęć reporterskich w internecie. Wykorzystanie technologii Blockchain zwiększa kontrolę nad obrotem treściami medialnymi, ale też staje się warunkiem wstępnym zmiany modelu biznesowego wytwórców treści i przywrócenia rentowności obrotu w branży medialnej.

W materiale dotyczącym poszukiwania ukrytych kompetencji, Aleksander Żołnierski, Adam Gulczyński i Jacek Gulczyński na przykładzie technologii grafenowych opisują możliwości ich

zastosowania w medycynie. Wyniki badań, jakie przedstawiają, wskazują na możliwość identyfikacji nowych kompetencji strategicznych w jednostkach naukowych zaangażowanych w badania nad tym materiałem, zaś zakres i możliwości wykorzystania grafenu w medycynie wskazują na pojawienie się kompetencji organizacyjnych w instytucjach, które dotąd nie były związane z medycyną.

W swoim tekście, Anna Jesionek podejmuje problematykę wulgaryzacji języka w materiałach dostępnych za pośrednictwem Internetu. Zdaniem autorki, proces wulgaryzacji języka może powodować wiele poważnych zaburzeń rozwojowych człowieka. Opracowany przez autorkę model regresji logistycznej wskazał, że wykorzystanie niektórych wulgarnych słów w tekście utworu zwiększa jego popularność.

Realizm krytyczny, którego dotyczy tekst Małgorzaty Stochmal, jest interesującym podejściem badawczym pozwalającym na dogłębne analizowanie rzeczywistości społecznej. Refleksja nad tym podejściem badawczym wskazuje na korzyści jego zastosowania, a wynikającym przede wszystkim z łagodzenia mankamentów podejścia pozytywistycznego i postpozytywistycznego. Autorka przedstawia ontologiczne, epistemologiczne i etyczne założenia realizmu krytycznego, które można z powodzeniem zastosować przy realizacji projektów analitycznych związanych z dużymi zbiorami danych.

Justyna Komorowska analizuje technologię zabezpieczeń baz danych w kontekście przedsiębiorstw. Dane użytkowane przez przedsiębiorstwa są wyjątkowe ze względu na ich bogactwo i wysoki poziom zróżnicowania informacji. Tworzenie skutecznej – z punktu widzenia bezpieczeństwa – struktury informatycznej oraz zastosowanie odpowiedniego oprogramowania ochronnego są jednymi z najważniejszych zadań stojących przed osobami zarządzającymi bazami danych. Współczesne narzędzia informatyczne wykorzystują algorytmy AI, tym samym ucząc się, na jakie zdarzenia reagować w sposób optymalny, tworząc jednocześnie spersonalizowane alerty wraz z ich priorytetyzacją.

Zamykający monografię tekst autorstwa Łukasza Piętaka przedstawia wykorzystanie modeli autokorelacji przestrzennej do analizy występowania efektu dyspersji (tzw. spillover effect). Przedstawione rozważania dotyczą programu ekonometrycznego R, który pozwala na wykonywanie zaawansowanych obliczeń modeli ekonometrycznych. Autor ilustruje możliwości rozwiązania potencjalnych ograniczeń wynikających z braków oprogramowania dla modeli

panelowych wykorzystujących elementy przestrzenne programu R, które nie zapewniają kompletnych rozwiązań, co może sprawiać wiele problemów podczas estymacji modeli.

Bogactwo poruszanych w monografii tematów wskazuje z jednej strony na różnorodność możliwości zastosowań sztucznej inteligencji i metod Big Data, a z drugiej oddaje – poprzez ekspresję zainteresowań badawczych autorów – szeroki wachlarz dominujących obecnie tematów, które kształtując naszą rzeczywistość nie pozostają obojętne badaczom i naukowcom.

Aleksander Żołnierski

Dariusz Jaruga

Dariusz Jaruga

Bezpieczeństwo a podatność oprogramowania komputerowego - analiza

Wstęp

Ada Loverance, Brytyjska matematyczka, żyjąca w pierwszej połowie XIX wieku współpracowała z matematykiem Charles Babbage, któremu przypisuje się wymyślenie koncepcji pierwszego automatycznego komputera cyfrowego. Dla wymyślonego przez Babbage koncepcji komputera, Ada Loverance opracowała prototyp cyfrowego programu komputerowego. Z tego powodu Ada Loverance nazywana jest pierwszą programistką na świecie[1]. Na jej cześć jeden z języków programowania opracowany na początku lat 80 dla Departamentu Obrony Stanów Zjednoczonych otrzymał nazwę Ada[2]. Wracając na chwilę do pierwszej połowy XIX wieku, nikt wówczas prawdopodobnie nie przypuszczał, że komputery i oprogramowanie będą stanowiły tak istotną część funkcjonowania świata w XXI wieku. Dziś urządzenia komputerowe to nie tylko standardowe komputery osobiste, laptopy i smartfony. Urządzenia posiadające procesor, pamięć i program komputerowy montowane są w wielu urządzeniach specjalistycznych jak i codziennego użytku. Do urządzeń specjalistycznych można zaliczyć aparaturę medyczną,

[1] *Ada Lovelace | Biography, Computer, & Facts | Britannica* [na:] https://www.britannica.com/biography/Ada-Lovelace, dostęp 11 listopada 2022 r.; *Charles Babbage | Biography, Computers, Inventions, & Facts | Britannica* [na:] https://www.britannica.com/biography/Charles-Babbage, dostęp 11 listopada 2022 r.
[2] *Computer programming language - SQL | Britannica* [na:] https://www.britannica.com/technology/computer-programming-language/SQL, dostęp 11 listopada 2022 r.

roboty przemysłowe, urządzenia IoT, komputery pokładowe we wszelkiego rodzaju pojazdach, a kończąc na przedmiotach codziennego użytku jak pralka czy lodówka.

W pewnym sensie otaczające nas przedmioty dość często wyposażone są w różnego rodzaju urządzenia komputerowe. Samo urządzenie komputerowe bez oprogramowania jest nieużyteczną materią fizyczną. Dopiero po zaprogramowaniu, urządzenia te zaczynają wykonywać przypisane im funkcje. Niestety nie zawsze urządzenie komputerowe działa prawidłowo. Jedną z przyczyn wadliwego działania urządzenia komputerowego są błędy w oprogramowaniu. Jednym z przykładów błędów w oprogramowaniu, które doprowadziło do katastrofy był dziewiczy lot rakiety Ariane 5. 4 czerwca 1996 roku pierwszy lot rakiety Ariane 5 po około 40 sekundach zakończył się niepowodzeniem. Na wysokości około 3700 metrów rakieta zboczyła z zaplanowanego toru, po którym miała się poruszać, po czym rozpadła się i eksplodowała. Kilka dni po tym wydarzeniu została powołana komisja specjalistów do zbadania sprawy katastrofy rakiety. W wyniku prac komisji w podsumowaniu raportu wskazano kilka powodów, które przyczyniły się do katastrofy. Między innymi jednym z nich był błąd w oprogramowaniu w komputerze pokładowym odpowiedzialnym za system naprowadzania i kontroli położenia rakiety. Błąd polegał na przekroczeniu dopuszczalnej wartości prędkości, która mogła być przechowywana w wewnętrznej zmiennej programu. To w konsekwencji spowodowało, że trajektoria lotu rakiety została wyliczona nieprawidłowo. Błąd pociągnął za sobą łańcuch zdarzeń na urządzeniach wykonawczych, które docelowo doprowadziły do zniszczenia rakiety Ariane 5. W podsumowaniu wskazano, że nikt nie przeanalizował kodu programu, który wcześniej był używany w identycznym komponencie rakiety Ariane 4 pod kątem zakresu zmiennych w programie i możliwych do osiągnięcia przez te zmiennych wartości maksymalnych. Rakieta Ariane 4 miała inną charakterystykę wznoszenia i inne wartości prędkości dla pierwszych 40 sekund lotu w porównaniu z rakietą Ariane 5. W związku z zaistniałą sytuacją komisja opracowała czternaście rekomendacji, których celem było znaczące zmniejszenie ryzyka wystąpienia awarii o podobnym charakterze w przyszłości. Wśród przedmiotowych rekomendacji bardzo dużo uwagi poświęcono analizie kodu programów komputerowych, ich testowania i analizy pod kątem przyjętych założeń i ograniczeń[3]. W tym konkretnym przypadku awaria rakiety Ariane 5 spowodowała tylko straty materialne, ale

[3] J.-L. Lions, *ARIANE 5 Failure - Full Report* [na:] http://sunnyday.mit.edu/nasa-class/Ariane5-report.html, dostęp 12 listopada 2022 r.

niestety nie zawsze tak było. Innym przypadkiem błędu w oprogramowaniu, które niestety kosztowało życie kilka osób, było oprogramowanie zastosowane w maszynie do radioterapii Therac-25. Pomiędzy czerwcem 1985 roku a styczniem 1987 roku z powodu wadliwego oprogramowania, przynajmniej sześć osób otrzymało zbyt dużą dawkę promieniowania rentgenowskiego z czego przynajmniej 5 z nich zmarło z powodu napromieniowania. Przypadek ten jest opisywany jako jeden z przykładów najgorszego przypadku jakie miały miejsce w akceleratorach medycznych[4]. Kolejnym przykładem błędu w oprogramowaniu mającego katastrofalne skutki była w 1999 roku awaria sondy Mars Climate Orbiter. Podczas wchodzenia sondy w orbitę Marsa z sondą utracono z nią kontakt i nigdy potem nie udało się go nawiązać. Późniejsze dochodzenie wykazało, że awaria i zniszczenie sondy nastąpiło w wyniku błędu w oprogramowaniu i błędu nawigacyjnego polegającego na tym, że z Ziemi wysłano polecenie w jednostkach angielskich (pound-seconds) zamiast SI (Newton-seconds)[5]. Wymienione w artykule przypadki dotyczą spektakularnych błędów, które odbiły się szerokim echem w świecie. Podobnych przypadków można znaleźć więcej. W zasadzie na powyższych przykładach można byłoby wyciągnąć błędne wnioski, że błędy w oprogramowaniu nie dotyczą przeciętnego zjadacza chleba. Tak jednak nie jest, każdy z nas używa różnych urządzeń działających w oparciu o oprogramowanie. Zdarzające się zawieszenia komputera, niebieski ekran „śmierci" w Windows, czy wadliwie działające smartfony to tylko wierzchołek góry lodowej.

Przypadek biblioteki Apache Log4j

We współczesnym świecie oprogramowanie budowane jest w oparciu o gotowe komponenty, które zostały napisane w konkretnym celu. Wiele innych programów bazując na takich komponentach realizuje własne zadania. Obowiązuje tu zasada niewyważania otwartych drzwi. Programiści chętnie korzystają z gotowych i wydawałoby się sprawdzonych fragmentów oprogramowania napisanego przez innych ludzi. Dobrym przykładem jest tutaj oprogramowanie dostępne na otwartej licencji z dostępnym kodem źródłowym. Jeśli w takim komponencie - bibliotece programistycznej pojawi się błąd, będzie on także obecny w innych

[4] N. Levenson, *Medical Devices: The Therac-25*, http://sunnyday.mit.edu/papers/therac.pdf.
[5] *In Depth | Mars Climate Orbiter* [na:] „NASA Solar System Exploration", https://solarsystem.nasa.gov/missions/mars-climate-orbiter/in-depth, dostęp 12 listopada 2022 r.

programach, które korzystają z takiej biblioteki. Jednym z niechlubnych i głośnych przykładów była biblioteka Apache Log4j, która do dziś jest często wykorzystywana przez programistów i służy do obsługi logowania zdarzeń w programach komputerowych. W grudniu 2021 roku w bibliotece Apache Log4j wykryto błąd polegający na tym, że umożliwiał on dostęp do systemu z pominięciem wszystkich zabezpieczeń. Oznaczało to, że możliwy był atak zewnątrz wymagający bardzo niewielu uprawnień. Dodatkowo w wyniku przeprowadzonego ataku włamywacz w dość prosty sposób mógł włamać się np. do serwera celem kradzieży loginów i haseł, mógł także ukraść cenne dane, czy zainfekować urządzenie złośliwym oprogramowaniem. Powszechność użytej biblioteki Log4j stanowiła, że zagrożony był każdy, od dużych instytucji po zwykłego użytkownika smartfonu czy komputera[6].

Zespół CERT Polska działający w strukturze Państwowego Instytutu Badawczego NASK także wydał ostrzeżenie w sprawie podatności Log4j wskazując, że prawdopodobieństwo wystąpienia zagrożenia tą podatnością w organizacji jest wysokie a wektorem ataku jest zdalne wykonanie kodu na urządzeniu ofiary. W komunikacie CERT mamy także informację o przypisanym do podatności identyfikatorze CVE, który jednoznacznie identyfikuje podatność, i o którym będzie dalej mowa.

W tym miejscu dotykamy problemu innej skali niż miało to miejsce w opisanych we wstępie przykładach największych błędów programistycznych związanych z Ariane 5, Therac-25, czy z marsjańską stacją pomiarową. W tamtych przypadkach, choć skutki były spektakularne to nie miały one zasięgu globalnego.

W przypadku błędu w bibliotece Log4j ze względu na jej powszechne użycie w systemach podłączonych do Internetu sprawa przedstawia się zgoła inaczej. Tutaj mamy do czynienie z globalnym zagrożeniem. Nietrudno sobie wyobrazić, że jakaś grupa hakerów włamuje się za pośrednictwem Log4j do kluczowych systemów cyfrowych państwa lub cyfrowej infrastruktury przemysłowej jak elektrownie, zakłady przemysłowe itp.

[6] *Log4j vulnerability - what everyone needs to know* [na:] https://www.ncsc.gov.uk/information/log4j-vulnerability-what-everyone-needs-to-know, dostęp 12 listopada 2022 r.

Rysunek 1. Komunikat o krytycznej podatności w bibliotece Apache j4log na stronie CERT Polska

Źródło: Krytyczna podatność w bibliotece Apache Log4j (CVE-2021-44228) [na:] https://cert.pl/posts/2021/12/krytyczna-podatnosc-w-bibliotece-apache-log4j/, dostęp 4 października 2022 r.

Fakt ten uzmysławia, że im większa liczba bibliotek programistycznych tym większa szansa na różnego rodzaju błędy. W sytuacji, gdy dany błąd powoduje nieprawidłowe działanie programu sytuacja jest zgoła inna, niż gdy mamy do czynienie z błędem, który nie zaburza jej działania ale otwiera drogę przestępcom cyfrowym do ataku. W sytuacji, gdy błąd w oprogramowaniu może być wykorzystany do ataku mamy do czynienia z podatnością oprogramowania. Powszechność Internetu i trend podłączania wszystkich możliwych urządzeń do tej sieci uwidacznia, że jako ludzkość mamy przed sobą stale rosnący nie lada problem do rozwiązania.

Błędy w oprogramowaniu – skala problemu

Mając na uwadze zasygnalizowane przypadki błędów w oprogramowaniu warto w tym miejscu podjąć próbę odpowiedzi na kilka kluczowych pytań m.in.: Czy programiści na przestrzeni ostatnich np. 20 lat nadal popełniają te same błędy? Jakimi metodami współczesny świat troszczy się o to, aby ryzyko obecności błędów w kodzie programów komputerowych było jak najmniejsze?

Na potrzeby niniejszego opracowania w ramach przedmiotowej analizy przyjęto, że okresem analizy błędów w oprogramowaniu będą lata 1999 - 2022, gdzie rok 2022 będzie obejmował

tylko trzy pierwsze kwartały. Powyższe założenie przyjęto w związku z dostępnością danych o błędach w oprogramowaniu zgromadzonych w programie CVE na moment pisania rozdziału, która została omówiona w dalszej jego części.

W związku z przyjętymi założeniami i pytaniami badawczymi przyjęto główną hipotezę badawczą, że na przestrzeni 24 lat, programiści stale popełniają błędy tej samej natury podczas tworzenia oprogramowania, przez co oprogramowanie posiada te same podatności.

Dodatkowo przyjęto także drugą hipotezę pomocniczą, że wektory ataku hackerów (crackerów) nie zmieniły się i są takie same dziś jak i 24 lata temu.

Common Vulnerabilities and Exposures (CVE) to inicjatywa – program – prowadzony przez amerykańską organizację non-profit MITRE[7]. Misją programu CVE jest identyfikacja i katalogowanie publicznie ujawnionych luk w zabezpieczeniach cybernetycznych. W ramach programu każdej wykrytej luce jest przypisywany indywidualny i unikalny numer identyfikacyjny. Opisywaniem luk zajmują się firmy, które współpracują z programem CVE[8]. W listopadzie 2022 roku liczba podmiotów współpracujących z CVE liczyła 258 pozycji. Wśród nich znajduje się wiele dużych podmiotów jak: Adobe, Apple, Apache Software Foundation, Google LLC, IBM, Meta Platforms, Inc., Microsoft, Oracle, Philips, Seagate Technology, Xiaomi Technology Co., Ltd., Yandex N.V. i inne[9]. Dzięki odpowiedniej organizacji pracy w ramach programu CVE, partnerzy współpracujący z CVE przekazują spójne opisy luk wykrytych w zabezpieczeniach. Katalogowanie podatności pozwala uporządkować zarządzanie wykrytymi podatnościami. Dodatkowo można przypisać określoną podatność do odpowiedniego podmiotu. To w znaczący sposób pozwala skutecznie skoordynować wysiłki podmiotów i ustawić priorytety usuwania luk w zabezpieczeniach[10]. Warto także dodać, że w bazie CVE nie są wliczane specjalnie wprowadzone do oprogramowania tzw. „tylne drzwi" – back door. Każda nowa podatność znaleziona w oprogramowaniu otrzymuje swój unikalny identyfikator składający się z trzech członów – CVE-RRRR-NNNN, gdzie CVE jest stałym prefiksem, RRRR – oznacza rok, a NNNN jest liczbą naturalną składającą się od 4 do 7 cyfr. W sytuacji, gdy podatność występuje na styku dwóch programów, wówczas nadawane są osobne numery CVE dla każdego programu z osobna. Aby zapewnić unikalność numerów CVE, każdy partner

[7] *Mitre Corporation*, [w:] *Wikipedia*, 2022.
[8] *Overview | CVE* [na:] https://www.cve.org/About/Overview, dostęp 12 listopada 2022 r.
[9] *List Of Partners | CVE* [na:] https://www.cve.org/PartnerInformation/ListofPartners, dostęp 12 listopada 2022 r.
[10] *Overview | CVE*.

(CVE Numbering Authorities - CNA) rezerwuje sobie pewną pulę numerów. Stąd istnieje ryzyko, że na koniec roku nie wszystkie identyfikatory mogą być wykorzystane[11]. Powyższe będzie miało istotny wpływ na realizację badania opisanego w dalszej części rozdziału. Wokół programu CVE powstało kilka równoległych inicjatyw, takie jak np. Krajowa (dotyczy USA) baza danych o lukach w zabezpieczeniach (National Vulnerability Database - NVD), Wspólny system oceny podatności na zagrożenia (Common Vulnerability Scoring System - CVSS)[12]. Zgodnie z informacją na stronie CVSS pomaga ocenić we właściwy sposób poziom zagrożenia: niski, średni, wysoki, krytyczny, wynikająca z cech podatności. Przyjęty przez CVSS sposób obliczania krytyczności podatności posiada wartość liczbową od 0 do 10 i obecnie składa się z ośmiu elementów takich jak:

- Wektor ataku (Attack Vector -AV);
- Złożoność ataku (Attack Complexity - AC);
- Wymagane uprawnienia (Privileges Required - PR);
- Interakcja użytkownika (User Interaction - UI);
- Zakres (Scope - S);
- Poufność (Confidentiality - C);
- Integralność (Integrity - I);
- Dostępność (Availability -A)[13].

W przedmiotowym przypadku Log4j podatność została określona na wartość maksymalną 10 punktów, w której wektorem ataku jest sieć. Złożoność ataku oceniono na poziom niski, co oznacza, że może go przeprowadzić osoba nieposiadająca wysokich kwalifikacji. Dodatkowo do przeprowadzenia ataku nie są wymagane żadne uprawnienia, nie jest także wymagana interakcja użytkownika, który musi coś zrobić, aby atakujący mógł z podatności skorzystać. Idąc dalej zakres podatności pozwala wpływać na inne komponenty w systemie, które nie wchodzą w skład oprogramowania posiadającego podatność. Także utrata poufności, integralności i dostępności jest na wysokim poziomie. Wektor został pokazany na ilustracji poniżej.

[11] *Home | CVE* [na:] https://www.cve.org/, dostęp 12 listopada 2022 r.
[12] *Related Efforts | CVE* [na:] https://www.cve.org/About/RelatedEfforts, dostęp 12 listopada 2022 r.
[13] *CVSS v3.1 Specification Document* [na:] „FIRST — Forum of Incident Response and Security Teams", https://www.first.org/cvss/v3.1/specification-document, dostęp 12 listopada 2022 r.

Rysunek 2. Komunikat o podatności w bibliotece Apache log4j na stronie NIST

Źródło: NVD - CVE-2021-44228 [na:] https://nvd.nist.gov/vuln/detail/CVE-2021-44228, dostęp 4 października 2022 r.

Powyższy przykład podatności pokazuje jak groźna jest podatność i w jak łatwy sposób możliwe jest skompromitowanie systemu korzystającego z Log4j wykorzystując jako wektor ataku sieć i możliwość uruchomienia zdalnie kodu na atakowanej maszynie bez konieczności posiadania uprawnień.

Podsumowując, dane zgromadzone w CVE w latach 1999 – 2022 (trzy pierwsze kwartały) stanowiły punkt wyjścia do analizy podatności oprogramowania komputerowego.

Analiza podatności w programach komputerowych

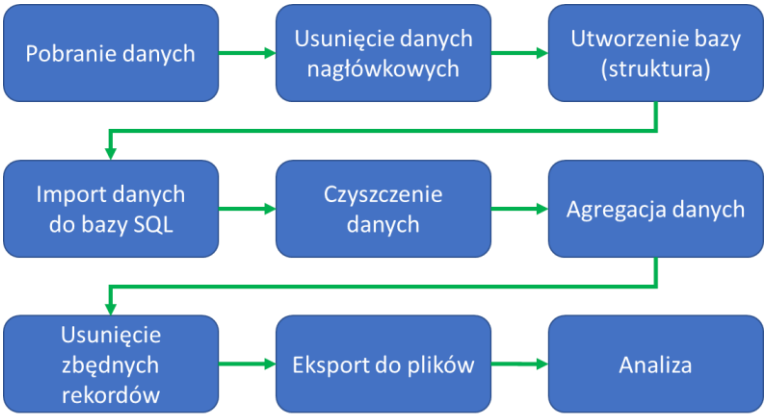
W niniejszej części opisano metodykę zastosowaną w badaniu. Opis w niniejszym podrozdziale został celowo uszczegółowiony z dwóch ważnych powodów. Pierwszym z nich jest zaprezentowanie, że stosunkowo niewielkim kosztem można przeprowadzić dość złożone operacje na danych będących surowcem badawczym. Drugi powód to zilustrowanie, że dany proces badawczy, ze względu na zastosowane narzędzia może być w pełni zautomatyzowany i tym samym wielokrotnie wykorzystywanym w przyszłości, co przekłada się na niższe koszty przyszłych badań i możliwość wykonywania analiz porównawczych za jakiś czas.

Trzeba także wyraźnie zaznaczyć, że zaproponowany sposób rozwiązywania problemu badawczego może być z powodzeniem użyty dla innych przypadków.

W przedmiotowym przypadku metodyka badawcza w zakresie zebrania i przekształcenia danych do postaci wymaganej dla analizy danych składała się z następujących kroków:

1. pobrania danych źródłowych ze strony https://cve.mitre.org/data/downloads/ index.html;
2. usunięcia z pliku źródłowego danych nagłówkowych;
3. utworzenie bazy danych wraz z wymaganą strukturą;
4. import danych źródłowych do bazy danych;
5. usunięcie zbędnych rekordów;
6. czyszczenie danych w bazie w pojedynczych rekordach z niepotrzebnych informacji;
7. eksport do plików zewnętrznych;
8. właściwa analiza przetworzonych danych źródłowych.

Rysunek 3. Etapy przetwarzania i analizy danych w badaniu



Źródło: opracowanie własne.

Ad1. Pobranie danych źródłowych ze strony https://cve.mitre.org/data/downloads/ index.html jest stosunkowo trywialną operacją. Na stronie udostępnione są linki do kilku formatów danych. Ze względu na potrzebę importu danych źródłowych do relacyjnej bazy danych najlepszym formatem był CSV. Takie też zalecenie znajduje się na stronie WWW cve.mitre.org. Mając na względzie problematykę zautomatyzowania procesu badawczego do pobrania danych wykorzystano system operacyjny Linux w dystrybucji Debian, wraz z programem wget, który

jest dedykowany do kolekcjonowania danych ze stron internetowych. W tym miejscu należy zwrócić uwagę na bardzo istotną rzecz związaną z kontrolowaniem bezbłędnej realizacji poszczególnych kroków procesu badawczego podlegającego automatyzacji. Co do zasady procesy uruchomione w systemie operacyjnym w trybie wsadowym muszą być kontrolowane również wsadowo. Badacz uruchamiający taki zautomatyzowany proces powinien na końcu otrzymać informację o poprawności zakończonego procesu albo informację o błędzie wraz z podaniem miejsca, w którym nastąpił problem.

Wychodząc z takiego założenia na potrzeby niniejszego badania przyjęto, że każde polecenie będące etapem wsadowej realizacji badania będzie opatrzone instrukcją warunkową, która będzie sprawdzała poprawność wykonanej czynności. W przypadku gdy nastąpi pierwszy błąd program automatyzujący proces badawczy wyśle stosowny komunikat do badacza po czym zakończy swoje działanie. W systemie Linux w powłoce BASH mamy kilka sposobów obsługi sytuacji wyjątkowych. W przedmiotowym badaniu zostaną zastosowane dwie metody: instrukcja warunkowa i przechwytywanie sygnału[14].

Ad2. W czasie przygotowania mechanizmu automatyzującego proces pobierania danych ważne jest, aby zapoznać się z formatem w jakim te dane zostały zapisane. Automatyzując proces należy sprawić, czy dane zawierają nagłówek, w jaki sposób poszczególne dane zostały podzielone na pola, kolumny itp. W przedmiotowym przypadku analiza danych wykazała, że plik CSV zawiera w nagłówku nadmiarowe wiersze danych, które nie są potrzebne i powinny zostać usunięte. Poszczególne pola w wierszu rozdzielono przecinkami a wartości w pojedynczym polu zostały określone znakami podwójnego cudzysłowu. Powyższa informacja będzie miała kluczowe znaczenie w momencie wykonywania importu danych CSV do tabeli w bazie danych.

Z pliku źródłowego dokonano usunięcia nadmiarowych danych z nagłówka pliku, które nie są potrzebne do analizy. Powyższą operację wykonano za pomocą polecenia tail w systemie Linux.

[14] Sygnał w systemach operacyjnych jest mechanizmem, obsługiwanym zazwyczaj na poziomie jądra systemu operacyjnego, który występuje asynchronicznie i może pojawić się w dowolnym momencie. Pojawienie się sygnału oznacza, że nastąpiła jakaś sytuacja wyjątkowa, którą należy obsłużyć w programie. Program komputerowy może być napisany w taki sposób, że będzie w stanie przechwytywać sygnały i w odpowiedni sposób je obsługiwać. Przykładowe sygnały w systemie Linux to m.in: SIGHUP - zerwanie łączności z terminalem, SIGINT - przerwanie programu, SIGKILL - unicestwienie procesu. Więcej: *Signal(7) - Linux manual page* [na:] https://www.man7.org/linux/man-pages/man7/signal.7.html, dostęp 23 września 2022 r.

Rysunek 4. Przykładowy fragment rekordu CVE w pliku CSV

```
CVE-1999-0001,Candidate,"ip_input.c in BSD-derived TCP/IP implementations allows remote attackers to cause a denial of service (crash or hang) via crafted packets.","BUGTRAQ:19981223 Re: CERT Advisory CA-98.13 - TCP/IP Denial of Service […]
```

Źródło: https://cve.mitre.org/data/downloads/allitems.csv [2022.09.30].

Tu warto nadmienić, że odrzucenie 11 wierszy nagłówkowych w pliku CSV można wykonać na etapie importu danych do bazy SQL. Zdecydowano się jednak na krok pośredni na wypadek, gdyby badacz chciał dokonać importu pliku CSV do innego systemu, który nie ma funkcji zignorowania określonej liczby wierszy z początku pliku.

Rysunek 5. Nagłówek pliku CSV zawierającego dane o podatnościach w programach komputerowych (CVE)

```
"CVE Version 20061101",,,,,
"Date: 20220929",,,,,
"Name","Status","Description","References","Phase","Votes","Comments"
"Candidates must be reviewed and accepted by the CVE Editorial Board",,,,,,
"before they can be added to the official CVE list.  Therefore, these",,,,,,
"candidates may be modified or even rejected in the future.  They are",,,,,,
"provided for use by individuals who have a need for an early",,,,,,
"numbering scheme for items that have not been fully reviewed by",,,,,,
"the Editorial Board.",,,,,,
,,,,,,
```

Źródło: https://cve.mitre.org/data/downloads/allitems.csv [2022.09.30].

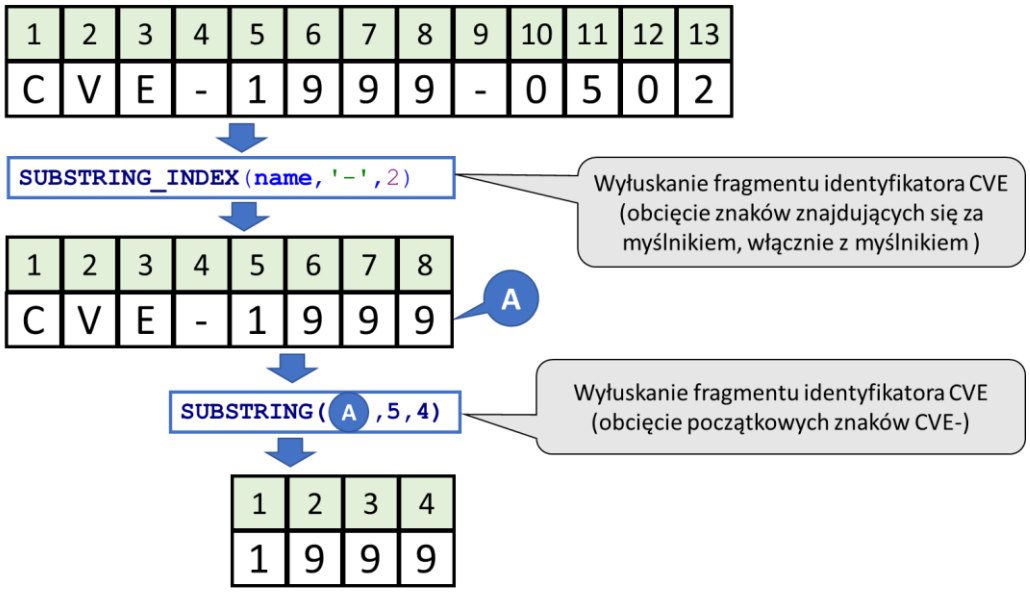
Ad3. Mając już wstępną wiedzę na temat danych źródłowych, chcąc zaimportować je do bazy danych, na serwerze baz danych utworzono nową pustą bazę danych a następnie zbudowano w niej właściwą dla pliku źródłowego CSV struktutrę - tabelę.

W przedmiotowym badaniu struktura pliku CSV zawiera siedem kolumn: "Name","Status","Description","References","Phase","Votes","Comments".

Nazwy kolumn zostały zaczerpnięte z nagłówka źródłowego pliku CSV – patrz Rysunek 5

Docelowo utworzona w bazie danych tabela, przeznaczona do importu danych źródłowych zawiera dziewięć kolumn, ponieważ dodano do niej dodatkową kolumnę, która odpowiada za liczbę porządkową oraz kolumnę year, w której na podstawie identyfikatora CVE został wstawiony rok wpisu podatności. Dodanie kolumny year było ważne z powodu konieczności analizowania danych w szeregu czasowym. Przyjęta w badaniu modyfikacja struktury danych uprości w przyszłości konstrukcję pytań w języku SQL bazy danych.

Rysunek 6. Wyłuskanie roku z identyfikatora CVE



Źródło: opracowanie własne.

Rysunek 7. Wyłuskanie roku z rekordu CVE w języku SQL

```
UPDATE cve_records SET year=SUBSTRING(SUBSTRING_INDEX(name,'-',2),5,4);
```

Docelowo przygotowana w bazie danych tabela miała strukturę jak na rysunku poniżej. Do utworzenia ww. tabeli użyto bazę danych na wolnej licencji MariaDB[15]. Ad4. Kolejnym etapem był import pliku CSV do utworzonej tabeli. Czynność importu można wykonać na kilka sposobów. W przypadku zastosowania trybu wsadowego użyto do tego celu polecenia LOAD DATA. Chcąc korzystać z tego typu ładowania danych należy pamiętać o tym, aby użytkownik bazy danych miał odpowiednie uprawnienia (FILE), bez których cała operacja zakończy się niepowodzeniem.

Dodatkowo podczas importu danych tekstowych należy zwrócić uwagę na standard kodowania pliku źródłowego. Jeśli jest on inny niż zastosowany w bazie (UTF-8) należy dokonać odpowiedniej konwersji kodowania pliku źródłowego. Taka zmiana gwarantuje nam, że dane tekstowe będą w bazie danych odwzorowane we właściwy sposób. Przykładowe polecenie dokonujące takiej zmiany może być następujące: *iconv -f ISO-8859-1 -t UTF-8 no_headers_allitems.csv -o no_headers_allitems_utf.csv* .

Rysunek 8. Struktura tabeli do przechowywania zaimportowanych rekordów CVE z pliku CSV

#	Nazwa	Typ danych	Długość/Zestaw	Domyślnie
1	**id**	**INT**	**11**	**AUTO_INCREME...**
2	name	VARCHAR	20	NULL
3	status	VARCHAR	20	NULL
4	description	TEXT		NULL
5	references	VARCHAR	2048	NULL
6	phase	VARCHAR	2048	NULL
7	votes	VARCHAR	2048	NULL
8	comments	VARCHAR	2048	NULL
9	year	INT	11	NULL

Źródło: opracowanie własne.

[15] *MariaDB Foundation - MariaDB.org* [na:] https://mariadb.org/, dostęp 8 września 2022 r.

Po zaimportowaniu danych tabela w bazie danych jest gotowa do dalszego przetwarzania. W niniejszym badaniu Łącznie zaimportowanych zostało 257 889 rekordów.

Ad5. Jak to zostało zaznaczone wcześniej każdy rekord w bazie CVE ma swój unikalny numer. Aby nie doszło do zduplikowania się numeracji, każdy podmiot posiadający uprawnienia do wpisywania informacji do bazy CVE rezerwuje pewną ich pulę (zakres numerów). Stąd pewna dość znacząca procentowo liczba rekordów jest nieistotna z punktu widzenia przedmiotowego badania. Należy także zwrócić uwagę na fakt, że część rekordów z różnych przyczyn zostały odrzucone, stąd one także kwalifikują się do usunięcia z przygotowanej do potrzeb badania tabeli. Kolejną grupą są rekordy, które z jakiś względów stanowią przypadki sporne, one także nie powinny być przedmiotem dalszej analizy i należy je odrzucić. Powyższe oznacza, że z tylko część zaimportowanych danych będzie podlegała analizie badawczej. Natomiast rekordy zbędne wymagają usunięcia. W przypadku przedmiotowych danych z tabeli usunięto rekordy zawierające w treści: „** RESERVED", „** REJECT", „** DISPUTED".

Rysunek 9. Liczba rekordów CVE przeznaczonych do usunięcia w ramach procesu oczyszczania danych

fraza	suma rekordów
** RESERVED	51706
** REJECT	11088
** DISPUTED	961
** UNSUPPORTED WHEN ASSIGNED	123
** PRODUCT NOT SUPPORTED WHEN ASSIGNED	6
** UNVERIFIABLE	5
**VERSION NOT SUPPORTED WHEN ASSIGNED	5
** UNVERIFIABLE, PRERELEASE	2
**DISPUTED	2
**Resolved	2
** SPLIT	1

Źródło: opracowanie własne.

Ad6. Po usunięciu zbędnych rekordów, kolejnym krokiem jest przeprowadzenie czyszczenia poszczególnych rekordów, konkretnie pola description z nadmiarowych spacji, które czasami pojawiają się w treści. Proces ten jest istotny z punktu widzenia dalszego procesu analizy danych. Dla programów komputerowych ciąg znaków np. dwa słowa rozdzielone jedną lub dwiema spacjami to dwa różne obiekty. Oczywiście problem różnej liczby spacji pomiędzy

wyrazami można rozwiązać w innych sposób stosując w ciągach wyszukiwania wyrażenia regularne, jednak trzeba wówczas mieć na względzie, że użycie wyrażeń regularnych nie zawsze jest możliwe. Poza tym użycie wyrażeń regularnych wymaga większego nakładu obliczeniowego co przekłada się na czas wykonywania polecenia. W sytuacji, gdy przetwarzanych jest bardzo duża liczba danych, czas przetwarzania może być istotnie dłuższy.

Ad7. Ostatnim krokiem w zakresie przygotowania danych było ich wyeksportowanie do plików zewnętrznych. W tym celu skorzystano z wewnętrznego polecenia bazy danych, które pozwala zapisać do zewnętrznego pliku zawartość określonego pola dla wskazanej liczby wierszy. Przykładowe polecenie eksportu danych (pole description z opisem podatności) do pliku tekstowego dla jednego roku przedstawiono poniżej:

SELECT description INTO OUTFILE '/tmp/cve-2012.txt' FROM cve_records_copy where name like 'CVE-2012%';

Dzięki wyeksportowaniu danych do plików tekstowych możliwe będzie skorzystanie z innych programów do analizy, które bazują na plikach i nie potrafią skorzystać ze źródła danych jakim jest baza danych. Do dalszej analizy zostały wykorzystane dane w plikach tekstowych wyeksportowane z bazy oraz z przedmiotowej tabeli w bazie danych. W zależności od sytuacji wykorzystano jeden z przygotowanych zbiorów.

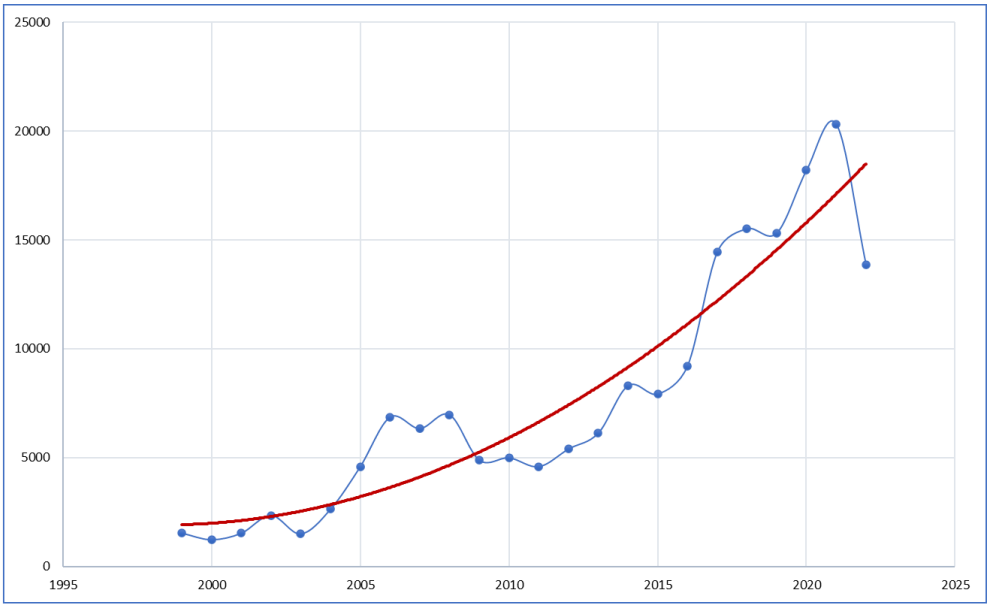
Analiza podatności oprogramowania

Przygotowane do dalszej analizy dane stanowiły punkt wyjścia. Mając na względzie zdefiniowane na wstępie hipotezy badawcze postawiono także kilka pytań, na które w niniejszym opracowaniu zostanie podjęta próba odpowiedzi. Wśród nich postawiono pytanie o liczbę zgłoszonych do repozytorium CVE podatności w kolejnych latach od 1999 do 2022 roku. Czy istnieje możliwość wytypowania najczęstszych podatności i jak zmieniała się ich liczba na przestrzeni ostatnich 23 lat. Czy od 1999 roku do 2022 roku nastąpiły zmiany w zakresie popełnianych przez programistów błędów, które mają wpływ na podatności w zakresie bezpieczeństwa oprogramowania. Powyższe pytania, jak i analiza danych będzie stanowiła także bazę do udzielenia odpowiedzi na postawione na wstępie hipotezy badawcze.

W pierwszym kroku, mając wprowadzone do bazy danych rekordy CVE z opisem podatności, przy pomocy zapytania do bazy danych SQL pobrano zestawienie zawierające rok wraz z sumaryczną liczbą podatności jaka została wprowadzona do przedmiotowego rejestru. Otrzymany wynik pokazuje, że liczba podatności w kolejnych latach ma trend wzrostowy. Na wykresie (Rysunek 10) widać drobne fluktuacje, przykładowo w 2007 roku wpisanych podatności w stosunku do 2006 roku było nieco mniej, tym nie mniej trend wielomianowy wyliczony na podstawie danych z kolejnych lat wyraźnie wskazuje na wzrost liczby podatności w kolejnych latach. Tu warto także zaznaczyć, że rok 2022 obejmuje dane z trzech kwartałów, więc liczba podatności na koniec 2022 roku prawdopodobnie będzie wyższa.

Rysunek 10. Liczba wpisanych podatności w repozytorium CVE w latach 1999 - 2022 (bez 4 kwartału 2022)

rok	liczba_podatnosci
1999	1540
2000	1236
2001	1534
2002	2351
2003	1501
2004	2635
2005	4592
2006	6862
2007	6342
2008	6973
2009	4888
2010	4994
2011	4591
2012	5402
2013	6133
2014	8312
2015	7928
2016	9220
2017	14453
2018	15516
2019	15320
2020	18210
2021	20315
2022	13869



Źródło: opracowanie własne.

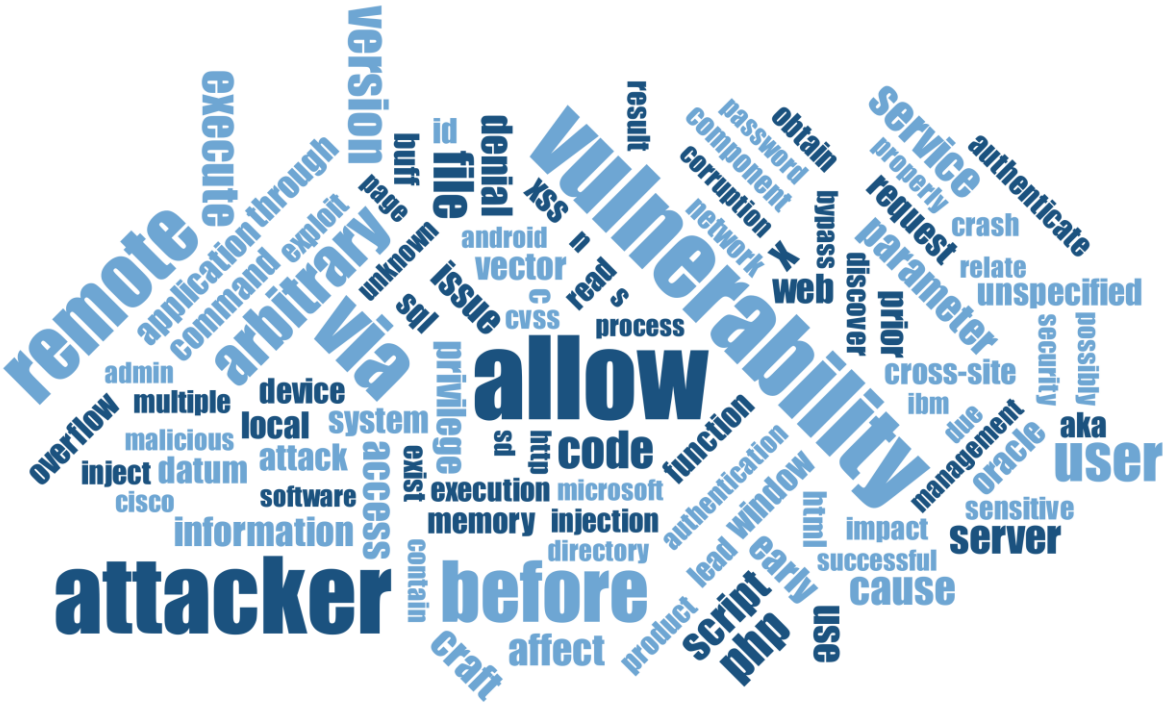
Ze względu na także edukacyjny charakter niniejszej publikacji przedstawiono sposób w jaki zostało skonstruowano zapytanie do bazy danych, aby otrzymać powyższą tabelę i wykres.

SELECT year, COUNT(id) as liczba_podatnosci FROM cve_records group BY year ORDER BY year;

Kolejnym krokiem była analiza częstotliwościowa treści (słów/wyrażeń) zawartej w opisach w kolejnych latach, celem wyłuskania najbardziej istotnych słów, które pojawiają się w opisach zgłoszonych do CVE podatności.

Przedmiotową analizę częstotliwościową wykonano na wyeksportowanych plikach za pomocą programu MaxQda[16], w którym wykorzystano narzędzie do generowania chmury tagów. Generując chmurę tagów w opcjach wybrano lematyzację słów dla języka angielskiego oraz dodano listę słów tzw. „stop words", które zostały wykluczone z analizy. Dzięki lematyzacji otrzymany wynik zawiera słowa sprowadzone do wyrazu podstawowego.

Rysunek 11. Chmura najczęściej występujących wyrazów w opisach CVE w okresie od 1999 r. do 2022 r. (bez czwartego kwartału)



Źródło: opracowanie własne.

W wyniku analizy częstotliwościowej dla całego zbioru danych obejmujących lata 1999 – 2022 (trzy kwartały) widać kilka wyrazów, które powtarzają się najczęściej, są to m.in. attacker, allow,

[16] *MAXQDA | All-In-One Qualitative & Mixed Methods Data Analysis Tool* [na:] „MAXQDA", https://www.maxqda.com/, dostęp 9 września 2022 r.

vulnerability, remote, via, user, arbitrary, version, execute, service. Ze względu na funkcję jaką pełni CVE, podany zbór wyrazów nie jest zaskoczeniem. Widać na nim pewne elementy składowe, które występują przy opisach podatności i pośrednio wskazują na rodzaj podatności. Nie jest zaskoczeniem także wystąpienie czterech kluczowych słów jak attacker, allow, vulnerability, remote, które wskazują, że za wykorzystaniem podatności stoi atakujący, który może wykonać jakąś operację w sposób zdalny. Pośrednio odzwierciedla to niekorzystne zjawiska zachodzące w internecie dla użytkowników, o których dość często można przeczytać choćby w wiadomościach prasowych.

Aby przyjrzeć się dokładniej i podjąć próbę zidentyfikowania najczęstszych podatności, przedmiotowy zbiór danych poddano analizie częstotliwościowej (n-gram) par, trójek, czwórek i piątek słów występujących najczęściej bezpośrednio w swoim sąsiedztwem – oddzielonych tylko spacją.

Na podstawie analizy n-grams (2-grams, 3-grams, 4-grams i 5-grams) wytypowano 7 określeń często występujących w opisach odnoszących się do konkretnych podatności:

1. buffer overflow;
2. sql injection;
3. xss (patrz: cross site scripting);
4. traversal;
5. denial of service;
6. cross site scripting (patrz: xss);
7. allows remote attackers to execute.

Patrząc na przedłożoną listę warto zauważyć, że niektóre podatności, przykładem jest punkt 3 i 6, który dotyczy tej samej podatności, ale nazwanej w odmienny sposób.

Krótka charakterystyka badanych podatności

Charakteryzując pokrótce każdą z ww. podatności:

buffer overflow – przepełnienie bufora jest to podatność polegająca na tym, że za pomocą interfejsu wejściowego program pozwala przekazać do programu więcej

danych niż przewidział na ten cel programista. W efekcie, w wyniku przepełnienia bufora, program przestaje działać prawidłowo a przygotowane i umiejętnie wprowadzone do pamięci komputera dane pozwalają wykonać przez atakującego jego kod, który pozwala na wykonanie niedozwolonej operacji włącznie z ew. możliwością przejęcia kontroli nad atakowanym systemem[17].

sql injection – polega na umiejętnym spreparowaniu zapytania do bazy danych w języku SQL, który pozwala uruchomić nieprzewidziane przez autora aplikacji polecenie. Efektem takiego działania może być m.in. możliwość uruchomienia dowolnego polecenia na atakowanym serwerze i w skrajnym przypadku przejęcie kontroli nad serwerem[18].

xss / cross site scripting – polega na umieszczenie złośliwego kodu programu na stronie WWW, który pobrany przez przeglądarkę klienta oglądającego zaatakowaną stronę, uruchomi się w przeglądarce klienta. Spreparowany kod może np. wykraść dane z komputera użytkownika, do których dostęp ma przeglądarka internerowa. Jest to szczególnie groźne, gdy tego typu złośliwe oprogramowanie wykradnie dane dostępowe przechowywane np. w plikach cookie[19].

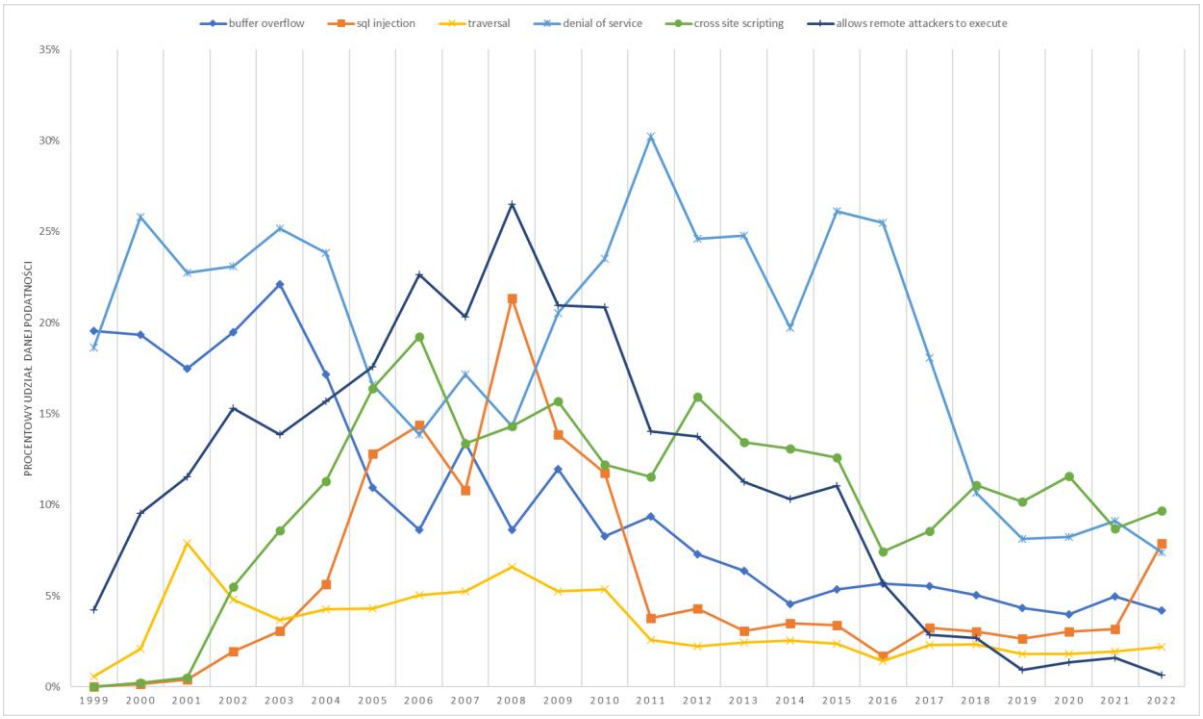
Traversal – polega na manipulowaniu ścieżkami dostępu do plików i katalogów w taki sposób, aby uzyskać dostęp do zasobów, do których nie powinno być dostępu. W skrajnym przypadku może to oznaczać, że osoba atakująca tą metodą może odczytać kluczowe informacje dostępu do usług serwera[20].

[17] Więcej na temat przepełnienia bufora jest dostępne w: *Buffer Overflow | OWASP Foundation* [na:] https://owasp.org/www-community/vulnerabilities/Buffer_Overflow, dostęp 28 października 2022 r.
[18] Więcej na temat wstrzykiwania kodu SQL jest dostępne w: *SQL Injection | OWASP Foundation* [na:] https://owasp.org/www-community/attacks/SQL_Injection, dostęp 28 października 2022 r.
[19] Więcej na temat XSS jest dostępne w: *Cross Site Scripting (XSS) | OWASP Foundation* [na:] https://owasp.org/www-community/attacks/xss/, dostęp 28 października 2022 r.
[20] Więcej na temat travelsar jest dostępne w *Path Traversal | OWASP Foundation* [na:] https://owasp.org/www-community/attacks/Path_Traversal, dostęp 28 października 2022 r.

denial of service [21]– jest to jedna z metod atakowania usług sieciowych, której podstawowym celem jest utrudnienie, bądź uniemożliwienie korzystania z atakowanych usług przez innych uprawnionych użytkowników.

allows remote attackers to execute – podatność polegająca na zdalnym uruchomieniu jakiegoś polecenia na serwerze lub komputerze ofiary. Tego typu ataki są możliwe w programach, gdzie walidacja danych wejściowych nie jest przeprowadzana lub poziom walidacji danych wejściowych jest niewystarczający[22].

Rysunek 12. Procentowy udział wybranych podatności na tle wszystkich pozostałych w latach 1999 - 2022 (trzy kwartały)



Źródło: opracowanie własne.

[21] Więcej na temat denial of service jest w: *Denial of Service | OWASP Foundation* [na:] https://owasp.org/www-community/attacks/Denial_of_Service, dostęp 28 października 2022 r.
[22] Więcej na ten temat jest dostępne w: *Command Injection | OWASP Foundation* [na:] https://owasp.org/www-community/attacks/Command_Injection, dostęp 28 października 2022 r.

Analizując dane źródłowe w opisach podatności CVE uwzględniając dane z kolejnych lat widać, że każda z nich charakteryzuje się dużą zmiennością. To co napawa optymizmem to fakt, że udział badanych podatności od 2019 roku jest znacznie mniejszy niż w latach poprzednich, co mogłoby sugerować, że podatność aplikacji na szkodliwe działania osób nieupoważnionych została ograniczona. Jednak, gdy te wyniki zestawimy z ogólną liczbą podatności w kolejnych latach, której trend jest rosnący, pojawia się pewna sprzeczność. Z jednej strony udział znanych od lat podatności w oprogramowaniu maleje pomimo, że sumaryczna liczba podatności wzrasta.

Pojawia się pytanie, co może być tego przyczyną. Istnieje podejrzenie, że pojawił się nowy element zagrożenia, którego udział w podatnościach wybija się podnad te przedstawione wczesniej (patrz: Rysunek 11).

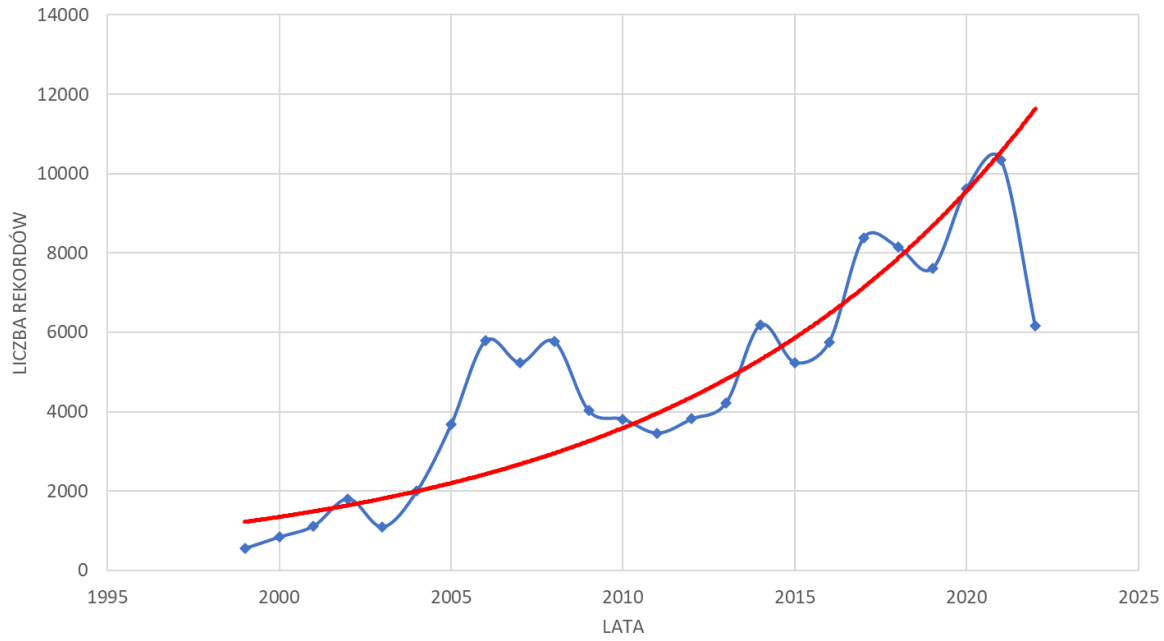
W związku z widocznym na wykresie (Rysunek 12) znaczącym zmniejszeniem udziału procentowego podatności określonej jako „denial of service" oraz w mniejszym stopniu pozostałych badanych podatności, obliczono częstotliwości występowania słów w opisach podatności z lat 2019-2022 (bez czwartego kwartału). Sposób analizy częstotliwościowej był analogiczny jak dla wcześniej opisanego pełnego okresu od 1999 roku.

Rysunek 13. Chmura najczęściej występujących wyrazów w opisach CVE w okresie od 2019 r. do 2022 r.

Źródło: opracowanie własne.

Tak jak poprzednio słowem wyróżniającym się spośród wszystkich jest vulnerability, a jako kolejne mające znaczący udział stanowią: attacker, version, allow. W związku z powyższymi wynikami słowo vulnerability oznaczające każdą podatność zostało pominięte, a do dalszej analizy liczby rekordów w kolejnych latach wybrano kolejny wyraz attacker (napastnik, agresor). Wynik został przedstawiony na poniższym wykresie.

Rysunek 14. Częstotliwość występowania słowa attacker w opisach podatności w rekordach CVE



Źródło: opracowanie własne.

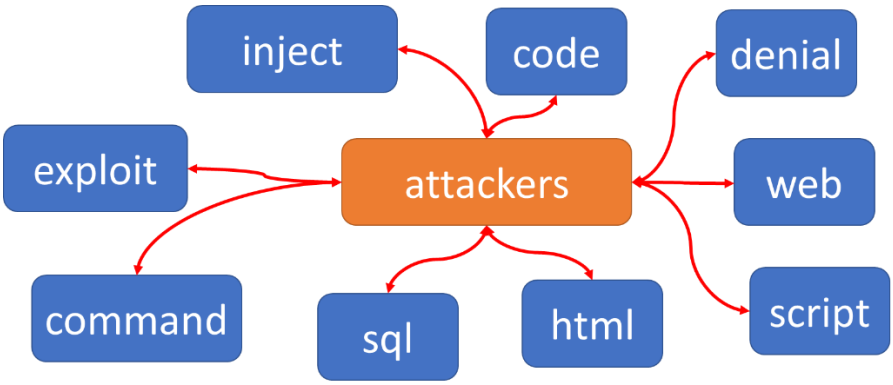
Otrzymany wynik wskazuje, że kluczową rolę w podatnościach odgrywa osoba występująca w roli agresora niezależnie od rodzaju podatności. Pozostałe version, allow nie wskazują na jakąś określoną metodę działania, poza tym, że odnoszą się do wersji oprogramowania oraz że coś jest dostępne na skutek czegoś co na danym etapie nie jesteśmy w stanie określić.

Dlatego też kolejnym krokiem było zbadanie jakie określenia wskazujące na podatność lub atak znajduje się w sąsiedztwie słowa attacker.

Do tego celu użyto system RING, który został zbudowany na potrzeby projektu badania trendów technologicznych w Katedrze Technologii Informacyjnych Mediów UW. W wyniku analizy

danych w systemie RING za cały okres od 1999 do 2022 roku (bez czwartego kwartału) ustalono, że w otoczeniu wyrazu attacker (po 60 znaków z lewej i prawej strony), często występującymi wyrazami były: inject, code, denial, web, script, html, sql, command, exploit.

Rysunek 15. Lista najczęściej występujących wyrazów wokół słowa attackers



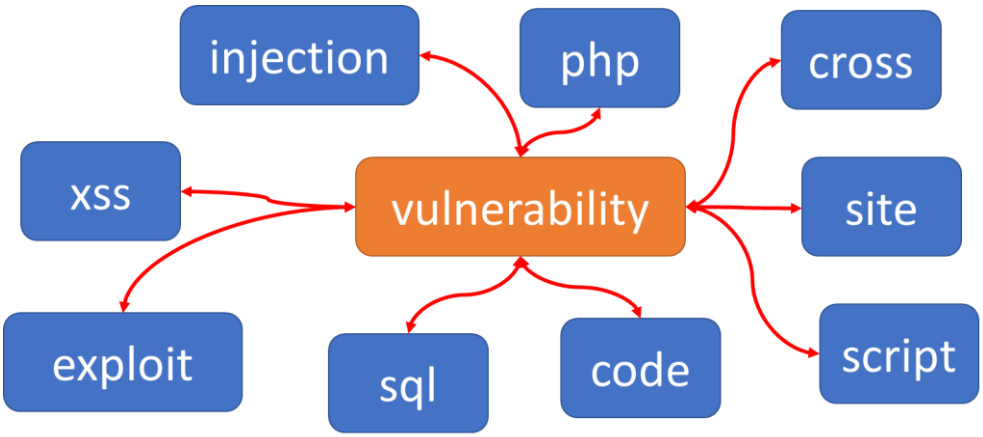
Źródło: opracowanie własne.

Na podstawie tak otrzymanych wyników można z pewnym prawdopodobieństwem przyjąć, że atakujący jako metodę ataku wykorzystują wstrzykiwanie kodu SQL (wyrazy: sql, inject). Wystąpienie słowa html i web świadczą o tym, że część ataków dotyczy stron internetowych. Słowo command wskazuje, że techniką wykorzystania podatności jest wykonanie polecenia na systemie posiadającym podatność. Występujące słowo denial oznacza, że przeprowadzone w wyniku ataku działania mogą doprowadzić do zatrzymania usługi sieciowej. Na uwagę zasługuje także wyraz exploit, który oznacza gotowy program komputerowy wykorzystujący jakąś podatność. Należy zatem przypuszczać, że jakaś część agresorów korzysta z gotowych narzędzi jak i również to, że do części podatności tworzone (code) jest oprogramowanie, które nawet niewprawnej osobie bez specjalistycznej wiedzy pozwoli zaatakować jakiś system. Jest to szczególnie groźna sytuacja, w której osoba bez posiadanej odpowiedniej wiedzy może wyrządzić znacznie więcej szkód niż świadomy hacker.

Patrząc na wszystkie wyrazy znajdujące się w najbliższym otoczeniu słowa attackers można wyciągnąć wniosek, że w większości przypadków celem ataków są serwisy internetowe i znajdujące się w tych serwisach podatności.

Bardzo podobne zestawienie słów otrzymano wokół wyrazu vulnerability, które świadczą o tym, że wektorem ataku są serwisy internetowe. W stosunku do otoczenia attackers mamy dodatkowo nowe wyrazy takie jak php, odnoszące się do najpopularniejszego języka używanego do konstrukcji stron WWW. Pojawia się także site, cross, script i xss wskazujące na metodę ataku polegającą na pewnego rodzaju wstrzyknięciu kodu najczęściej w Java Script do legalnie działającej strony WWW.

Rysunek 16. Lista najczęściej występujących wyrazów wokół słowa vulnerability



Źródło: opracowanie własne.

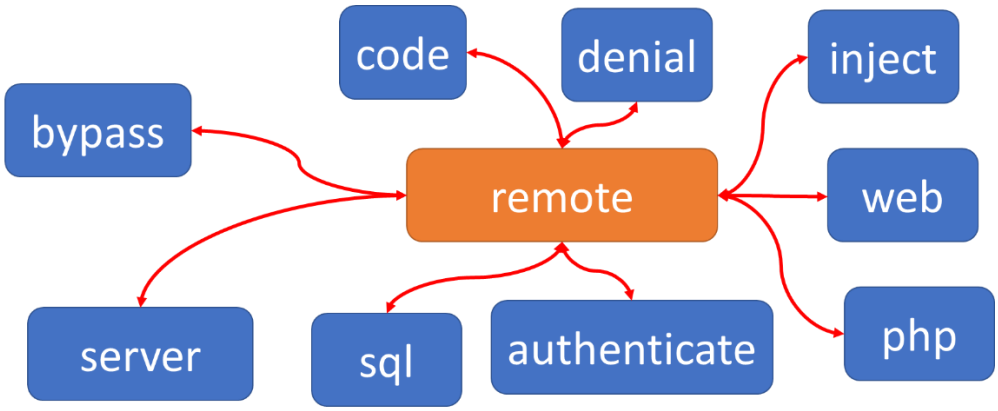
Omówione przykłady otoczenia słów skłaniają do sformułowania wniosku, że atakujący wykorzystują sieć jako medium, za pomocą którego próbują przejąć kontrolę nad zewnętrznymi systemami. Tym samym interesującym elementem jest analiza otoczenia słowa „remote". Poza już wcześniej wymienionymi słowami code, denial, web, php itd. pojawiają się trzy nowe słowa server, authenticate, bypass. W połączeniu wskazują m.in. na próbę obejścia zabezpieczeń systemu.

W kontekście ostatniego słowa „remote" warto spojrzeć na mapę bibliometryczną wygenerowaną za pomocą narzędzia VOSviewer[23] dla rekordów CVE z okresu od 2019 do 2022 (bez czwartego kwartału) ukazującą istotne klastry, słowa (skupienia) i powiązania pomiędzy nimi.

[23] *VOSviewer - Visualizing scientific landscapes* [na:] „VOSviewer", https://www.vosviewer.com//, dostęp 14 września 2022 r.

Na wykresie wyróżniającym się wyrażeniem jest „remote attacker” co potwierdza, że dominującą forma wykorzystania podatności (błędu) w oprogramowaniu dokonywana jest zdalnie za pośrednictwem sieci komputerowej. Z atakiem powiązany jest konkretny produkt, a ten z kolei z komponentem. Tu widać pewną analogię do przypadku krytycznej podatności j4Log (biblioteki/komponentu) omówionej na początku rozdziału.

Rysunek 17. Lista najczęściej występujących wyrazów wokół słowa remote



Źródło: opracowanie własne.

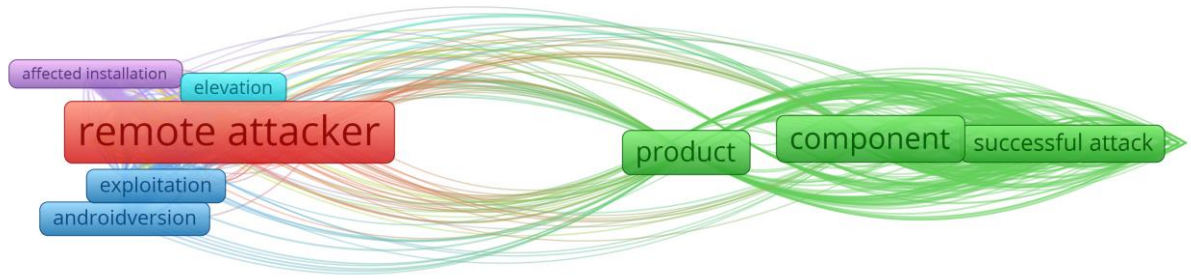
Słowo „component” jest nieco większy niż „product”. Można przyjąć, że ta różnica wynika z tego, że określony komponent może być wykorzystywany w wielu produktach. Niewykluczone, że w części opisów podatności w rekordach CVE używana jest nazwa konkretnego rozwiązania zamiast słowa produkt. Tuż obok na rysunku uwagę zwraca „successful attack” co pokazuje, że do przeprowadzenia skutecznego ataku potrzebne jest oprogramowanie i komponent posiadający podatność.

Podsumowując można zatem z pewnym prawdopodobieństwem przyjąć, że obecnie ataki zdalne realizowane za pomocą sieci komputerowej w ostatnich prawie czterech latach stanowią dominującą metodę ataku.

Postawiona na wstępie główna hipoteza badawcza, że na przestrzeni 24 lat, programiści stale popełniają błędy tej samej natury podczas tworzenia oprogramowania, przez co oprogramowanie posiada te same podatności, nie została potwierdzona. Z otrzymanych wyników nie widać, aby wskazane na wstępie opracowania rodzaje podatności jak np. buffer overflow była dominująca.

W kwestii drugiej hipotezy pomocniczej, że wektory ataku hackerów (crackerów) nie zmieniły się i są takie same dziś jak i 24 lata temu nie została także potwierdzona. Dziś widać wyraźnie, że dominującą metodą ataków jest atak zdalny poprzez sieć. Na drugi plan schodzi rodzaj wykorzystanej do ataku podatności, których różnorodność wzrosła, przez co przestają one być widoczne. Liczba podatności takich jak: buffer overflow, sql injection, xss i cross site scripting, traversal, denial of service; allows remote attackers to execute na przestrzeni lat zmalała (Rysunek 12) jednak liczba ataków w tym samym przedziale sumarycznie wzrosła (Rysunek 10). To oznacza, że pojawiły się nowe rodzaje podatności w stosunku do tych będących przedmiotem badania. Tym samym to także dowodzi to, że hipoteza pomocnicza nie została potwierdzona.

Rysunek 18. Mapa bibliometryczna dla wybranych do badania rekordów CVE z okresu od 2019 do 2022 (bez czwartego kwartału)



Źródło: opracowanie własne z użycie programu VOSwiever (https://www.vosviewer.com/).

Podsumowując można wyciągnąć następujące wnioski końcowe:

- współcześni programiści powinni posiadać wiedzę w zakresie wykorzystywania podatności w oprogramowaniu, których wektorem ataku jest sieć;
- nadal konieczna jest edukacja programistów w zakresie znanych od lat podatności takich jak: buffer overflow, sql injection, xss i cross site scripting, traversal;
- istotnym zadaniem na przyszłość jest zidentyfikowanie nowych rodzajów podatności, które pojawiły się na przestrzeni lat, w szczególności po 2019 roku;
- w miarę możliwości wskazane jest udoskonalanie języków programowania, kompilatorów i środowisk dla programistów, aby były bardziej „odporne” na znane już błędy popełniane przez programistów;

- administratorzy systemów komputerowych powinni na bieżąco aktualizować wykorzystywane w produkcji oprogramowanie;
- administratorzy powinni także zwracać baczną uwagę na konfigurację usług sieciowych. Rosnąca liczba ataków zdalnych świadczy o tym, że jest to dominująca metoda skompromitowania systemu. Tą drogą przestępca możne nie tylko skorzystać z podatności (błędu) w oprogramowaniu, ale także wykorzystać luki w konfiguracji oprogramowania.

Trywialnym przykładem takiej luki może być wystawienie usługi sieciowej na zewnątrz bez właściwych zabezpieczeń, które są dostępne w danym programie. Warto rozważyć, by oprogramowanie było coraz szerzej doposażone w dodatkowe zabezpieczenia na okoliczność błędnej konfiguracji wykonanej przez administratora. Już w tej chwili niektóre programy nie uruchomią się, jeśli założone warunki zabezpieczeń nie zostaną spełnione.

Przedmiotowa analiza pokazuje, że kwestie związane z tworzeniem bezpiecznych aplikacji są ciągle aktualne. Pomimo upływu lat, oprogramowanie zawiera błędy wynikające z ludzkiej natury. Część podatności znanych ponad ćwierć wieku temu nadal pojawiają się w oprogramowaniu. Rozwój sieci i usług zdalnych świadczonych przez serwery rozszerzył katalog potencjalnych podatności, które mogą znaleźć się w programach komputerowych. Optymistycznym aspektem, który może przyczynić się do podniesienia jakości oprogramowania w zakresie podatności, jest obecność kluczowych inicjatyw, do których należy m.in. inicjatywa Common Vulnerabilities and Exposures (CVE) oraz towarzyszące jej inne rozwiązania jak National Vulnerability Database - NVD), czy wspólny system oceny podatności na zagrożenia (Common Vulnerability Scoring System - CVSS). Tego typu inicjatywy i powstające obszerne repozytoria danych otwierają drogę do bardziej zaawansowanych analiz problematyki podatności programów komputerowych. Biorąc pod uwagę rosnącą rolę oprogramowania we współczesnym świecie, niniejsze opracowanie wskazuje potencjalną, choć nie jedyną, drogę dla tego typu analiz.

Bibliografia:

[1] *Ada Lovelace | Biography, Computer, & Facts | Britannica* [na:] https://www.britannica.com/biography/Ada-Lovelace, dostęp 11 listopada 2022 r.

[2] *Buffer Overflow | OWASP Foundation* [na:] https://owasp.org/www-community/vulnerabilities/Buffer_Overflow, dostęp 28 października 2022 r.

[3] *Charles Babbage | Biography, Computers, Inventions, & Facts | Britannica* [na:] https://www.britannica.com/biography/Charles-Babbage, dostęp 11 listopada 2022 r.

[4] *Command Injection | OWASP Foundation* [na:] https://owasp.org/www-community/attacks/Command_Injection, dostęp 28 października 2022 r.

[5] *Computer programming language - SQL | Britannica* [na:] https://www.britannica.com/technology/computer-programming-language/SQL, dostęp 11 listopada 2022 r.

[6] *Cross Site Scripting (XSS) | OWASP Foundation* [na:] https://owasp.org/www-community/attacks/xss/, dostęp 28 października 2022 r.

[7] *CVSS v3.1 Specification Document* [na:] „FIRST — Forum of Incident Response and Security Teams", https://www.first.org/cvss/v3.1/specification-document, dostęp 12 listopada 2022 r.

[8] *Denial of Service | OWASP Foundation* [na:] https://owasp.org/www-community/attacks/Denial_of_Service, dostęp 28 października 2022 r.

[9] *Home | CVE* [na:] https://www.cve.org/, dostęp 12 listopada 2022 r.

[10] *In Depth | Mars Climate Orbiter* [na:] „NASA Solar System Exploration", https://solarsystem.nasa.gov/missions/mars-climate-orbiter/in-depth, dostęp 12 listopada 2022 r.

[11] Levenson N., *Medical Devices: The Therac-25*, http://sunnyday.mit.edu/papers/therac.pdf.

[12] Lions J.-L., *ARIANE 5 Failure - Full Report* [na:] http://sunnyday.mit.edu/nasa-class/Ariane5-report.html, dostęp 12 listopada 2022 r.

[13] *List Of Partners | CVE* [na:] https://www.cve.org/PartnerInformation/ListofPartners, dostęp 12 listopada 2022 r.

[14] *Log4j vulnerability - what everyone needs to know* [na:] https://www.ncsc.gov.uk/information/log4j-vulnerability-what-everyone-needs-to-know, dostęp 12 listopada 2022 r.

[15] *MariaDB Foundation - MariaDB.org* [na:] https://mariadb.org/, dostęp 8 września 2022 r.

[16] *MAXQDA | All-In-One Qualitative & Mixed Methods Data Analysis Tool* [na:] „MAXQDA", https://www.maxqda.com/, dostęp 9 września 2022 r.

[17] *Mitre Corporation*, [w:] *Wikipedia*, 2022.

[18] *Overview | CVE* [na:] https://www.cve.org/About/Overview, dostęp 12 listopada 2022 r.

[19] *Path Traversal | OWASP Foundation* [na:] https://owasp.org/www-community/attacks/Path_Traversal, dostęp 28 października 2022 r.

[20] *Related Efforts | CVE* [na:] https://www.cve.org/About/RelatedEfforts, dostęp 12 listopada 2022 r.

[21] *Signal(7) - Linux manual page* [na:] https://www.man7.org/linux/man-pages/man7/signal.7.html, dostęp 23 września 2022 r.

[22] *SQL Injection | OWASP Foundation* [na:] https://owasp.org/www-community/attacks/SQL_Injection, dostęp 28 października 2022 r.

[23] *VOSviewer - Visualizing scientific landscapes* [na:] „VOSviewer", https://www.vosviewer.com//, dostęp 14 września 2022 r.

Bartłomiej Moszoro

Cyberbezpieczeństwo w firmie

Wstęp

Temat cyberbezpieczeństwa można rozpatrywać z różnych perspektyw. Jedna z nich to aspekt prawny, który narzuca przedsiębiorstwom pewne standardy. W Polsce jest to ustawa o krajowym systemie cyberbezpieczeństwa, która weszła w życie 28 sierpnia 2018 r. Wdraża ona do polskiego porządku prawnego Dyrektywę Parlamentu Europejskiego i Rady w sprawie środków na rzecz wysokiego, wspólnego poziomu bezpieczeństwa sieci i systemów informatycznych na terytorium Unii Europejskiej, zwaną Dyrektywą NIS.

W ustawie tej zdefiniowano cyberbezpieczeństwo jako „odporność systemów informacyjnych na wszelkie działania naruszające poufność, integralność, dostępność i autentyczność przetwarzanych danych lub związanych z nimi usług oferowanych przez te systemy". Aby dobrze zrozumieć obowiązki ustawodawcy należy uściślić pojęcie „systemy informacyjne". Przekładając na język praktyczny, „systemem informacyjnym firmy są używane w tej firmie serwery, komputery, telefony, tablety (i inne podobne urządzenia przenośne) wraz z oprogramowaniem służącym do ich używania, połączone siecią teleinformatyczną (w praktyce, nieco trywializując, połączone telekomunikacyjnie, sieciowo, internetowo) oraz dane, które są przetwarzane za pomocą tych systemów." [1]

Dalszymi elementami definicji cyberbezpieczeństwa są pojęcia ściśle związane z charakterystyką dobrze działających systemów informacyjnych, tj.: poufność danych

– tzn. zapewnienie, że dane nie są ujawniane w sposób nieautoryzowany; integralność danych – czyli zapewnienie ich kompletności i dokładności; dostępność danych – tj. zapewnienie, że dane są dostępne dla każdego z firmy w miejscu i czasie, w jakim są potrzebne; autentyczność danych – kryterium pewności, że przetwarzane dane są prawdziwe, tj. są danymi, które w sposób autoryzowany zostały wprowadzone do systemu. W definicji cyberbezpieczeństwa najważniejsze jest określenie przedmiotu ochrony, tj. danych lub związanych z nimi usług.

Ustawa ustanawia krajowy system cyberbezpieczeństwa, którego zadaniem jest zapewnienie cyberbezpieczeństwa na poziomie krajowym. Tworzy on sieć zespołów reagowania na incydenty oraz wyznacza zadania odpowiednim organom. Ponadto ustawa nakłada konkretne obowiązki operatorom usług kluczowych, dostawcom usług cyfrowych oraz podmiotom świadczącym usługi z zakresu cyberbezpieczeństwa i podmiotom publicznym, w tym m.in.: jednostkom sektora finansów publicznych, instytutom badawczym; NBP, BGK, Urzędowi Dozoru Technicznego; Narodowemu Funduszowi Ochrony Środowiska i Gospodarki Wodnej oraz wojewódzkim funduszom ochrony środowiska i gospodarki wodnej oraz spółkom prawa handlowego wykonującym zadania o charakterze użyteczności publicznej. Powinni one wdrożyć system zarządzania bezpieczeństwem, który wiąże się przede wszystkim z koniecznością prowadzenia systematycznego szacowania ryzyka wystąpienia incydentu i dostosowania do niego środków bezpieczeństwa, takich jak bezpieczna eksploatacja systemu, bezpieczeństwo fizyczne systemu (w tym kontrola dostępu), bezpieczeństwo i ciągłość dostaw usług, które mają wpływ na świadczenie usługi kluczowej, utrzymanie planów działania umożliwiających ciągłość świadczenia usług, ciągłe monitorowanie systemu zapewniającego świadczenie usług. Audyty bezpieczeństwa systemu informacyjnego mają być przeprowadzone regularnie i w określonym odstępie czasowym, operatorzy są nadzorowani przez konkretne organy, uprawnione do przeprowadzenia kontroli i nakładania kar pieniężnych przewidzianych w ustawie.

Pozostałe przedsiębiorstwa, czyli wszystkie firmy niewchodzące w grupę dostawców kluczowych czy operatorów usług kluczowych nie mają obowiązków wprowadzenia powyżej opisanych standardów stawianych przez ustawę o krajowym systemie cyberbezpieczeństwa.

Rola dyrektora ds. bezpieczeństwa informacji (CISO)

Zarządzanie przedsiębiorstwem w obecnym świecie wymaga wypracowania podejścia do tematyki cyberbezpieczeństwa. Istnieje szereg światowych standardów, których celem jest

sformalizowanie i zapewnienie wspólnych ram dla oceny i zarządzania ryzykiem cybernetycznym. Jednym z nich jest „CyBOK Cyber Security Body Of Knowledge" [7]. Aktualna wersja podręcznika zawiera podstawowe pojęcia i praktyki w tej dziedzinie. Zaprezentowane tam podejście zaczyna się od „Governance, risk management, and compliance" (GRC) – czyli strategii zarządzania lub ładu korporacyjnego, kontroli ryzyka oraz zgodności z przepisami. Strategia przedsiębiorstwa określająca procesy, strukturę i narzędzia wykorzystywane do kierowania i zarządzania wszystkimi działaniami organizacji powinna uwzględniać bezpieczeństwo informacji.

W celu odpowiedniego zabezpieczania się przed cyfrowym niebezpieczeństwem należy najpierw zrozumieć istotę przedsiębiorstwa. Następujące pojęcia pomogą zrozumieć proces zarządzania bezpieczeństwem informacyjnym:

- Business drivers – główne obszary działalności;
- Information security drivers – czynniki wpływające na bezpieczeństwo informacji;
- Risk management – zarządzanie ryzykiem.

Główne obszary działalności – pod tym pojęciem znajduje się odpowiedź na pytanie, dlaczego organizacja istnieje i jaką prowadzi działalność. Od struktury organizacyjnej, jej hierarchiczności oraz branży, w której firma działa i jej dojrzałość, zależy sposób zarządzania przedsiębiorstwem. Modele mogą się różnić w zależności od typu własności (od jednego właściciela, przez partnerów aż do udziałowców i dużej korporacji).

Kolejnym krokiem jest zaprojektowanie procesów biznesowych i technologii informacyjnych, aby wspierały te cele organizacji. Biorąc pod uwagę również takie zagrożenia, które mogą zakłócić te procesy. Często w strukturze organizacyjnej większych korporacji pojawia się nowa rola związana z bezpieczeństwem informacyjnym. Nazwy i zakres odpowiedzialności mogą być różne. Poniżej przedstawiono typowo występujące stanowiska osób odpowiedzialnych za obszar cyberbezpieczeństwa w przedsiębiorstwach - nazwy angielskojęzyczne z odpowiednikami w języku polskim. Osoby te mogą wchodzić w skład zarządu lub być poza nim jako specjaliści w danym zakresie z funkcją doradczą.

Wiele z obszarów działań dyrektora ds. bezpieczeństwa (CISO) wiąże się z koniecznością przestrzegania przepisów prawa, regulacji i standardów. Te regulacje stanowią podstawę do opracowania polityki, standardów i procedur bezpieczeństwa informacji w organizacji.

Tab. 1. Przykładowe role w przedsiębiorstwie związane z cyberbezpieczeństwem

CISO - Chief information security Officer	Dyrektor ds. bezpieczeństwa informacji (CISO)
CIO – Chief Information Officer	Dyrektor ds. informatyki lub dyrektor IT, odpowiada za stan, rozwój i wdrożenia technologii informacyjnych.
CSO – Chief Security Officer	Dyrektor ds. bezpieczeństwa, członek kadry zarządzającej odpowiedzialny za fizyczne i cyfrowe bezpieczeństwo firmy.
CTO – Chief Technology Officer	Dyrektor ds. technologii, nadzoruje rozwój i poprawność działania systemów informatycznych z punktu widzenia realizacji strategii firmy.
CDO – Chief Data Officer	Dyrektor ds. danych, osoba na szczeblu zarządczym, która jest odpowiedzialna za wykorzystanie danych i zarządzanie nimi.
CRO – Chief Risk Officer	Dyrektor ds. ryzyka, stara się zapewnić zgodność firmy z regulacjami prawnymi.

Źródło: opracowanie własne.

Praktycznym mechanizmem umożliwiającym takie działanie może być stworzenie programu zarządzania bezpieczeństwem informacji lub sformułowanie polityki bezpieczeństwa. Jest to ogólna strategia, określająca w jaki sposób firma będzie wdrażać zasady i technologie bezpieczeństwa informacji. Można powiedzieć, że w zasadzie jest to biznesplan, który dotyczy aspektów bezpieczeństwa informacji w firmie.

Polityka bezpieczeństwa musi w szczególności odnosić się do następujących kwestii:

- zakres, w jakim każdy pracownik ma udział w bezpieczeństwie informacji organizacji,
- poufność i prywatność informacji,
- zasadę najmniejszego dostępu do informacji,
- integralność informacji,
- dostępność informacji,
- równowagę między narażeniem na ryzyko a kosztami jego ograniczenia,

- wdrożenie środków bezpieczeństwa,
- klasyfikacja informacji,
- znaczenie świadomości bezpieczeństwa i zarządzania informacjami.

Ze względu na fakt, że istnieją róże typu organizacji, istnieją także różne standardy i wiodące praktyki w zakresie polityki bezpieczeństwa. Jedna praktyka może być odpowiednia dla firmy z sektora opieki zdrowotnej skupionej na utrzymaniu prywatności pacjentów, jednak ta sama praktyka może być nieodpowiednia dla firmy z sektora portali społecznościowych opartych na idei dzielenia się danymi osobowymi pomiędzy klientami. Dyrektor ds. bezpieczeństwa CISO dba o proces wyboru odpowiednich standardów i ram oraz tworzenia najlepszych praktyk dla swojej organizacji.

Zarządzanie ryzykiem polega na identyfikacji, ocenie i ustaleniu priorytetów ryzyka, a następnie skoordynowanym i oszczędnym wykorzystaniu zasobów w celu zminimalizowania, monitorowania i kontrolowania prawdopodobieństwa i/lub wpływu niefortunnych zdarzeń lub maksymalizacji wykorzystania szans. Rola dyrektora ds. bezpieczeństwa CISO jest bardzo złożona. Obraca się wokół kwestii wewnętrznej polityki organizacyjnej, odnalezienia wspólnego języka w zarządzie, ograniczeń budżetowych, koncepcji własności ryzyka oraz gotowości kierownictwa wyższego szczebla - nawet zarządu - do przyjęcia odpowiedzialności za zarządzanie ryzykiem (risk treatment). Stanowisko takie skierowane jest także na realizację wymagań nałożonych przez wszystkich różnych interesariuszy zaangażowanych w proces zarządzania ryzykiem.

Cyberbezpieczeństwo w polskich przedsiębiorstwach: obecny stan badań

Mając na uwadze prawne uwarunkowania narzucające niektórym polskim przedsiębiorstwom standardy działania cyberbezpieczeństwa oraz opisane praktyki cyberbezpieczeństwa, będące codziennym rozwiązaniem dla wielu firm globalnych, przyjrzymy się teraz najnowszym badaniom dotyczącym cyberbezpieczeństwa w polskich przedsiębiorstwach.

Pierwszy raport to „Barometr cyberbezpieczeństwa. Ochrona cyfrowej tożsamości". Przygotowany został przez z firmę KPMG [4,2]. Badanie zostało przeprowadzone w maju 2022 r. Zrealizowane było metodą wywiadów telefonicznych CATI wśród osób odpowiedzialnych za bezpieczeństwo IT (członków zarządu, dyrektorów s.. bezpieczeństwa, prezesów, dyrektorów

IT lub innych osób odpowiedzialnych za ten obszar) w 100 firmach o przychodach powyżej 50 mln zł.

69% badanych firm odnotowało w 2021 r. przynajmniej jeden cyberatak polegający na naruszeniu danych. Oznacza to niewielki wzrost, tj. 5% prób cyberataków w porównaniu do 2020 roku. Warto zaznaczyć, że ponad dwukrotnie wzrosła liczba firm, które zaobserwowały 30 i więcej incydentów bezpieczeństwa, co może świadczyć o wzmożonej aktywności cyberprzestępców.

Z badania wynika, że organizacje biorące w nim udział zadeklarowały, iż największym cyberzagrożeniem są dla nich wycieki danych za pośrednictwem złośliwego oprogramowania (*malware*) oraz *phishing* – czyli wyłudzenia danych uwierzytelniających. Przedsiębiorstwa najbardziej obawiają się zagrożeń ze strony szeroko rozumianej cyberprzestępczości, na którą wskazuje 92% badanych, z czego blisko 70% największego zagrożenia upatruje w zorganizowanych grupach cyberprzestępczych. Ponadto przedsiębiorstwa wyraźnie obawiają się tak zaawansowanych ataków ze strony profesjonalistów, jak i kradzieży danych przez pracowników.

Bardzo interesujący jest wgląd w motywację przedsiębiorców w obszarze cyberbezpieczeństwa. Dla specjalistów odpowiedzialnych w firmach za bezpieczeństwo informacji najważniejszym czynnikiem wpływającym na decyzję rozpoczęcia inwestycji w procesy zarządzania tożsamością i dostępem jest osiągnięcie wzrostu bezpieczeństwa przetwarzanych informacji – na co wskazało 73% respondentów. Dla 57% jedną z głównych motywacji okazała się potrzeba zapewnienia zgodności z regulacjami, a kolejne 41% widzi w takich inwestycjach szansę na optymalizację kosztową w efekcie odciążenia personelu. Inną motywacją do zabezpieczania się stanowią profity z nim związane. Mniej więcej co trzecia firma wśród najważniejszych potencjalnych korzyści wymienia zwiększenie jakości zarządzania uprawnieniami lub wygodę użytkowników – jak wskazano w badaniu.

Stan bezpieczeństwa cybernetycznego polskich firm nie jest – w świetle wyników tych badań – tragiczny. Ponad trzy czwarte respondentów na początku 2022 roku zadeklarowało pełną dojrzałość zabezpieczeń w połowie analizowanych obszarów. Ale już tylko 19% firm deklaruje dojrzałość zabezpieczeń w większości analizowanych obszarów, a jedynie 4% we wszystkich.

Ciekawym obszarem pozwalającym zrozumieć złożoność zagadnienia są trudności we wprowadzeniu praktyk cyberbezpieczeństwa. Jak wynika z badania, 64% firm przyznało, iż największą barierą utrudniającą budowanie odpowiedniego poziomu zabezpieczeń są trudności w znalezieniu oraz utrzymaniu odpowiednio wykwalifikowanych pracowników. Oznacza to wzrost aż o 14 pkt. procentowych w stosunku do poprzedniej edycji badania. Jak można było się również spodziewać, 57% organizacji deklaruje, że problemem jest brak wystarczających budżetów.

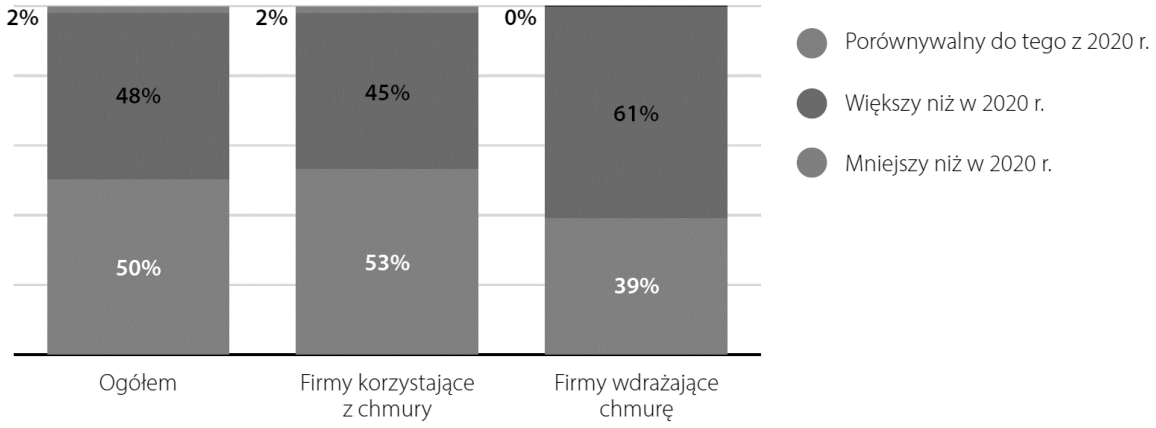
Istotną kwestią w zarządzaniu ryzykiem jest sama decyzja rozpoczęcia inwestycji. Dla specjalistów odpowiedzialnych w firmach za bezpieczeństwo informacji najważniejszym czynnikiem wpływającym na decyzję rozpoczęcia inwestycji w procesy zarządzania tożsamością i dostępem jest osiągnięcie wzrostu bezpieczeństwa przetwarzanych informacji – na które wskazało 73% respondentów. Dla 57% jedną z głównych motywacji okazała się potrzeba zapewnienia zgodności z regulacjami, a kolejne 41% widzi w takich inwestycjach szansę na optymalizację kosztową w efekcie odciążenia personelu.

Kolejnym raportem przedstawiającym stan polskich firm jest „Zwarci, silni, gotowi? Polskie firmy w obliczu cyberzagrożeń", opracowany przez CyberDefence24.pl oraz DGTL Kibil Piecuch i Wspólnicy [3]. Wyniki tego raportu pokazują aktualność obaw ze strony przedsiębiorców dotyczących cyberataków jak i rangę tego problemu. Prawie 2/3 respondentów wskazało, że niebezpieczeństwa czyhające w sieci stanowią jedno z trzech najważniejszych ryzyk w ich działalności. Badania ukazały też fakt, że na rynku brakuje obecnie jednej dominującej wizji obszaru zajmującego się cyberbezpieczeństwem firmy. Blisko jedna trzecia ankietowanych przedsiębiorstw nie posiada wyodrębnionej struktury organizacyjnej odpowiedzialnej za ten obszar funkcjonowania organizacji. Pozytywną wiadomością wynikającą z badania jest, że 70% firm wprowadziło Business Continuity Plan (BCP) uwzględniający planowanie wznowienie działalności firmy w razie wystąpienia cyberataku, a prawie 50% wdrożyło u siebie normę ISO27001.

W raporcie podano, że 70% badanych firm odnotowało incydenty z zakresu cyberbezpieczeństwa. Przy czym, fakt że 30% respondentów takich incydentów nie odnotowało, nie musi oznaczać, że nie padły one ofiarami takich ataków. Średni czas na wykrycie incydentów oraz zlikwidowanie jego skutków zajmuje bowiem średnio 287 dni. Co można odbierać jako bardzo długi okres. Badania wskazały, że firmy decydują się także na

przeprowadzanie testów. Mowa tu o jednorazowym sprawdzeniu pod kątem bezpieczeństwa nowych usług lub produktów (86,5%) oraz okresowych testów cyberbezpieczeństwa (92,3%), np. skanów podatności czy testów penetracyjnych systemów informatycznych. Coraz powszechniejszą praktyką są szkolenia z cyberbezpieczeństwa dla pracowników. W przypadku ankietowanych firm polegały one na wystąpieniach specjalistów, którzy omawiali pojawiające się w sieci cyberzagrożenia i sposoby walki z nimi. Niemniej jednak, aż 15% przedsiębiorstw nie przeprowadza w ogóle szkoleń z cyberbezpieczeństwa. Warto również zaznaczyć, że dla blisko 80% ankietowanych prowadzenie biznesu w oderwaniu od Internetu i technologii jest niemożliwe.

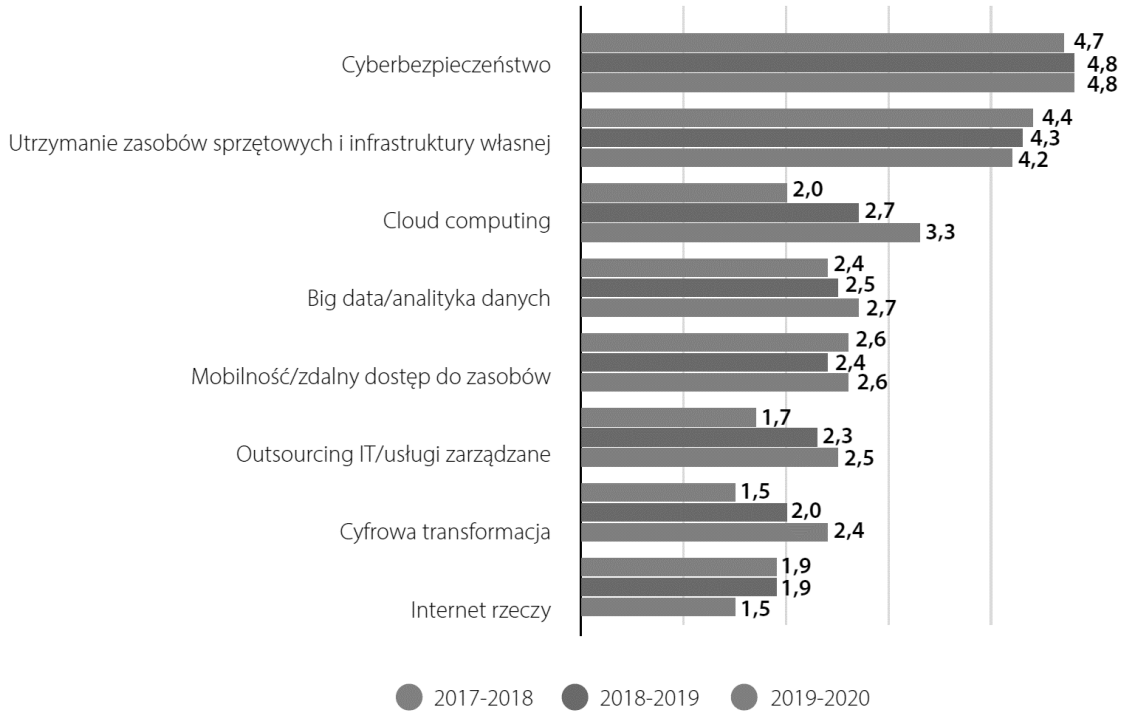
Rys. 1. Zmiany w budżecie na obszar cyberbezpieczeństwa w Polsce w opinii średnich i dużych przedsiębiorstw (%), 2021. Źródło: PMR



Źródło: opracowanie własne.

Innym cennym źródłem informacji na temat cyberbezpieczeństwa polskich firm jest raport PMR i Netii „Chmura i cyberbezpieczeństw w Polsce 2021"[5]. Z raportu wynika, że 54% badanych wskazuje poprawę bezpieczeństwa danych jako jedną z głównych korzyści z wdrożenia rozwiązań chmurowych. Taki kierunek działań jest potwierdzony przez 80% respondentów, którzy zadeklarowali inwestycje w kolejnych latach w rozwiązania chmurowe. Mają one m.in. zwiększać ochronę zasobów firmy i zapewniać jej ciągłość działania na wypadek incydentu.

Rys. 2. Priorytetowe działania z obszaru IT dla dużych przedsiębiorstw w Polsce w 2017-20220. Źródło: PMR



Źródło: opracowanie własne.

Ostatnie badanie, które zostaną przedstawione to raport PARP o stanie sektora MŚP w Polsce, wg danych GUS dotyczących sektora telekomunikacyjnego i cyberbezpieczeństwa (sierpień 2021) [6]. Jest to bardzo szczegółowe źródło informacji, które pozwala wyciągnąć wnioski nie tylko o realności zagrożeń, ale także z jakich usług z zakresu cyberbezpieczeństwa korzystają polskie firmy. Według danych prezentowanych przez GUS, na koniec września 2021 r., w branży telekomunikacji i cyberbezpieczeństwa działało 15 438 firm. Pozytywnym aspektem wskazanym przez przedsiębiorców (56% z sektora telekomunikacji oraz 71% z sektora cyberbezpieczeństwa) był wzrost liczby usług świadczonych przez firmy.

Rys. 3. Przedsiębiorstwa w Polsce, w których zadania związane z cyberbezpieczeństwem wykonywane były przez pracowników lub podmioty zewnętrzne, 2020. Źródło: GUS

Wyszczególnione	Zadania z zakresu cyberbezpieczeństwa wykonywane przez pracowników wewnętrznych	Zadania z zakresu cyberbezpieczeństwa wykonywane przez podmioty zewnętrzne
Ogółem	34,1%	68,6%
Duże przedsiębiorstwa	83,1%	76,2%
Średnie przedsiębiorstwa	48,2%	76,3%
Małe przedsiębiorstwa	29,4%	66,9%

Źródło: opracowanie własne.

W połowie przedsiębiorstw z sektora telekomunikacji oraz trzech na pięciu z sektora cyberbezpieczeństwa, w czasie pandemii nawiązano także współpracę z nowymi partnerami biznesowymi. 95% badanych przedsiębiorstw zastosowało środki bezpieczeństwa teleinformatycznego. Odnotowuje się duży brak specjalistów. W większości przedsiębiorstw pojawiały się problemy związane ze znalezieniem odpowiedniego wykwalifikowanego pracownika. W sektorze telekomunikacji wynikały one głównie z braku odpowiednich kompetencji, których oczekiwano w przedsiębiorstwach od kandydatów, natomiast w sektorze cyberbezpieczeństwa, z niewielkiego zainteresowania ofertą pracy.

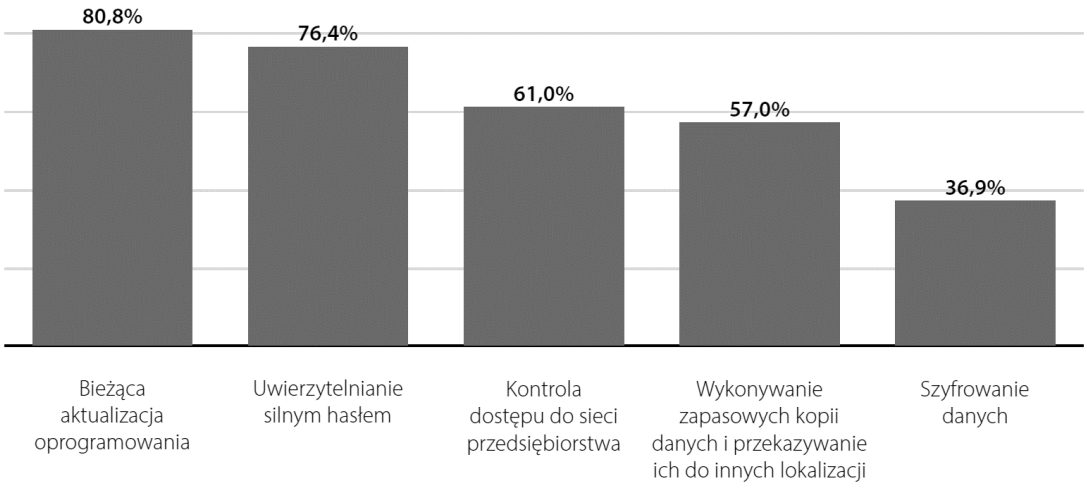
Rys. 4. Przedsiębiorstwa w Polsce doświadczające następstw incydentów związanych z cyberbezpieczeństwem, 2020. Źródło: GUS

Wyszczególnione	Niemożność korzystania z zasobów	Zniszczenie lub uszkodzenie danych	Ujawnienie poufnytch danych
Ogółem	8,6%	7,9%	1,2%
Duże przedsiębiorstwa	16,4%	16,9%	3,0%
Średnie przedsiębiorstwa	12,2%	11,6%	1,4%
Małe przedsiębiorstwa	7,5%	6,8%	1,1%

Źródło: opracowanie własne.

W 2020 r. odsetek przedsiębiorstw stosujących jakiekolwiek środki bezpieczeństwa ICT wyniósł 95,2%. Tego rodzaju środki wykorzystywano najczęściej w dużych przedsiębiorstwach (99,6%), a biorąc pod uwagę rodzaj prowadzonej działalności – w sekcji finanse i ubezpieczenia (99,5%).

Rys. 5. Przedsiębiorstwa w Polsce stosujące środki bezpieczeństwa ICT według wybranych rodzajów środków, 2020. Źródło: GUS



Źródło: opracowanie własne.

Najczęściej stosowanymi środkami bezpieczeństwa ICT w Polsce były bieżąca aktualizacja oprogramowania oraz uwierzytelnianie silnym hasłem (odpowiednio 83,1% i 78%). Mniej popularne jest wykonywanie zapasowych kopii danych i przekazywanie ich do innych lokalizacji (61,3%). Natomiast najrzadziej korzystano z identyfikacji i uwierzytelniania metodami biometrycznymi (7,1%). W 2020 r. 28,4% przedsiębiorstw przeprowadziło audyt bezpieczeństwa systemu informacyjnego firmy. Najczęściej przeprowadzały je podmioty duże (70,8%), a najrzadziej małe (23,7%).

Podsumowanie

Cyberbezpieczeństwo jest tak szybko rozwijającą się dziedziną, że musimy zaakceptować, że nie możemy być w pełni bezpieczni w cyberprzestrzeni. Polskie przedsiębiorstwa doświadczają przyrost cyberataków. Największym cyberzagrożeniem są dla nich wycieki danych za pośrednictwem złośliwego oprogramowania (*malware*) oraz *phishing* – czyli wyłudzenia danych uwierzytelniających. Wielu firmom brakuje zasobów do zabezpieczania wszystkich obszarów działalności a nikła część jest w pełni gotowa we wszystkich obszarach.

Wciąż brakuje specjalistów i firmy nie mogą skompletować swojej kadry. Coraz bardziej powszechne stają się szkolenia z cyberbezpieczeństwa dla pracowników. Niemniej, istnieje też grupa przedsiębiorstw, która w ogóle nie przeprowadza szkoleń z cyberbezpieczeństwa.

Pomimo faktu, że wzrost bezpieczeństwa przetwarzanych informacji stanowi główny czynnik skłaniający do inwestycji w bezpieczeństwo, wiele firm, która się na to decyduje myśli przede wszystkim o spełnieniu zgodności z regulacjami (compliance).

Cyberbezpieczeństwo nadal jest uważane za kwestię techniczną. Zrozumienie cyberbezpieczeństwa na poziomie strategicznym przyczyni się do lepszego dopasowania struktury zajmującej się tym obszarem w przedsiębiorstwie. Tworzenie „mostu" pomiędzy biznesem a działaniami z zakresu bezpieczeństwa danych zwiększy odporność polskich firm na cyberataki.

Bibliografia

[1] Dziomdziora, W. „Cyberbezpieczeństwo w firmie. Ustawa o krajowym systemie cyberbezpieczeństwa wdrażająca Dyrektywę NIS" https://www.parp.gov.pl/component/content/article/54056:cyberbezpieczenstwo-w-firmie-ustawa-o-krajowym-systemie-cyberbezpieczenstwa-wdrazajaca-dyrektywe-nis [14.09.2022].

[2] Grzegórska-Szpyt L., „Większość firm w Polsce odnotowało w 2021 r. przynajmniej jeden cyberatak", Gazeta Prawna, https://www.gazetaprawna.pl/wiadomosci/kraj/artykuly/8420387,firmy-w-polsce-2021-r-odnotowany-cyberatak-kpmg-cybezbezpieczenstwo.html [7.09.2022].

[3] Raport DGTL „Zwarci, silni, gotowi? Polskie firmy w obliczu cyberzagrożeń.", https://dgtl.law/publikacje/cyber-security/ [8.09.2022].

[4] Raport KPMG „Barometr cyberbezpieczeństwa. Ochrona tożsamości cyfrowej", 2022 https://home.kpmg/pl/pl/home/insights/2022/05/barometr-cyberbezpieczenstwa-ochrona-cyfrowej-tozsamosci.html [8.09.2022].

[5] Raport Netia Chmura i cyberbezpieczeństwo w Polsce – 2021, https://www.netia.pl/pl/srednie-i-duze-firmy/produkty/bezpieczenstwo [8.09.2022] https://www.netia.pl/pl/srednie-i-duze-firmy/lp/pobierz-raport-pmr [8.09.2022].

[6] Raport PARP o stanie sektora MŚP w Polsce, https://www.parp.gov.pl/component/content/article/80661:branza-telekomunikacji-i-cyberbezpieczenstwa-premiera-wynikow-badan-i-edycji-branzowego-bilansu-kapitalu-ludzkiego [8.09.2022].

[7] The Cyber Security Body Of Knowledge www.cybok.org. Obecnie najnowsza wersja (Version 1.1.0) dostępna na tej stronie jest datowana na 31.07.2021 r.

Wioletta Matosek, Agnieszka Heba

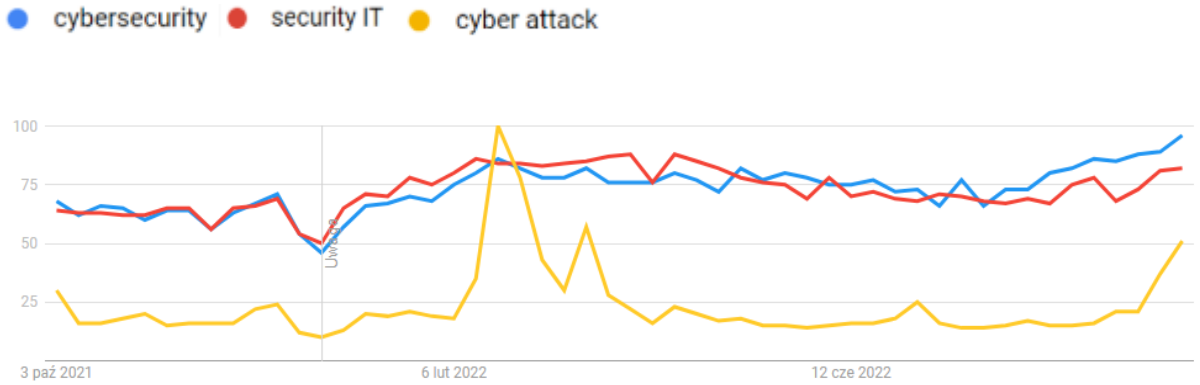
Bezpieczeństwo cyberprzestrzeni - badania nad nowym kierunkiem studiów

Wstęp

Cyberbezpieczeństwo jest obszarem, którego znaczenie stale rośnie. Na świecie obserwuje się wzrost zainteresowania tematyką związaną z bezpieczeństwem IT. Trendy wzrostu popularności wyszukiwania słów: „cybersecurity", „security IT", „cyber attacks" w wyszukiwarce Google w ciągu ostatnich dwóch i pięciu lat przedstawia Wyk. 1. i Wyk. 2.[2] Wzrost znaczenia bezpieczeństwa informatycznego, powoduje wzrost zapotrzebowania na specjalistów security IT.

Program studiów jest odpowiedzią na rosnące wciąż zapotrzebowanie na wykwalifikowanych pracowników zajmujących się bezpieczeństwem IT. Główne założenia programu skierowane są na rozpoznawanie potrzeb organizacji w zakresie bezpieczeństwa IT, kształtowanie polityki cyberbezpieczeństwa na poziomie państwowym oraz zarządzanie bezpieczeństwem informacyjnym UE. W systemie cyberbezpieczeństwa najsłabszym ogniwem wciąż pozostaje człowiek, dlatego tak ważną kwestią jest cyberedukacja. Najbardziej zaawansowane technologie z zakresu bezpieczeństwa IT mogą okazać się nieskuteczne w przypadku braku wystarczającej wiedzy użytkowników w zakresie bezpiecznego korzystania z narzędzi informatycznych i rozwiązań sieciowych.

Wyk. 1. Trendy popularności wyszukiwania: cybersecurity, security IT, cyber attacks w okresie dwóch lat



Źródło: opracowanie własne.

Program ukierunkowany jest na rozwój kompetencji w zakresie podnoszenia poziomu świadomości występowania cyberzagrożeń i możliwości zapobiegania ich gospodarczym, społecznym, psychologicznym i politycznym konsekwencjom. Umożliwia zdobycie wiedzy z zakresu szerokiego spektrum rozwiązań technologicznych security IT, niezbędnej do definiowania zagrożeń w cyberprzestrzeni i stosowania środków zapobiegawczych.

Wyk. 2. Trendy popularności wyszukiwania: cybersecurity, security IT, cyber attacks w okresie pięciu lat

Źródło: opracowanie własne.

Studenci nauczą się diagnozować i analizować zagrożenia związane z bezpieczeństwem cyberprzestrzeni, a także stosować narzędzia służące do ich ograniczania i eliminacji. Kształtowanie umiejętności zawodowych będzie miało na celu sprostanie stale zmieniającym się wymaganiom dzisiejszego cyfrowego świata.

Związek koncepcji i celów kształcenia na projektowanym kierunku studiów ze strategią UW w obszarze studiów

Planowany kierunek studiów II stopnia jest zgodny z generalnymi założeniami strategii UW, czyli czynnikami, które stanowią o sile i prestiżu uczelni. Nauczyciele akademiccy przewidywani do kształcenia na nowo tworzonym kierunku tworzą silny zespół, który oprócz licznych osiągnięć naukowych i dydaktycznych posiada duże doświadczenie w realizacji projektów B+R. Zarys programu studiów opracowany został po szczegółowej analizie takich samych lub podobnych kierunków studiów prowadzonych na innych uczelniach krajowych i zagranicznych. W ramach wymiany doświadczeń zaplanowano również wizytę studyjną na dwóch uniwersytetach: Masarykova Univerzita i Univerzita Obrany w Brnie. Celem wizyty jest zapoznanie się z aspektami funkcjonowania studiów z zakresu cyberbezpieczeństwa takimi jak: działalność naukowo-badawcza, realizowane projekty, współpraca z otoczeniem gospodarczym oraz opracowanie katalogu metod wykorzystywanych do nauczania przedmiotów. Wizyta służyła będzie integracji środowiska akademickiego przez nawiązanie wzajemnej współpracy dotyczącej wymiany doświadczeń naukowych, udziału we wspólnych konferencjach naukowych.

Zadaniem projektowanego kierunku jest kształcenie absolwentów liczących się na rynku pracy. Projektowany program jest odpowiedzią na zapotrzebowanie społeczne i gospodarcze. Gwałtowny w ostatnich latach wzrost popularności pracy zdalnej i hybrydowej spowodował wzrost zapotrzebowania firm i organizacji na specjalistów znających zasady polityki cyberbezpieczeństwa i sposoby jej kształtowania, mających wiedzę z zakresu stosowania mechanizmów, technologii i systemów zabezpieczeń przed cyberzagrożeniami. Praca zdalna obnażyła niedostateczny poziom bezpieczeństwa w wielu firmach i spowodowała konieczność edukowania rynku, uświadamiania o problemach związanych z zagrożeniami w obszarze IT i zwiększania poziomu świadomości wśród kadry managerskiej i specjalistów. Relatywnie niska świadomość rynkowa dotycząca konieczności stosowania rozwiązań z zakresu

cyberbezpieczeństwa połączona ze wzrostem zapotrzebowania na pracę zdalną i rozwiązania chmurowe oraz szybki wzrost rynku IoT, wymusiły konieczność zwiększania świadomości o bezpieczeństwie danych. Przeciętny obywatel nie wie, w jaki sposób chronić swoje dane i prywatność w sieci, często nie jest świadom utraty danych, bądź tego, że stał się ofiarą ataku. Edukacja oraz zwiększanie świadomości zarówno o zagrożeniach jak i rozwiązaniach zabezpieczających jest sposobem ochrony przed cyberprzestępcami [6].

W trakcie projektowania programu studiów II stopnia, a także podczas realizacji toku studiów zostaną wykorzystane wzorce międzynarodowe oraz własne doświadczenia. Podstawą sposobu sformułowania efektów uczenia się oraz doboru metod nauczania będzie ich projektowanie zgodnie z Europejskimi Ramami Kwalifikacji (ERK), zapewniającymi porównywanie poziomów kwalifikacji w różnych systemach edukacyjnych. Taki system pracy pomoże między innymi w możliwościach realizacji nowych wizji prowadzenia procesu dydaktycznego bazującego na teorii i praktyce w zakresie cyberbezpieczeństwa. Pomoże także w modularyzacji oraz możliwości racjonalnego wykorzystania Europejskiego Systemu Transferu Punktów (ECTS). Przy określaniu zakładanego poziomu wiedzy i umiejętności przypisanych poszczególnym kwalifikacjom, zostaną wykorzystane sposoby podziałów programu studiów na kategorie odnoszące się do postanowień europejskich występujących w deskryptorach dublińskich (Dublin Descriptors). Deskryptory te bazują na pięciu aspektach kształcenia, są to: wiedza i rozumienie; wykorzystanie w praktyce wiedzy i rozumienia; ocena i formułowanie sądów; umiejętności komunikacji; umiejętności uczenia się. Modularyzacja (stworzenie modułów, w skład których wejdą różne treści i formy kształcenia), pomoże między innymi w pogłębieniu efektywności studiów. To z kolei pozwoli na akumulację i transfer osiągnięć studentów, pogłębienie ich wiedzy (określonych w kategoriach efektów kształcenia), ale także większą indywidualizację programów studiów.

Zdobyta przez studentów wiedza i umiejętności pozwolą także w przyszłości na wykorzystanie ich kompetencji na użytek społeczno-gospodarczy oraz generowanie nowych zapotrzebowań w informacyjnej sferze nauki, edukacji i biznesu. Począwszy od potrzeb małych społeczności i przedsiębiorstw (diagnozowanie trendów rozwoju w danej dziedzinie czy obszarze gospodarki, nauki, kultury, edukacji), skończywszy na wyzwaniach organizacyjnych i zarządczych stojących przed poszczególnymi sektorami gospodarki, ale także określeniu kierunków rozwoju nauki

i technologii, a więc także potencjału sektora B+R+I. Jako ważną część procesu dydaktycznego przewiduje się kształtowanie umiejętności samodoskonalenia.

Planowane studia będą cyklicznie weryfikowane na podstawie przeprowadzanych okresowych analiz i ocen dotyczących rynkowego zapotrzebowania na specjalistów z zakresu cyberbezpieczeństwa. Istotne w tym względzie będą także, zgłaszane w trakcie całego procesu edukacyjnego, potrzeby samych słuchaczy. W trakcie prowadzenia poszczególnych zajęć wykorzystane zostaną dotychczasowe doświadczenia pracowników naukowych UW oraz współpracujących z nimi specjalistów spoza uniwersytetu (instytucji państwowych oraz przedsiębiorstw).

Główne założenia programu studiów II stopnia

Program studiów zakłada kształcenie absolwentów w oparciu o dwa poniższe filary:

- I filar – wiedza z zakresu kształtowania i stosowania polityki cyberbezpieczeństwa na poziomie firm, organizacji, instytucji państwowych i europejskich,
- II filar – wybrane obszary IT istotne w tworzeniu polityki cyberbezpieczeństwa,
- I filar programowy obejmuje następujące zagadnienia:
    - rozpoznawanie potrzeb oraz kształtowanie polityki i strategii cyberbezpieczeństwa na poziomie firm, organizacji, instytucji państwowych i europejskich,
    - sposoby tworzenia zasad kontrolnych i ochronnych dla zasobów technologicznych i informacyjnych przedsiębiorstw,
    - rozwiązania z zakresu cyberbezpieczeństwa funkcjonujące na szczeblu państwowym,
    - zasady polityki UE mające na celu zwiększenie cyberodporności, walkę z cyberprzestępczością, wzmocnienie cyberdyplomacji i cyberobrony,
    - prawo karne i ochrona cyberprzestrzeni,
    - zarządzanie ryzykiem i ocena bezpieczeństwa systemów IT,
    - ochrona prywatności w Internecie,
    - społeczna percepcja zagrożeń,

- dezinformacja oraz narzędzia i techniki manipulowania opinią publiczną.

Dyscypliny, do których projektowany kierunek studiów zostanie przyporządkowany przedstawia Wyk.3.

Wyk.3. Dyscypliny, do których projektowany kierunek studiów zostanie przyporządkowany

Źródło: opracowanie własne.

Powyższe zagadnienia obejmują tworzenie i kształtowanie polityki bezpieczeństwa w następujących obszarach:

- bezpieczeństwo systemów informatycznych,
- zabezpieczenia systemów operacyjnych,
- bezpieczeństwo komunikacji sieciowej i elektronicznej,
- technologie i zabezpieczenia internetowe,
- bezpieczeństwo pracy w chmurze,
- bezpieczeństwo IoT.

II filar to wybrane zagadnienia dotyczące obszarów IT, które wspierają operacje cyberbezpieczeństwa i których poznanie jest konieczne do realizacji zagadnień z I filaru, takie jak:

- sztuczna inteligencja,
- Big Data i analiza danych,

- technologie budowy serwisów i portali internetowych,
- podstawy programowania w cyberprzestrzeni,
- web 2.0 i media społecznościowe.

Zagadnienia programowe będą realizowane przy użyciu tradycyjnych metod dydaktycznych, takich jak: wykłady problemowe, konwersatoria, ćwiczenia. Zastosowane zostaną również metody oparte na dyskusji, angażujące studentów w wymianę poglądów, wymagające wewnętrznej aktywności oraz zachęcające do zajęcia własnego stanowiska (metoda sytuacyjna, giełda pomysłów, SWOT czy dyskusja panelowa, okrągłego stołu lub seminaryjna). Istotnym narzędziem będzie również zespołowe, pod kierunkiem prowadzącego zajęcia, analizowanie i rozwiązywanie konkretnych oraz rzeczywistych problemów, z którymi studenci mogą się spotkać w przyszłej działalności projektowej lub pracy zawodowej. Ważną częścią procesu dydaktycznego będzie również kształtowanie umiejętności samodoskonalenia.

Sylwetka absolwenta

Absolwenci będą wyróżniać się wiedzą w zakresie planowania i wdrażania prewencyjnych środków cyberbezpieczeństwa w celu ochrony przed cyberatakami. Będą znać zasady i regulacje kształtowania polityki i strategii cyberbezpieczeństwa na poziomie firmy, kraju i EU. Będą posiadać umiejętności identyfikowania obszarów występowania cyberzagrożeń i doboru metod przeciwdziałania ich występowaniu. Ważną kwestią będzie rozpoznawanie potrzeb organizacji w zakresie bezpieczeństwa IT, formułowanie, aktualizowanie i stosowanie rozwiązań organizacyjnych i technologicznych służących podnoszeniu poziomu zabezpieczeń. W aspekcie społecznym, ważnym elementem będzie propagowanie zrozumienia występowania i oceny zagrożeń cyberbezpieczeństwa oraz kształtowanie i uświadamianie potrzeby ograniczania ryzyka ich występowania.

Absolwenci studiów będą posiadać wiedzę w zakresie:

- funkcjonowania systemu cyberbezpieczeństwa na poziomie firmy i kraju oraz rozwiązań międzynarodowych w tym zakresie,
- oceny bezpieczeństwa systemów IT,
- podstaw prawnych ochrony informacji i systemów informatycznych,

- znaczenia sztucznej inteligencji w ograniczaniu ryzyka występowania cyberzagrożeń i ich zapobieganiu,
- bezpieczeństwa sieci komputerowych i komunikacji elektronicznej,
- wyzwań związanych z zapewnieniem bezpieczeństwa pracy systemów chmurowych,
- zagrożeń związanych z wdrażaniem nowych usług sieciowych, np. IoT.

Podczas trwania studiów studenci będą mogli kształtować następujące umiejętności w zakresie:

- analizowania sytuacji stwarzających ryzyko występowania cyberzagrożeń,
- diagnozowania zagrożeń w cyberprzestrzeni, na jakie są narażone są obecnie organizacje, państwa i ich obywatele,
- zarządzania ryzykiem i wdrażania strategii zapobiegawczych w celu zapewnienia bezpieczeństwa przedsiębiorstw,
- wykorzystywania narzędzi do przeciwdziałania zagrożeniom i destrukcyjnemu oddziaływaniu na informację i systemy informatyczne,
- ochrony systemów informatycznych zgodnie z aktualnymi aktami prawnymi oraz normami międzynarodowymi,
- współpracy w grupie i wspólnego rozwiązywania problemów,
- samodzielnego zdobywania wiedzy i kierowania rozwojem swoich umiejętności.

Kompetencje społeczne absolwentów będą dotyczyły następujących obszarów:

- propagowania i rozpowszechniania wiedzy dotyczącej możliwości występowania zagrożeń w cyberprzestrzeni,
- uświadamiania potrzeby ograniczania ryzyka zagrożeń,
- kształtowania odpowiedzialnych postaw dotyczących korzystania z cyberprzestrzeni,
- uznania znaczenia wiedzy w krytycznym odnoszeniu się do problemów bezpieczeństwa IT i w ich rozwiązywaniu w życiu społecznym i gospodarczym,
- konieczności uczenia się przez całe życie i ponoszenia swoich kwalifikacji wobec zmieniającego się otoczenia.

Nowy kierunek studiów w kontekście dotychczasowej oferty studiów UW

Analizę oferty edukacyjnej UW wykonano na podstawie danych zawartych w systemie RAD-on [5]. Uczelnia w swojej ofercie nie posiada kierunku, którego treści programowe byłyby poświęcone w całości cyberbezpieczeństwu. Wybrane zagadnienia z zakresu cyberbezpieczeństwa realizowane są jedynie w ramach pojedynczych przedmiotów wybranych kierunków takich jak np. Cyberbezpieczeństwo na Wydziale Prawa i Administracji lub Bezpieczeństwo zasobów cyfrowych, Bezpieczeństwo cybernetyczne i Warsztaty cyberbezpieczeństwa dla humanistów na WNPiSM. W ofercie uczelni znajduje się płatny kurs Cybersecurity realizowany w ramach współpracy Wydziału Matematyki Informatyki i Mechaniki z wiodącym izraelskim instytutem szkoleniowym HackerU [7]. W zależności od wyboru oferty programowej, kurs trwa maksymalnie 6 miesięcy i kończy się uzyskaniem certyfikatu ukończenia. Utworzenie i prowadzenie płatnego kursu z zakresu cyberbezpieczeństwa świadczy o dużym zapotrzebowaniu na specjalistów z tej dziedziny i chęci kształcenia się w tym zakresie. Projektowany kierunek studiów stanowiłby unikalną ofertę uczelni, doskonale uzupełniającą ofertę programową i stwarzającą możliwość zdobycia wiedzy z zakresu cyberbezpieczeństwa dla wszystkich studentów, również dla tych, którzy nie mają możliwości wyboru płatnych studiów.

Zgodnie z informacjami zawartymi w systemie RAD-on, studia II stopnia na kierunku cyberbezpieczeństwo oferują cztery uczelnie w Tab.1. Jedna z uczelni prowadzi zajęcia w języku angielskim, pozostałe w języku polskim. Profil kształcenia wszystkich oferowanych studiów to profil ogólnouniwersytecki. Cyberbezpieczeństwo występuje również jako specjalność na innych uczelniach krajowych.

W celu opracowania nowego programu studiów dokonano przeglądu dostępnych na stronach uczelni zakresów nauczania i profili kandydatów dla kierunków zawartych w Tab.1. Szczególną uwagę zwrócono na kierunki, które zostały przyporządkowane do dyscyplin: nauki o bezpieczeństwie, nauki po polityce i administracji oraz informatyki. Treści zawarte w ofertach programowych zostały skonfrontowane z proponowanym programem studiów nowego kierunku. Na uwagę zasługuje fakt, że znaczna część kierunków studiów z Tab.1. została uruchomiona w ostatnich czterech latach. Świadczy to o rosnącej potrzebie kształcenia specjalistów z zakresu cyberbezpieczeństwa.

Tab. 1. Wykaz uczelni wyższych w Polsce prowadzących studia II stopnia na kierunku związanym z cyberbezpieczeństwem

Lp	Nazwa kierunku studiów	Instytucja prowadząca	Poziom	Dyscypliny	Data uruchomienia
1	Bezpieczeństwo informacyjne i cyberbezpieczeństwo	Akademia Sztuki Wojennej	II stopnia	nauki o bezpieczeństwie (100%)	2021-10-01
2	Cyberbezpieczeństwo	Politechnika Wrocławska	II stopnia	informatyka techniczna i telekomunikacja (100%)	2020-11-05
3	Kryptologia i cyberbezpieczeństwo	Wojskowa Akademia Techniczna	II stopnia	informatyka techniczna i telekomunikacja (100%)	2018-03-01
4	IT cyber security	Uniwersytet Marii Curie-Skłodowskiej w Lublinie	II stopnia	nauki o polityce i administracji (55%), informatyka (45%)	2020-10-01

Źródło: https://radon.nauka.gov.pl, dostęp [2022-04-14].

Analizę uczelni zagranicznych oferujących studia II stopnia w zakresie cyberbezpieczeństwa przeprowadzono w oparciu o dane zawarte na międzynarodowej platformie Studyportals, skupiającej ponad 3750 instytucji edukacyjnych w 110 krajach. Platforma jest wspierana przez Komisję Europejską i inne krajowe instytucje szkolnictwa wyższego [4]. W zasobach platformy, w dyscyplinie Cyber Security, zamieszczone są 553 oferty studiów II stopnia [3] obejmujące zarówno studia ukierunkowane na zdobycie kompleksowej wiedzy z zakresu bezpieczeństwa, jak również studia skupiające się na konkretnych aspektach z zakresu cyberbezpieczeństwa lub łączące cyberbezpieczeństwo z innymi zagadnieniami np. zarządzanie cyberbezpieczeństwem, sieci komputerowe i cyberbezpieczeństwo.

Analiza dostępnych na platformie ofert programowych studiów II stopnia pozwoliła na sformułowanie głównych obszarów w zakresie cyberbezpieczeństwa:

- informacje w cyberprzestrzeni i bezpieczeństwo cyberprzestrzeni,
- bezpieczeństwo sieci komputerowych i pracy w chmurze,
- zarządzanie danymi i bazami danych, w tym zbiorami Big Data,
- komunikacja mobilna i programowanie,
- prawo i ochrona w cyberprzestrzeni,
- ochrona prywatności w Internecie,
- społeczeństwo informacyjne i socjologia cyberprzestrzeni,
- technologie i zabezpieczenia internetowe,
- podstawy Internetu rzeczy – IoT,
- inżynieria oprogramowania,
- bezpieczeństwo systemów informatycznych,
- bezpieczeństwo komunikacji elektronicznej,
- metody badawcze i planowanie projektów,
- administracja systemów informatycznych,
- uczenie maszynowe i sztuczna inteligencja,
- kryptografia.

Szczegółowej analizie poddano programy nauczania uczelni, które zgodnie z informacją zawartą na stronach portalu, zostały uznane w 2021 r. za jedne z najlepszych szkół cyberbezpieczeństwa na świecie:

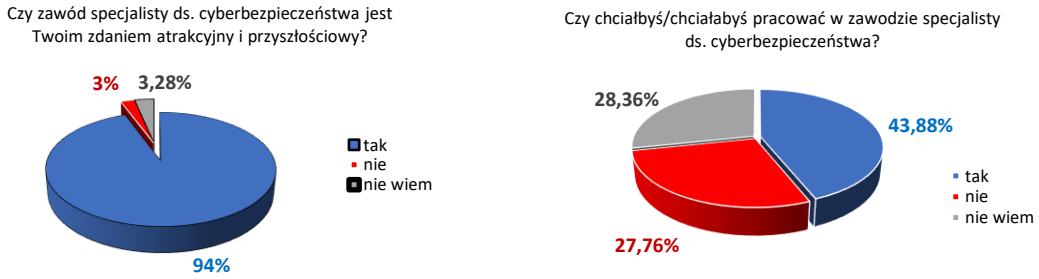
- Georgia Institute of Technology, USA,
- King's College Londyn, Wielka Brytania,
- Uniwersytet w Leiden, Holandia,
- ETH Zurych, Szwajcaria,
- Uniwersytet w Aalborgu, Dania,
- Uniwersytet Nowej Południowej Walii, Australia,
- Uniwersytet Jiao Tong w Szanghaju, Chiny [8].

Przeprowadzona analiza była punktem wyjścia do sformułowania nowego programu studiów II stopnia.

Potrzeby otoczenia społeczno-gospodarczego i studentów

W celu zbadania stopnia zainteresowania studentów nowo tworzonym kierunkiem studiów II stopnia, przeprowadzono ankietę wśród studentów studiów I stopnia [9]. Prośby o wypełnienie ankiety zostały wysłane do wszystkich studentów studiów I stopnia WNPiSM UW (za pośrednictwem Biura promocji i komunikacji) oraz zostały skierowane do wybranych studentów studiów I stopnia UW przy wykorzystaniu służbowych i osobistych kontaktów pracowników UW zaangażowanych w przygotowanie projektu studiów. W badaniu wzięło udział 335 osób, w tym 207 kobiet i 128 mężczyzn. Większość badanych osób (70%) to osoby mieszkające w miastach powyżej 500 tys. mieszkańców. Wszystkie badane osoby są studentami UW, a znaczna większość – 217 osób studiuje na WNPiSM, na którym planowane jest otwarcie nowego kierunku. Wyniki ankiety jednoznacznie wskazują, że studenci postrzegają zawód specjalisty ds. cyberbezpieczeństwa jako atrakcyjny i przyszłościowy, aż 43,88% badanych osób chciałoby pracować w tym zawodzie - Wyk. 4.

Wyk. 4. Zainteresowanie studentów zawodem specjalisty ds. cyberbezpieczeństwa



Źródło: opracowanie własne.

Średnie wynagrodzenie w branży cybersecurity to około 15 000 zł miesięcznie [1]. 96,12% respondentów uznało je za atrakcyjne - Wyk. 5. Kluczowe pytanie ankiety miało na celu zbadanie zainteresowania i ewentualnego wyboru nowo utworzonego kierunku przez studentów. Aż 47,16% respondentów zdecydowałoby się na wybór kierunku Cyberbezpieczeństwo, gdyby ten znalazłby się w ofercie studiów II stopnia - Wyk. 6. Na uwagę

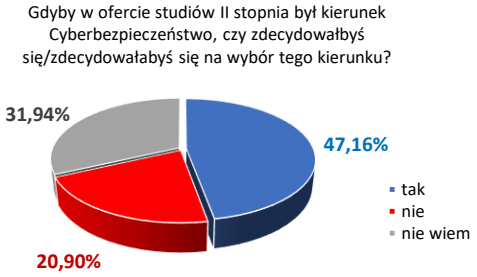
zasługuje również fakt, że 217 badanych osób to studenci studiów I stopnia WNPiSM, na którym nowy kierunek ma zostać utworzony.

Wyk. 5. Atrakcyjność wynagrodzenia w branży cybersecurity

Źródło: opracowanie własne.

Jedną z kluczowych kwestii przy projektowaniu nowego kierunku jest opracowanie programu studiów zgodnego z oczekiwaniami studentów. Tylko tak skonstruowany program może zapewnić stałe zainteresowanie ofertą programową i być atrakcyjny dla studentów.

Wyk. 6. Wybór nowego kierunku Cyberbezpieczeństwo na studiach II stopnia



Źródło: opracowanie własne.

Zdaniem studentów, najważniejsze jest bezpieczeństwo danych, a następnie bezpieczeństwo ludzi i bezpieczeństwo społeczne - Wyk.7. W opinii ankietowanych (236 osób), program studiów powinien uwzględniać zagadnienia związane z nowymi technologiami - sztuczną inteligencją, uczeniem maszynowym, zagadnienia Big Data itp. Ostatnim elementem badania było ustalenie preferowanych przez studentów kanałów promocji kierunku, mające na celu dotarcie

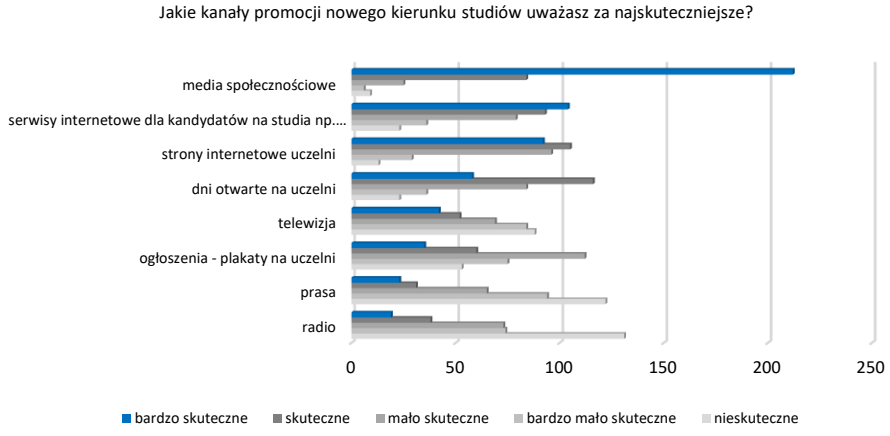
z informacją do jak największej liczby potencjalnych studentów nowego kierunku, w przypadku, gdy kierunek ten zostałby utworzony.

Wyk. 7. Preferowane przez studentów treści programu nauczania

Źródło: opracowanie własne.

Zdaniem studentów, najbardziej skutecznym kanałem promocji są media społecznościowe, następnie serwisy internetowe dla kandydatów na studia oraz strony internetowe uczelni - Wyk.8. Ważnym elementem promocji kierunku są również dni otwarte na uczelni. Pozostałe kanały promocji, w tym zwłaszcza media tradycyjne, zostały uznane przez studentów jako mało skuteczne.

Wyk. 8. Preferowane przez studentów kanały promocji nowego kierunku



Źródło: opracowanie własne.

W ramach przeprowadzenia rozeznania potrzeb/oczekiwań otoczenia społeczno-gospodarczego pozyskano rekomendacje następujących instytucji:

- Komendy Głównej Policji,
- Departamentu Spraw Obronnych, Zarządzania Kryzysowego i Bezpieczeństwa Ministerstwa Klimatu i Środowiska,
- NASK'u - Państwowego Instytutu Badawczego.

Bibliografia:

[1] CyberDefence24, https://cyberdefence24.pl/bezpieczenstwo-informacyjne/branza-cyberbezpieczenstwa-pilnie-poszukuje-pracownikow-kusi-wysokimi-zarobkami, dostęp [2022-07-28].

[2] GTrends, https://trends.google.pl, dostęp [2022-09-28]

[3] Platforma Studyportals dla studiów II stopnia - https://www.mastersportal.com, dostęp [2022-09-28].

[4] Platforma Studyportals, https://studyportals.com dostęp [2022-09-28].

[5] Radon, raporty, analizy, dane, https://radon.nauka.gov.pl, dostęp [2022-09-28].

[6] Raport Cyberbezpieczeństwo: Trendy 2021, https://xopero.com/pl/dokumenty/raport-cyberbezpieczenstwo-trendy-2021/, dostęp [2022-09-28].

[7] Strona kursu Cybersecurity, https://cybersecurity.mimuw.edu.pl, dostęp [2022-09-28]

[8] Why You Should Study a Cyber Security Degree in 2022, https://www.mastersportal.com/articles/2722/why-you-should-study-a-cyber-security-degree-in-2022.html, dostęp [2022-04-14].

[9] Wyniki ankiety z dnia 22 kwietnia 2022 r.

Nataliya Poplavska

Infomedia literacy of the audience as a key to countering disinformation: from work experience

Introduction

In the times of the war currently ravaging Ukraine, the wide media audience is facing new and new challenges, especially those regarding its ability to understand information properly. The CEDOS analytical center, founded by the International Renaissance Foundation with the financial support of the Embassy of the Kingdom of Sweden in Ukraine, has cooperated with the Prague Civil Society Centre and the Heinrich Böll Foundation to implement a two-stage research analyzing the influence of the large-scale war on the Ukrainian society, and they have discovered what thoughts, feelings and actions characterized the Ukrainian people within the first two weeks after February 24th.In order to understand the dynamics of changes in emotional state, decision-making and adaptation of everyday life to wartime conditions, the second round of the research was performed in May 2022.The research made use of a questionnaire, based on Google Forms, to collect the necessary data. The subsequent report outlined certain conclusions: it mentions the main tendency that people felt detachment, being plucked out of the current reality, “as if the time had stopped,” and empathy with the compatriots’ grief. At the same time, the second round of the research showed that respondents’ emotions were becoming more self-targeted. “Although the tendency to plan the future for a day or two was still widespread among the respondents, it was no longer dominant, unlike it was in the first round. The planning horizon changed with the understanding that this

war was going to continue. The respondents mostly mentioned having both short-term and long-term plans. Long-term plans were most frequently related to events postponed till the time after the victory: homecoming, restoring the usual routine, meeting family members" [6].Such willingness to perceive a potentially negative situation certainly increases resistance to influences and protects people from excessive traumatization. The effects on the audience are evidently boosted by the media as well which can also inflict psychological damage on the people. According to Lyubov Naydionova, "a media trauma is a kind of stress caused by media experience, empathy to traumatizing events shown in the media (primarily in the form of photo and video). Media traumas were mentioned for the first time when after being showed acts of terrorism on television, doctors started increasingly seeing people with symptoms similar to PTSD (Post-Traumatic Stress Disorder), although the patients themselves were not directly exposed to the traumatizing events" [8].The issue of recognizing media traumatization as the grounds for a diagnosis is still under discussion, but currently PTSD caused by media cannot be recognized since a person can stop consumption of such media products at any time and therefore remove the traumatizing factor. Meanwhile, the media are recognized as being even more stress-provoking than other live events as indicated by results of Ukrainian research [13].The media's high potential for manipulation gives us sufficient grounds to view them as factors capable of exerting explicit and implicit psychological pressure on the people being influenced by them. In wartime conditions, certain factors can be activated in the information realm that can destroy some of the realm's structures or initiate their self-destruction processes. It may happen not only due to the enemy's propagandist narratives; there is also counterpropaganda, certain restrictions in the freedom of speech due to the war, when information that can be used by the enemy to advantage is not reported in real time. That is why the audience is under the threat of disinformation influences.

To counter destructive influences of the media, the audience needs self-regulation for their media consumption, and such situation demonstrates that the problem of media education does exist, and its relevance is constantly increasing.The main thesis is also vital: a person unprepared to perceive information in various forms cannot be expected to understand and analyze it properly, is unable to resist manipulative influences of the media as well as express their thoughts and feelings independently.Critical perception of media information and the skill of analyzing media products gains particular importance, especially in conditions of the full-

scale war that our country is involved in, and this makes the problem *relevant*. The purpose of the research is to separate out the main tools that can facilitate formation of critical perception skills in mass audience in order to counter disinformation influences, to discover basics of media text decoding, and to bring out the importance of media education in this context.

Research results

Scientists are unanimous in their recognition of manipulative nature of the processes taking place in the media space, and they point out its plasticity and relation to the social space. In this regard, S. Hrytsay emphasizes: “Creating new conditions for vital activities of the society, the new reality does not only influence its organizational and communicational peculiarities, but also thenature of social relations which are becoming more rapid, media-saturated, varied and intensive, thus giving rise to new socio-psychological and information-psychological phenomena: information phobias, information pressures and tensions, aggressiveness and information crime”[4].

The means of mass communication that for, the information space have already become an inseparable component of everyone’s professional and personal life.Information knows no borders, geographic distances or any other limits, and it has the cumulative property of permanent accumulation and self-restoration.It has become so easily accessible, and its volumes so large, that it is not surprising that people sometimes feel lost in it.That is why the media influence on a personality is growing, and they become influential social regulators, gaining new dimensions and functionality. The following tendencies can be observed: increasing scope of audience, promptnessin delivering information to the audience, increasing speed and sophistication of information flows, which can be explained by wider access to information, extended range of information sources, oversaturation by information, growing information noise and complication of control over mass media. Thanks to modern technologies and freedom of communication, anyone can use media channels and media platforms, and generate media texts which differ fundamentally from professional journalism products.

Moreover, one may also add the tendency for recipient's activity in establishing feedback to media, that is, a consumer's active participation in creating the products they are interested in. A manifestation of this phenomenon is the emergence of powerful amateur or user content in modern media environment, and such content also competes with the products created by journalists. The situation is even more complicated because certain information can be distorted, manipulative, or merely false. Large volumes of information can turn our choice of what and whom to trust into a truly exhaustive task.In such situation, mass audience is not always able to distinguish obviously biased material or fake information from professional journalistic text.

All these factors result in media's ability to manipulate people's opinion, and exert targeted influence on the audience in general, or on specific groups within such audience.In today's complicated information world, they become an important tool to form political, economic and social agenda.There are quite many of those who would like to format this agenda to suite their private needs and interests.Such influencers may include politicians, officials, media owners and even journalists themselves.However, the knowledge of telling a quality text from a superficial material can help us resist the attempts to manipulate the audience.

In Ukrainian media environment, changes are also determined by the social factor, namely the economic crisis and the war which also influences the development of media.In its situation of full-scale war with Russia (and therefore, permanent information attacks), Ukraine is especially vulnerable to propaganda manipulations, both internal and those generated by the aggressor.That is why proper understanding of media information and the skill of analyzing and critically perceiving mass media products becomes the knowledge of great importance in wartime conditions.

It is quite disappointing to admit that such program is not in full force and effect yet, despite active efforts by the Ukrainian academic community and civic organizations to work in their own directions in the field of media education.After the Concept of Media Education Implementation in Ukrainewas approved in 2010, the time has come to develop a complex government program that would encompass all age categories – children, youth and adults – which would be regarded as an element of the government education policy.Since today such education policy is needed not only for schoolchildren but for the entire population, and a person's media and information literacy should be developed throughout their life regardless

of their age, status, occupation, and place of residence.The problem of systemic media education for Ukrainian audience has become especially urgent in wartime conditions, since it determines how well everyone is protected against harmful disinformation influences attempted by the enemy.

Unfortunately, there are no unified definitions for the terms “media education” and “media literacy” yet. The International Encyclopedia of the Social & Behavioral Sciences suggests this definition of media education: “The study of media which is different from learning with the use of media, is simultaneously related to understanding how media texts are created and disseminated, and to the development of analytical abilities to interpret and evaluate their contents”[12].A Ukrainian scientist G. Onkovych believes that it is “an active process of a personality’s development and self-development based on mass media support and use of their materials” [9].

O. Konovets emphasizes that it is one of the directions or components of the so-called citizen science (“science for everyone,” citizenship lessons), which is an important element of civil society and purportsto form an active, smart, independent, critically thinking consumer of mass information who is capable of public communication, and to build certain psychological protection against manipulation or exploitation on the part of mass media, while developing a person’s information culture [5].

Having analyzed all existing definitions of media education, we can generalize them and define that media education should give people the knowledge of how to analyze, critically consider and create media texts, identify their sources while taking into account political, social, commercial, cultural interests and contexts; interpret media texts and values carried by the media. It has to be noted, that media education is directed at achieving media literacy; it does not suggest teaching with the help of media, but it facilitates development and formation of one’s personality while using media materials.

As to the terms of “media literacy,” the Grunwald Declaration of 1982 defines it as “a person’s ability to understand the role and functions of mass media; to analyze and evaluate media contents critically; to use mass media for democratic participation, intercultural dialogue and education; to produce their own media contents (create media products of their own); master information and communications technologies, and other media skills” [3].The European Commission suggests such a definition for this term: “The ability to access the media, to

understand and to critically evaluate different aspects of the media and media contents and to create communications in a variety of contexts" [7].

"The ability to use individual media unaided, to understand, and bring critical assessment to bear on, the various aspects of media [...], to separate out information from the new media's flood [...] and to categorize that information" is the definition given by the European Parliament in its "Resolution on media literacy in a digital world" [2].Hence media literacy is the skill of synthesizing and analyzing the space-time reality, and the skill of reading media texts. It results from the process of media education and consists in knowing the peculiarities of media structure, functions and tools, in being able to analyze and evaluate phenomena, and participate actively in social, cultural and political life.

Considering that "individuals acquire professional features either in the period of basic training or in the course of retraining, or directly on workplaces, therefore the list of professional competences changes depending on the demand of the society and in accordance with peculiarities of a certain country or region and job market trends" [11], and critical thinking is evaluative, reflexive, does not accept dogmas and develops by way of imposing new information on a person's life experience, it is important that recipients should be able to acquire communicative competence which helps them to overcome communicative obstacles, and use it in their lives. That is why the goal of forming basic media literacy and practical skills of effective and safe interaction with media is the development of reflexive, logical, critical and creative thinking, and formation of the ability to understand and analyze media texts.

Media text is a multi-genre media product in a printed, video or audio format directed at formation of a person's conceptual worldview and its social regulation.In the present-day information society, it is used as a unique means of interpreting and representing reality, i.e., to form media reality.That is why it shapes the recipient's worldview and therefore enables mass media to influence important social processes.Based on the above, all didactic strategies in media education activity should be directed towards critical analysis of media texts, studying social and political contexts of their creation, forming practical skills and abilities to protect oneself from harmful media influences, and using media technologies to generate one's own texts. In-depth analysis of media text and its effects on how the audience thinks should become one of the important elements to increase effectiveness of media literacy.

An important aspect of media education activity isthe “Learn to Discern: info-media literacy” project implemented in Ukraine by the International Research & Exchanges Board (IREX) with the support from the U.S. and UK embassies in Ukraine, and in partnership with the Ministry of Education and Science of Ukraine, and the Academy of Ukrainian Press (AUP).The project started in 2017.Its main goal is to effectively introduce the youth to such important competence as critical thinking, and help them realize the value of quality information. The teaching staff of the Department of Journalism at Ternopil Volodymyr Hnatiuk National Pedagogical University joined the project in 2017.Project participants underwent training in various educational activities. They included trainings for integration of infomedia literacy into teaching materials, and specialized training sessions for teaching infomedia literacy with the use of interactive methods and online tools. It allowed us to borrow effective approaches to analyzing information, and methodological tools for teaching media education in order to improve audience resilience to propaganda and disinformation. Also, it helped us to spread positive experience of critical perception of information among future educators, students and teachers, united territorial communities, and members of government agencies.

Based on the methods for practical implementation of infomedia literacy among potential media audience developed by IREX and practically tested, we created the curriculum for the subject “Infomedia Literacy” which has been certified successfully by the International Research & Exchanges Board, and has been taught for two years as an elective course for students seeking Bachelor’s and Master’s degrees in both pedagogical and non-pedagogical specialties offered by our university.Also, this subject is introduced as a separate content module to students of the Postgraduate Education Center at our university, and the State Institution of Postgraduate Education –Ternopil Regional Center for Retraining and Advanced Training of Local Government Employees. We also worked with spontaneous audiences – for instance, with population in certain settlements. In terms of quantity, every group typically consisted of 30 to 40 persons. Since 2020, such meetings have been conducted mostly in remote format, which also enabled its participants to improve their infomedia literacy.

The goal of the subject “Infomedia Literacy” is to form a basis of theoretical knowledge for elementary infomedia literacy and practical skills in handling information through the use of modern information and communications technologies, to develop critical thinking, and to resist manipulations and propaganda on the part of the media.It “emphasizes the need for

permanent (long-term) attention to challenges of modern information society, provides guidance on relevant knowledge, skills and abilities, as well as attitudes suitable for permanent use and further improvement"[1].The key tasks of the course are to develop recipients' ability to think critically and to use this skill in everyday life, to enable them to find their bearings in the information environment, analyze the media and their content.The course participants should realize the role played by the media in shaping a person's worldview, present-day tendencies in development of media environment, use of information, and peculiar features of such environment, methods for protection from possible negative influences, understanding and using of methods for search, selection, and systematization of information and use of audiovisual and printed content; the ability to demonstrate a proper level of critical thinking and understand the extent of information influence on one's personality; the skillto verify a media product using modern information technologies; the ability to interpret relevant phenomena (fake news, information attacks, manipulations etc.) of the information environment for a wider audience, and improve the level of their media literacy; the ability to demonstrate the skill of acting professionally in various communicative situations and form one's culture of communication in an information society; they should also understand the technologies of creating a media product, and overcome negative tendencies generated by mass media.

Their attention is focused on the following thematic blocks: 1. Media environment of Ukraine: present-day challenges (it envisages learning about peculiarities of media text and its types, such terms as *media education* and *infomedia literacy*, the most widespread techniques of deviating from professional standards of journalism in mass media, media manipulations, peculiar features of modern mass media); 2. Ways of forming critical thinking (forms the development of critical thinking, the skill of reviewing a media text critically, identifying manipulative technologies); 3. How to spot a fake (it suggests clarifying the notion of fake, understanding the typology of fake messages, and distinctive features of fake news); 4. Verification of online content (it suggests teaching recipients to verify a website, photos, videos); 5. Social media: peculiarities of communication (it provides guidance on trolls and bots in social media, social media users and their activity, classification of social media and their functions); 6. Hate speech in mass media (it explores criteria for identifying hate speech, and provides recommendations as to using and selecting correct vocabulary).

Although course participants are introduced to a wide range of theoretical materials, the block of practical activity is advisable and effective. For example, while exploring the topic "Media environment of Ukraine: Present-day challenges" and considering that televisions remains to be a powerful media resource and one of the main means of influencing the audience, it is advisable that course participants acquire practical skills in analysis and critical evaluation of television tests (as exemplified by news items), learn to distinguish facts from opinions in media products, be able to identify manipulations aimed at media product consumers, determine sources of media texts and their context (their authors, owners, and the parties that benefit from news presented in a certain way). This block consists of at least three stages and suggests work in groups.

Stage I. *Analytical viewing of news programs on one of the TV channels, in groups.* The process of performing this task can be transformed into an experimental game.Suggest that one group should analyze news texts for observance of the standard for completeness of information, the second group – for observance of the standard for separation of facts from comments, and the third group – for observance of the standard for accessible presentation of the information. After viewing the news and discussing it in groups, it is worth eliciting critical comments to audiovisual products regarding their observance of professional standards.

Stage II. In order to teach the course participants to identify emotional and manipulation influences in media texts, they are suggested to watch an item from the final news program of the day on one of the TV channels. Within the framework of this task, course participants are invited to answer such questions: *Are facts replaced with opinions? Do the facts correspond to the problem being considered? Is emotionally colored and evaluative lexicon used? For what purpose? Does the material appeal more to people's intellect or emotions? What values does it appeal to?* It is also suggested to analyze one's emotions that arise while watching the news.

Stage III. In order to help course participants to form the skill of identifying how mass media influence people's consciousness, what manipulative technologies are used, and how the peculiarities of their editorial policy can be explained, it is necessary to prepare (for at least three groups) a short selection of news items from top rated TV channels, where signs of media manipulation are present. We suggest that every group should analyze a selection of news pieces by answering the following questions: *Aren't the media paying too much attention to this topic? Why are the media discussing it right now? Why weren't they discussing it when the*

actual event happened? Is this topic truly so important to receive so much attention? Who benefits from spreading this news right now? Who is this news directed against? Why should the society know about it? Is it important for me? What are they trying to hide with this topic? Is there any news that can be more important than that? What important topics are ignored in the news, while news pieces like this are put in the foreground?

After this block, every course participant is given a sheet of paper with printed after-class reflection matrix, and they are suggested to complete the unfinished statements:*I learned to…, I learned that…; I found out that...; I found a confirmation that…;I like that…; I was disappointed by the fact that…;The most important thing in class for me was…;Or write your own opinion.*

Therefore, mastering key knowledge and skills of critical commenting as exemplified by news texts enables us to train a person to consciously perceive modern media content, prepares people for the need to think independently and form their own opinion of the present-day events, teaches them to use analytical tools that can be used in any life situation.

In order to solidify the knowledge gained during the course, it is important that the participants engage in self-guided work, that is, use the acquired knowledge in everyday life, which envisages analysis and critical evaluation of any media texts, searching for and commenting on materials that contain manipulative technologies (we recommend using such information platforms as politobzor.net, nahnews.org), recognizing signs of fake news, distinguishing fakes in social media posts, Internet publications, checking illustrations and videos of specific websites or social media pages with the use of verification tools suggested at the lessons, working on negatively colored vocabulary and searching for appropriate counterparts, exploring modern Ukrainian media for presence of hate speech, studying specialized social networks, evaluating their content and circle of members, reviewing technologies applied for information influence through network communities (screenshots of corresponding content and posted comments).

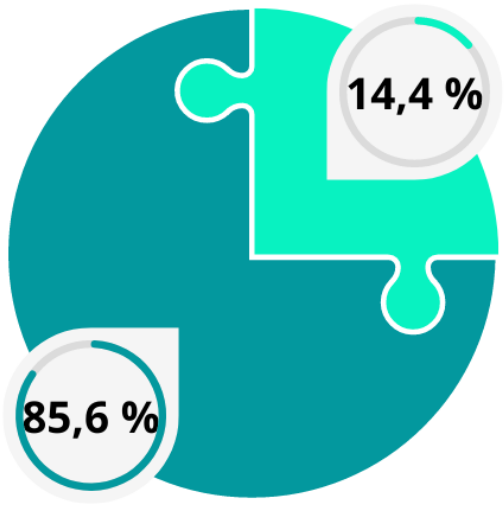
As to the special nature of the course, its teaching takes into account a personality's critical autonomy and critical analysis of media text. In the course of work on shaping the audience's system of values, we prefer group and interactive forms of training – exercises, discussions, brainstorms, case studies, analyzing problematic situations, and reflections. Their essence consists in organizing the educational process based on principles of interaction, which encourages the course participants to think critically, to find solutions to complicated problems

while possessing relevant information, to consider alternative opinions, take balanced decisions, and participate in discussions.

As shown by our practice, these are the methods that facilitate interest for infomedia literacy, create comfortable conditions for everyone to perceive the material, and allow the transition from monologue-based system to dialogue, where course participants do not only exchange ideas and opinions concerning reality, but also train to engage in discussions with the moderator, whose task in this situation is to organize search for information and discussion. Such forms of learning also suggest that course participants can learn from each other as well.

An essential and final aspect of our cooperation with the audience to implement the goal of the course is an anonymous survey with Google Forms. The block of questions includes as follows: 1. The level of teaching and course accessibility (rated from 2 to 5); 2. Informativity and practical value of its materials for everyday life (yes, no, I don't know); 3. Have you gained certain skills in analysis of media texts? (yes, no). These problems are believed to be the most relevant for improving the course and aiming it at the mass audience, which evidently faces such terms as *media education, media, infomedia literacy, media content* for the first time. For analysis, we have used the materials of our Google Form surveys conducted in three different groups that became participants of the Infomedia Literacy course.

Fig. 1 Course level and availability



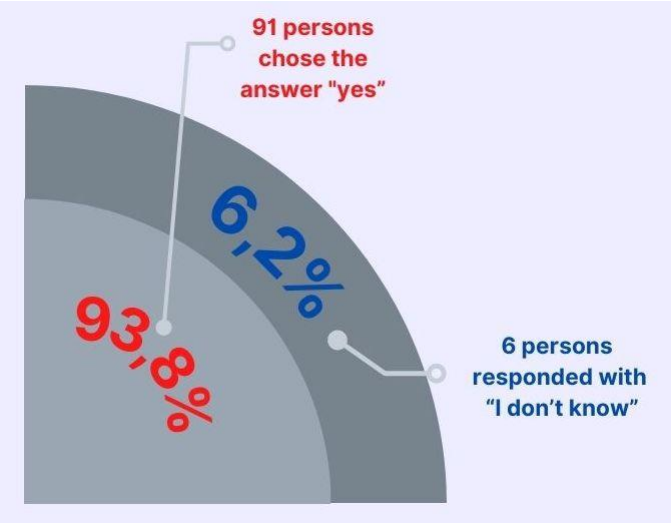
Source: own study.

They include students of the State Institution of Postgraduate Education –Ternopil Regional Center for Retraining and Advanced Training of Local Government Employees (34 persons), the Postgraduate Education Center of Ternopil University (37 persons), and officers from various departments of united territorial communities of Ternopil Region (26 persons). The total number of participants who took part in the survey is 97 persons.

All survey participants rated the level of teaching and course accessibility with 4 and 5 points. 83 persons rated them with 5 points, which equals 85.6%, and 14 persons rated them with 4 points (14.4%). Results of the survey are presented in the diagram (Fig. 1 Course level and availability). With all that, in free conversations course participants shared their positive impressions, pointed out to the high level of quality and informativeness of the presentations, the approach, accessibility and methodology of presenting the materials.

As regards informativeness and practical value of materials for everyday life, 91 persons chose the answer “yes” (93.8%) and only 6 persons (6.2%) responded with “I don’t know” (Fig.2 Informativeness and practical value of materials for everyday life). Responses to the question “Have you gained certain skills in analysis of media texts?” are distributed as follows: 87 persons gave positive answers, which equals 89.6%. Meanwhile, 10 (10.1%) of the course participants haven’t acquired any skills in analyzing media texts (Fig.3).

Fig. 2 Informativeness and practical value of materials for everyday life



Source: own study.

The results can be explained by a multitude of factors: varied level of participants' training, their willingness to join the course, motivation, peculiarities of their worldview etc.

Fig. 3 Have you gained certain skills in analysis of media texts?

Source: own study.

Hence, we can see that the majority of course participants have improved their knowledge and skills in regard to understanding information and its effects on one's personality, verifying media products, being able to recognize fake news, information attacks, manipulations etc., being able to demonstrate professional skills in various communicative situations, understanding the technology behind creation of media products, and overcoming negative tendencies created by mass media.

A relevant project within the context of activities performed by the Department of Journalism at TernopilVolodymyrHnatiuk National Pedagogical University to popularize infomedia literacy among potential mass media audience is V ETERI (On the Air)online media school, which has been arranged for two consecutive years(2021, 2022). Course participants are mostly young people.Its goal is to provide course participants with an opportunity to try their hand at

journalism and master the basics of media activity.The lessons are given in the form of webinars and give participants a chance to learn more about backstage secrets of journalism, to master the art of creating media text, to learn the tools used to create photo and video content, and this activity also creates a platform for interactive communication between course participants and coaches, media professionals and experts. The course duration is 12 weeks. For example, course participants registered for the lessons in 2022 joined regular Zoom meetings every Wednesday at 5 p.m. from 9.02.2022 to 27.04.2022. It is important that the main webinar speakers were students of Journalismwho had undergone relevant training in the lessons given by the teaching staff of the Department of Journalism. (Fig. 4 Advertisement of V ETERI media school in 2022 [10]).

Fig. 4 Advertisement of V ETERI media school in 2022

Source: own study.

The program consists of two modules. Module 1. It includes the following topics: Topic 1.Secrets of successful communication: the skill of speaking; Topic 2. How to handle public speaking;Topic 3. How to avoid taking the bait of fake news;Topic 4. Secrets of storytelling; Topic 5. The art of self-advertising. Module 2. Topic 1. How to write great texts;Topic 2. The art of photography; How to boost your audience with photos; Topic 3. Deadlines: how to do it all

in time; Topic 4. How to become a successful TV journalist; Topic 5. Journalist's rights and responsibilities of a journalist: how to stay out of trouble; Topic 6. How to create your own media business.

The main tasks of such lessons include the participants' mastering fundamental knowledge of information security and gaining the ability to protect themselves against information aggression.Such forms of work in implementation of vitally important competences for mixed-age audience in the circumstances of full-scale warthat Russian wages against Ukraine are very relevant, but the materials used by moderators for illustrative purposes need to be constantly updated and adapted to actual conditions and challenges.

Conclusions

Media education activities for mass audience are the key to forming people's critical thinking, ability to evaluate media texts, identify manipulative media content, the skills to analyze preconditions and causes of events, discern facts from judgments, identify propaganda, fakes and manipulations, analyze causes of events and their possible consequences, and avoid negative influences.That is why media education should not only be a topic for scientific and theoretical discussion, but also become the subject of system development at the national level for mixed-age audience, as an urgent need in the present-day wartime conditions.Its importance is obviously incontestable.

References:

[1] Daschenko, Nataliia. „Infomedia literacy in the university educational process: results of implementation". The Scientific Issues of Ternopil Volodymyr Hnatiuk National Pedagogical University. Series: Pedagogy, 1(1), 96–106. https://doi.org/10.25128/2415-3605.22.1.12 [in Ukrainian].

[2] European Parliamentre solution of 16 December 2008 on media literacy in a digital world (2008/2129 (INI)) European Parliament : portal. https://www.europarl.europa.eu/doceo/document/TA-8-2018-0485_EN.html.

[3] Hriunvaldskaia deklaratsyia YuNESKO po medya obrazovanyiu. (b. d.). Ynformatsyonnaia hramotnost i medyaobrazovanye dlia vsekh. http://www.mediagram.ru/documents/documents_23.html.

[4] Hrytsai, S. „Defining the concept of ‘media space’ from the perspective of multidisciplinary approach”. Visnyk Kharkivskoi derzhavnoi akademii kultury, 36(2012), 235–243 [in Ukrainian].

[5] Konovets, A. Media education and formation of innovative environment in educational institutions. Topical issues of mass communication, 8, (2007). 31-33 [in Ukrainian].

[6] Kulchytskyi, S. (2022, 17 August). “Three months of full-scale war in Ukraine: thoughts, feelings, actions”. International Renaissance Foundation. https://cedos.org.ua/en/researches/three-months-of-full-scale-war-in-ukraine-thoughts-feelings-actions/.

[7] MediaLiteracyProfileEurope [Electronicresourse] EuropeanCommission. http://ec.europa.eu/culture/library/studies/literacy-trends-profiles_en.pdf.

[8] Mediatravmaviiny: zviazok psykholohichnoho blahopoluchchia i mediakompetentnosti (L.A.Naidonova) [Video]. (2021, 28 September). Youtube. https://www.youtube.com/watch?v=DasjYYYCpno.

[9] Onkovych, H. “Professionally-oriented media education in higher school”. Higher education of Ukraine. Theoretical and scientific-methodical journal, 2(2014), 80–87 [in Ukrainian].

[10] Onlain-mediashkola „V ETERI”. (b. d.). Sait kafedry zhurnalistyky of Ternopil Volodymyr Hnatiuk National Pedagogical University. http://kafgyrn.tnpu.edu.ua/uk/онлайн-медіашкола-в-етері/.

[11] Poplavska N. M., Dashchenko N. L., Medynska O. Ya. „Key educational competencies and competencies of information literacy”. „Infomediina hramotnist – nevidiemna skladova navchalnoho protsesu zakladuv yshchoiosvity: zbirnyk statei”. red. V. F. Ivanov ta in. Kyiv: Akademia ukrainskoi presy, IREX, Tsentrvilnoipresy, 2021. ss. 83–108.ISBN 978-617-7370-26-9[in Ukrainian].

[12] Smelser,Neil J,Baltes, Paul B. „International Encyclopedia of the Social & Behavioral Sciences”Elsevier. ScienceDirect.https://www.sciencedirect.com/science/referenceworks/9780080430768.

[13] Study of the psychological state of the population in the conditions of a full-scale war. (2022, 13September). Institute of Social and Political Psychology. National Academy of Educational Sciences of Ukraine. https://ispp.org.ua/2022/09/13/doslidzhennya-psixologichnogo-stanu-naselennya-v-umovax-povnomasshtabnoii-vijni/.

Tomasz Gruszka

Wykorzystanie technologii Blockchain w celu ograniczenia dezinformacji i redefinicji modelu biznesowego przedsiębiorstw medialnych na przykładzie projektu Pix.T

Wstęp

Rozwój Internetu przyniósł duże zmiany w sposobie w jaki korzystamy z mediów oraz warunków prowadzenia biznesu przez przedsiębiorstwa medialne. W powszechnym odbiorze zmiany te są postrzegane jako pozytywne. Internet pozwolił na szybki, szeroki i tani dostęp do informacji. Stał się niestety również katalizatorem negatywnych zjawisk takich jak dezinformacja i erozja modelu biznesowego tradycyjnych przedsiębiorstw medialnych w wyniku spadku wartości rynkowej informacji oraz nowej konkurencji. Nowa konkurencja ze strony firm cyfrowych takich jak wyszukiwarki, media społecznościowe czy portale często związana jest z naruszaniem praw autorskich będących fundamentem modelu biznesowego tradycyjnych firm medialnych.

Określenie "post-prawda" zostało uznane za słowo roku 2016 przez redakcję Oxford Dictionaries. Powodem był gwałtowny wzrost jego popularności w związku z referendum ws. Brexitu oraz sukcesem Donalda Trumpa w czasie wyborów prezydenckich w USA. Określenie "post-prawda" opisuje świat, w którym osobiste przekonania i emocje są ważniejsze

od faktów. Czas ten określany jest również jako *age of customer*, gdzie to klienci mają przewagę nad przedsiębiorcą, państwem i renomowanymi ekspertami [1].

Wraz z pojawieniem się stron internetowych takich jak Twitter, Facebook, YouTube i TikTok rosnąca liczbą użytkowników pozyskuje informacje z serwisów społecznościowych. Dwie trzecie Amerykanów uzyskują wiadomości za pośrednictwem internetowych serwisów społecznościowych [2].

Natura social mediów czyni z nich miecz obosieczny. Z jednej strony są one dogodną platformą do rozpowszechniania użytecznych informacji o ważnych problemach takich jak Covid -19 z drugiej zaś stanowią zagrożenie ze względu na brak standardów i wymagań profesjonalnych stawianych dziennikarzom w tradycyjnych mediach. Użytkownicy social mediów sami decydują o wyborze źródeł informacji zamykając się w bańkach informacyjnych. Bańki te powodują, że w rezultacie naszych wyborów źródeł i charakteru aktywności w social mediach docierają do nas tylko takie informacje, które odpowiednie algorytmy określiły jako zbieżne z naszym światopoglądem oraz zainteresowaniami. Rozprzestrzenianie się zwodniczych lub wprowadzających w błąd informacji stanowi zagrożenie społeczne, gospodarcze i polityczne. Mylna informacja rozprzestrzenia się szybko i tanio. To rozprzestrzenianie się jest szczególnie skuteczne, jeśli treść rezonuje z uprzedzeniami użytkownika i uprzedzeniami społeczności. Nieprawdziwe informacje rozpowszechniają się szybciej niż informacje prawdziwe [3].

Za dezinformacją i wrogimi działania w przestrzeni informacyjnej mogą stać państwa oraz prywatne grupy interesu. Z analizy CitizenLab wynika, że z 218 przebadanych prób nielegalnego uzyskania dostępu do prywatnych komputerów (w ciągu jednej akcji) 21 proc. ataków dotyczyło biznesu, a 24 proc. – przedstawicieli rządowych [4].

Fałszywe informacje sprzyjają kształtowaniu postaw konfrontacyjnych i radykalnych, a także manipulacji nastrojami społecznymi. Czynniki te wpływ na stabilność państwa i potencjał mobilizacji społeczności, na przykład na protest przeciwko projektom gospodarczym [4].

Fałszywe informacje mogą mieć wpływ na wycenę aktywów na giełdach papierów wartościowych czyniąc znaczące szkody akcjonariuszom. Przykładem może być zdarzenie listopada 2022. Amerykańska firma farmaceutyczna straciła miliardy dolarów wyceny po opublikowaniu fałszywej informacji przez podszywające się pod spółkę konto. Użytkownik podający się za Eli Lilly and Company napisał, że „insulina od teraz będzie darmowa”.

Konsekwencje dla firmy farmaceutycznej były bardzo dotkliwe - cena wartości pojedynczej akcji spółki Eli Lilly na amerykańskiej giełdzie spadła w ciągu doby o około 20 dolarów, z poziomu 366 do 345 dolarów. Według szacunków, firma mogła stracić nawet 20 mld dolarów kapitalizacji rynkowej [5].

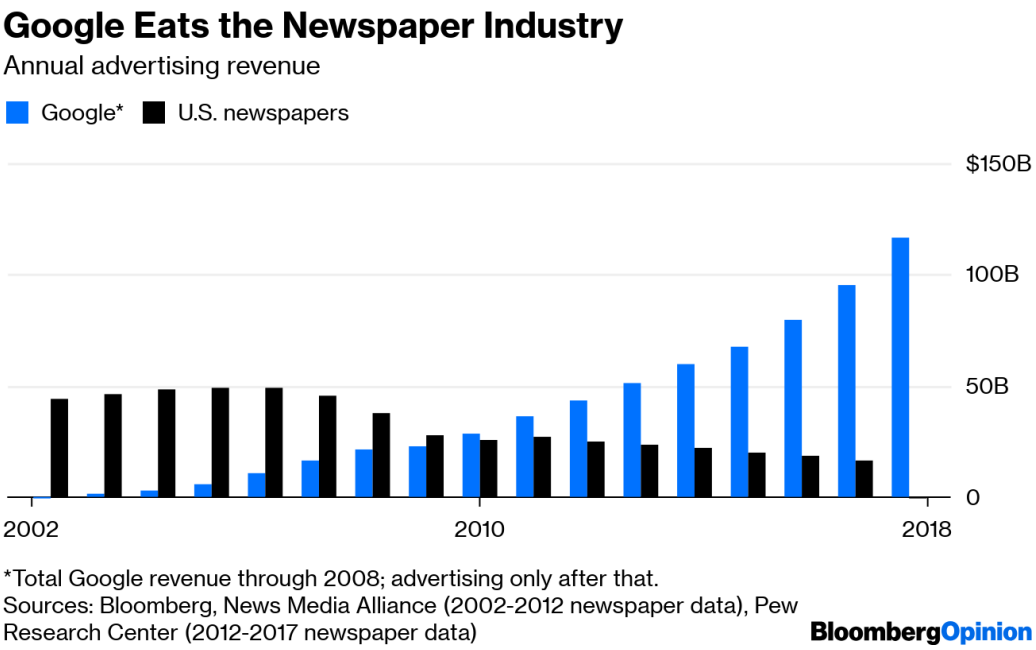
Cyfryzacja stała się źródłem wielu zagrożeń dla tradycyjnych przedsiębiorstw medialnych. Technologia zawsze była ważnym elementem rozwoju tej branży. Począwszy od druku, poprzez wykorzystywanie telegrafu, telefonu, internetu technologia zmieniała modele biznesowe mediów.

Upowszechnianie się dostępu do internetu, które nabrało szczególnego przyspieszenia po wzroście wykorzystania smartfonów – mobilnych telefonów będących komputerami – nie było obojętne dla przedsiębiorstw zajmujących się przetwarzaniem informacji. Wydawcy prasowi rozpoczęli udostępnianie swoich treści online. Pojawili się również nowi konkurenci, którzy oferowali informację online – portale, blogi, serwisy społecznościowe. Model biznesowy tradycyjnych wydawców prasy został podważony przez nowych konkurentów w zakresie dystrybucji informacji a także przez konkurentów w zakresie pozyskiwania reklamodawców. Ścieżka klienta (*customer journey*) znacząco się zmieniła w cyfrowym świecie co spowodowało przeniesienie pieniędzy reklamodawców do dużych platform cyfrowych takich jak Google (wyszukiwarka), Amazon (ecommerce) czy Meta (media społecznościowe).

Miarą zamian na rynku mediów jest porównanie przychodów reklamowych Google i branży amerykańskich wydawców prasowych zestawionych przez agencję Bloomberg. Przychody wydawców od 2010 roku są mniejsze niż Google [6].

Działalność nowych podmiotów na rynku mediów takich jak Google czy Facebook (Meta) rodzi szereg kontrowersji. Tradycyjni wydawcy twierdzą, że model biznesowy firm technologicznych narusza reguły konkurencji poprzez bezprawne i bez wynagrodzenia wykorzystywanie treści wydawców w swoim biznesie. W raporcie opublikowanym przez News Media Alliance w oparciu wywiady i konsultacje z wieloma członkami organizacji wydawcy domagają się, aby Google przestał nadużywać dominującej pozycji w relacjach z nimi, wynagradzać je sprawiedliwie za ich treści i pozwolić im na kontrolę nad konkretnym wykorzystaniem ich artykułów przez Google [7].

Rysunek 1. Przychody reklamowe Google i amerykańskiej branży wydawców prasowych.



Źródło: Bloomberg.

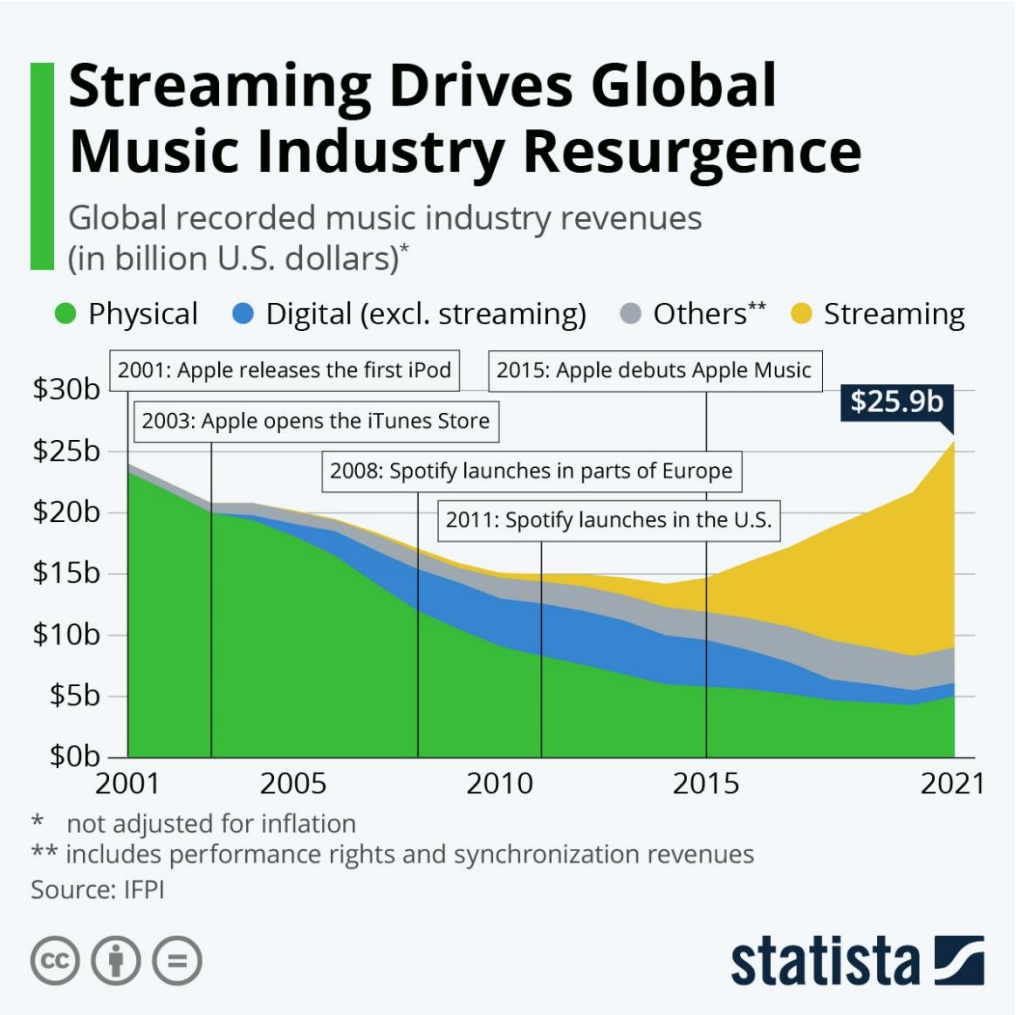
Profesor Matthew Elliott z University of Cambridge w publikacji przygotowanej na zlecenie News Media Association zwraca uwagę, że treści newsowe mają kilka cech, które czynią je szczególnie ważne z punktu widzenia budowania zaangażowania użytkowników oraz pozyskiwania danych użytkowników na potrzeby reklamy online. Po pierwsze dostawcy newsów dostarczają wiarygodnych informacji co jest związane ze znacznymi nakładami finansowymi. Po drugie dostawcy newsów dostarczają szeroki zakres tematów. Po trzecie dostarczają szybką aktualizację (fresh content). Po czwarte zainteresowanie użytkownika konkretnym tematem jest wartością dla reklamodawcy i sygnalizuje zainteresowanie zakupem samochodu, wycieczki zagranicznej albo pozwala określić wiek czy grupę zawodową. Wg badań prof. Elliotta 70 proc. użytkowników w Wielkiej Brytanii szuka w wyszukiwarce Google informacji newsowych a 90 proc. użytkowników znajduje informacje newsowe szukając czegoś innego. Wartość newsów dla Google można określić również przez analizę zapytań do wyszukiwarki. Wg analiz The News Media Alliance ponad 50 proc. wyników wyszukiwania zawiera newsy a wg badań prof. Elliotta w UK 67 proc. wyników wyszukiwania zwiera newsy.

70 proc. użytkowników powiedziało, że najprawdopodobniej zaczęliby używać innej wyszukiwarki, gdyby ta, której używają przestał używać newsy w wynikach wyszukiwania. Prof. Elliott ocenia wartość newsów dla Google i Facebook na rynku brytyjskim na ok 1 mld funtów co stanowi ok. 10 proc. ich przychodów na tym rynku [8].

Argumenty powyższe są wykorzystywane przez wydawców prasy i agencje informacyjne – najważniejszych dostawców newsów na rynkach medialnych do wpływania na regulatorów rynku i kształtowania ram regulacyjnych wymuszających na firmach technologicznych zawieranie licencji na wykorzystanie treści medialnych. Ratio legis nowego prawa wydawców – ustawy implementującej dyrektywę o prawach autorskich i prawach pokrewnych jest spostrzeżenie, że cudze treści wspomagają rozwój ekosystemów internetowych graczy. W ten sposób dochodzi do powstania tzw. „value gap", czyli sytuacji, gdy wartość tworzona przez jednych (wydawców) jest przechwytywana przez drugich (wielkie platformy), natomiast sami wydawcy nie partycypują w odpowiednim zakresie w realizowanych z tego tytułu przychodach. Założeniem nowej regulacji jest w rezultacie to, by wydawcy uzyskiwali stosowne wynagrodzenie za użytek, jakie inne podmioty czynią z ich publikacji prasowych w Internecie. Tymi podmiotami zobowiązanymi są dostawcy usług społeczeństwa informacyjnego, zwielokrotniający lub publicznie udostępniający online takie publikacje. W Polsce trwają obecnie prace legislacyjne nad wdrożeniem unijnej regulacji: w czerwcu 2022 został opublikowany projekt nowelizacji do niej się odnoszący [9].

Problemy te są przedmiotem wielu rozważań i debat. Często spotykanym poglądem jest przekonanie, że w tylko nowa technologia pomoże w rozwiązaniu wyzwań stojącymi przez wydawcami, do których przyczyniły się zmiany technologiczne ostatnich 20 lat. Analogii upatruje się na rynku muzycznym. Branża muzyczna jest ciekawym polem zilustrowania zmian, które przyniosła digitalizacja. 2021 rok by dobrym rokiem dla branży, 18,5 proc. stopa wzrostu pozwoliła na osiągnięcie wartości rynku na poziome 25.9 mld USD. To był siódmy rok wzrostu po blisko dwóch dekadach spadków. Po złotych latach 90 i upowszechnianiu CD pojawienie się formatu mp3 w internecie spowodowało 60 proc. spadek sprzedaży i przychodów branży spowodowanych pojawieniem się pirackich serwisów takich jak Napster i jego następców. W 2021 roku 65 proc. przychodów pochodziło ze streamingu. 523 mln ludzi na świecie płaci za subskrypcję muzyki. Streaming jako technologia w połączeniu z subskrypcyjnym modelem biznesowym przyniósł sukces branży po 20 latach spadków [10].

Rys. 2. Wartość globalnego rynku muzycznego w latach 2001 - 2021



Źródło: Statista.

Technologia stała się źródłem spadku wartości rynku wynikającego z utraty kontroli nad dystrybucją treści a później przyczyną wzrostu, który jest związany z przywróceniem kontroli nad dystrybucją muzyki.

Ten rodzaj myślenia przyświecał twórcom projektu - opartego na technologii Blockchain protokołu i platformy, której celem jest wywołanie systemowej zmiany w dystrybucji fotografii profesjonalnej poprzez stworzenie i przyjęcie nowego standardu certyfikacji, sprzedaży i dystrybucji fotografii.

Innowacyjny protokół technologiczny oraz współpraca medialnego ekosystemu europejskiego zapewnione przez projekt Pix.T mają w założeniu strukturalnie wpłynąć na profesjonalne praktyki dziennikarskie oraz przekształcić modele biznesowe w celu wzmocnienia rentowności, budować odporność i zapewniać wysokiej jakości treści dziennikarskie, które są niezbędne dla demokracji. Blockchain, inaczej łańcuch bloków, jest to technologia, która służy do przechowywania oraz przesyłania informacji o transakcjach zawartych w Internecie. Informacje te są ułożone w postaci następujących po sobie bloków danych. Jeden blok zawiera informacje o określonej liczbie transakcji, następnie tworzy się kolejny blok danych, a za nim kolejny, tworząc pewien rodzaj łańcucha. Mogą być przesyłane w nim informacje o różnych rodzajach transakcji, np. handlowych, kupnie lub sprzedaży walut, także kryptowalut.

Główną istotą działania blockchain jest utrzymanie wspólnej i zbiorowej księgi rachunkowej transakcji w cyfrowej postaci (ledger), rozproszonej po sieci, w takich samych kopiach. Technologia ta jest oparta na sieci peer-to-peer bez komputerów centralnych, systemów zarządzających oraz weryfikujących transakcje. Każdy komputer podłączony do sieci może brać udział w przesyłaniu oraz uwierzytelnianiu transakcji. Księga ta dzięki narzędziom kryptograficznym jest w pełni zabezpieczona przed niepowołanym dostępem, a zarazem otwarta jest dla wszystkich. Użytkownik może przejrzeć i zweryfikować całą historię transakcji od samego początku istnienia blockchain.

Technologia blockchain jest postrzegana jako technologia o dużym potencjale transformacyjnym dla branż i gospodarki. Rodzi wiele pytań o możliwość jej szerokiego wykorzystania w biznesie. Centralnym pytaniem jest możliwość zaprojektowania nowego modelu biznesowego uwzględniającego wykorzystanie technologii Blockchain w dystrybucji treści medialnych.

Przegląd literatury

W literaturze znajdujemy zarówno opisy implementacji Blockchain w zwalczaniu dezinformacji jak i próby konceptualizacji nowych modeli biznesowych uwzględniających nową technologię. Kathryn Harrison i Amelia Leopold opisują potencjał Blockchain w zwalczaniu dezinformacji. Autorki dostrzegają ważną cechę nowej technologii - zdolność Blockchain do zapewnienia dzięki zdecentralizowanej walidacji i przejrzystemu łańcuchowi dostaw potencjalnie skuteczne

narzędzie do śledzenia nie tylko środków finansowych ale wszelkiego rodzaju formy treści. Część tego, co sprawia, że tak trudno jest walczyć z deepfake'ami i innymi typami dezinformacji polega na tym, że obecnie nie ma spójnych standardów lub najlepsze praktyki w zakresie identyfikacji, etykietowania, śledzenia i reagowanie na zmanipulowane media na platformach cyfrowych. Poprzez zapewnienie większej przejrzystości w cyklu życia treści, Blockchain może zaoferować mechanizm przywracania zaufania do technologii cyfrowej. W szczególności istnieją trzy kluczowe sposoby. Rozwiązania oparte na blockchain mogą sprostać wyzwaniom stawianym przez te nowe formy cyfrowej dezinformacji:

1. Weryfikacja pochodzenia - śledzenie i weryfikacja źródeł i inne krytyczne informacje dla mediów internetowych. Publikacje wymogą korzystać blockchain, aby utworzyć rejestr wszystkich opublikowanych obrazów, dokonywanie informacji, takich jak napisy, lokalizacje, zgoda na bycie sfotografowanym, prawa autorskie i inne metadane weryfikowalne przez każdego. Na przykład New York Times bada wykorzystanie tego podejścia w ramach projektu News Provenance, który wykorzystuje Blockchain do śledzenia metadanych, takich jak źródła i edycje zdjęć do wiadomości, zapewniając czytelnikom szerszy kontekst i przejrzystość czasu i sposobu tworzenia treści.
2. Utrzymanie tożsamości i reputacji online - Kiedy czytelnicy czerpią wiadomości głównie z mediów społecznościowych, może to poważnie utrudnić im zdolność rozróżniania wiarygodnych dziennikarskich marek od treści propagandowych. W tym właśnie może pomóc Blockchain. System oparty na łańcuchu bloków może zarówno weryfikować tożsamość twórcy treści, jak i śledzą ich reputację, zasadniczo eliminując potrzebę zaufanej, scentralizowanej instytucji.
3. Mechanizmy zachęt do tworzenia treści wysokiej jakości - w obecnym krajobrazie medialnym twórcy i dystrybutorzy są silnie zachęcani do zwiększania liczby kliknięć za wszelką cenę a kliknięcia najczęściej pochodzą z treści o sensacyjnym charakterze. Jednak inteligentne kontrakty (smart contracts) zbudowane na Blockchainie oferują mechanizm automatyzacji płatności za treści, które są weryfikowane wg z góry określonych standardów jakości. Na przykład uruchomienie łańcucha bloków Civil uruchomiono w 2017 r., aby zachęcać do wiarygodności w dziennikarstwie. Użytkownicy publikujący wiarygodną informację otrzymują rekompensatę

w kryptowalucie a użytkownicy, którzy naruszają standardy społeczności publikując nieprawdziwe informacje są karani [11].

Przykład wykorzystania Blockchain w zwalczania dezinformacji opisują Mary Lacity i Dan Conway. W swojej pracy autorzy zajmują się rozwiązaniem adresującym problem fabrykowanych informacji (fake news fabrication). Autorzy przedstawiają studium przypadku jak Agenzia Nazionale Stampa Associata (ANSA) — największy włoski serwis informacyjny — oraz Ernst &Young (EY) zapobiega podszywaniu się, rodzajowi fałszywych wiadomości, które wydają się pochodzić z prawdziwych źródeł informacyjnych.

Według artykułu w czasopiśmie Science oszuści są „szczególnie zgubni, ponieważ pasożytują na standardach serwisów informacyjnych, jednocześnie korzystając z ich wiarygodności i podważając ją" Rozwiązanie, tzw. ANSAcheck weryfikuje, czy ANSA stworzyła wiadomość, aby zagwarantować to wydawcom i czytelnikom że to, co czytają, pochodzi od ANSA. Rozwiązanie wykorzystuje technologię Blockchain do tworzenia odpornych na manipulacje dowodów autentyczności źródła wiadomości i wszelkich aktualizacji.

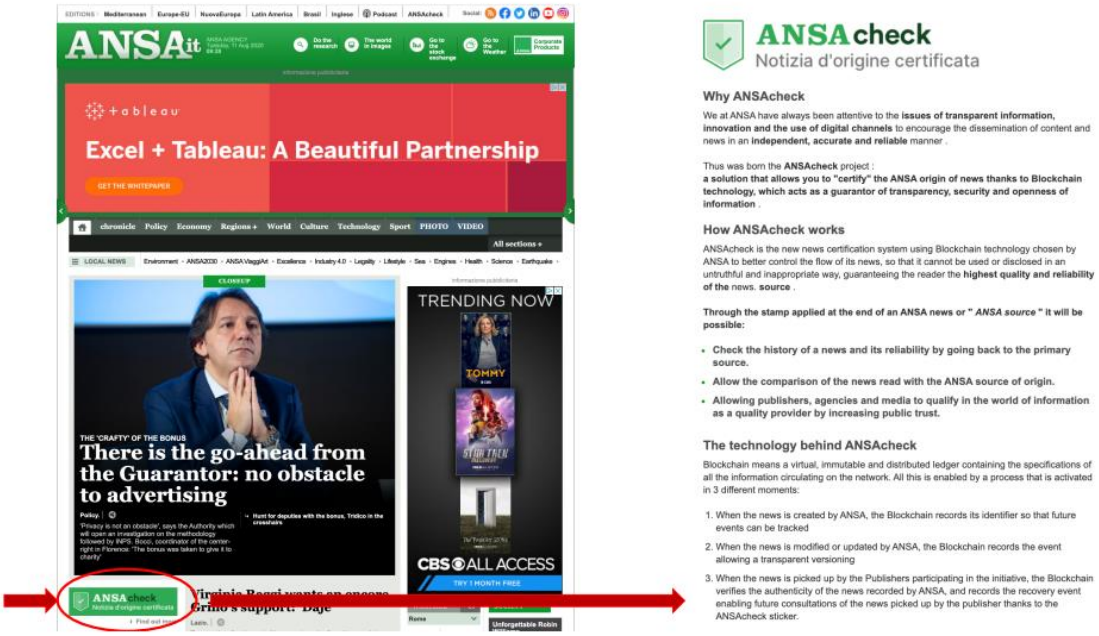
Celem implementacji rozwiązania ANSAcheck było zapewnienie gwarancji pochodzenia informacji stworzonej przez ANSA i prześledzenie tej informacji przez całą historię aktualizacji i republikacji. EY wiedziała, że technologie Blockchain mogą zapewnić autentyczności informacji odpornej na manipulacje, zapewniają identyfikowalność informacji w czasie i umożliwiają czytelnikom lub wydawcom zweryfikować informację w dowolnym momencie. ANSA uwzględniła w swoim rozwiązaniu wykorzystanie publicznej platformy Blockchain Ethereum.

Zgodnie z najlepszymi praktykami tworzenia rozwiązań opartych na Blockchain, ANSA i EY rozpoczęły przedsięwzięcie od najmniejszego możliwego środowiska (MVE – minimum viable ecosystem) w celu stworzenia i wdrożenia minimalnego opłacalnego produktu (MVP – Minimum Viable Product). Celem MVP jest przetestowanie produktu przy minimalnych zasobach, przyspieszenie nauki dzięki informacjom zwrotnym od pierwszych użytkowników i budowanie marki. ANSA, jej klienci i czytelnicy przekazują informacje zwrotne w celu ulepszenia produktu, takich jak proponowanie nowych cech i funkcjonalności.

W fazie II EY i ANSA zaproszą innych wydawców do przyłączenia się do platformy. Od września 2020 r. EY współpracuje z trzema innymi włoskimi wydawcami, którzy publikują informacje

ANSA. Rozwiązanie ANSAcheck się rozwija nie tylko we Włoszech, ale także do innych wydawców i platform mediów społecznościowych w USA i Europa. Wiele przedsiębiorstw, w tym Facebook, widziało demonstracje rozwiązania. EY zaprojektował trzyetapowy proces łączenia ANSA z wydawcami w ich ekosystemie. Na początku ANSA publikuje i autoryzuje historię, publikując ją na Blockchainie Ethereum, co było ukończone w fazie I. W trzecim kroku inni wydawcy, którzy chcą ponownie opublikować historię z systemem weryfikacji ANSAcheck zarejestruje się, aby korzystać z platformy. Wydawcy uruchomią wówczas transakcję na łańcuchu bloków przy użyciu swojego klucza prywatnego, gdy ponownie publikują historię, która została poświadczona notarialnie w systemie. W przyszłości wydawcy będą mogli dostosować oznaczenie ANSA do swojej marki, aby wyświetlała się np. na przykład oznaczenie z logo Il Corriere della Sera na stronie internetowej Il Corriere della Sera.

Rysunek 3. Strona główna ANSA z odnośnikiem do informacji o ANSAcheck



Źródło: ANSA.

„Do tej pory reakcja była dobra” – mówi cytowany w artykule De Alessandri, prezes ANSA. „Otrzymałem wiele komplementów i zainteresowanie tym rozwiązaniem. W naszym kraju jako pierwsi weszliśmy na rynek. Klienci doceniają to, że certyfikujemy się. Przekazujemy najświeższe

wiadomości, a jeśli musimy je poprawiać, poprawiamy je wkrótce potem. Jeśli zmieniamy podane przez nas wiadomości, rejestrujemy sprostowanie na Blockchainie. Jesteśmy teraz bardziej rzetelni i ostrożni niż byliśmy w przeszłości" [12].

Oprócz ANSAcheck wdrożono lub wdraża się inne rozwiązania oparte na technologii Blockchain. Wg szacunków Gartnera do 2023 r. 30 procent newsów ze świata (w tym wideo) będzie opierać się na technologiach Blockchain w zakresie uwierzytelniania. Prawdopodobnie będzie wiele rozwiązań obsługujących Blockchain, które zapewniają usługi takie jak ustalanie autentyczności treści, śledzenie pochodzenia treści w czasie, umieszczanie oszustów na czarnej liście, wykrywanie „deepfake'ów" (treści zmanipulowanych przez sztuczną inteligencję) i łączenie treści cyfrowych z świecie fizycznym, na przykład oznaczając lokalizację GPS zdjęcia.

Przedsiębiorstwa używają nowej technologii do aktualizacji modelu biznesowego albo podważenia reguł działania całych sektorów (*disrupting their industries*) [13].

Rysunek 4. Dwie klasy innowacji modelu biznesowego

TWO CLASSES OF BUSINESS MODEL INNOVATION

Blockchain is driving two classes of business model innovation in the media and entertainment industries: disruptive models, which represent potential threats to leading players, and sustaining models, which allow established companies to strengthen their businesses.

	BUSINESS MODEL	WHO IT SERVES	WHAT IT PROVIDES	HOW IT USES BLOCKCHAIN	VALUE IT GENERATES FOR THE COMPANY
DISRUPTIVE BUSINESS MODELS (THREATS)	**Monetizing content for both creators and curators**	Social media users Content creators and curators	Monetary incentives for posting and voting A decentralized, censorship-free platform	Blockchain content ledger Micropayments Cryptocurrency	Selling the power to influence Transaction fees, commissions
	Building a one-stop content shop	Digital content creators Digital content consumers	Single place for publishing, distributing, and consuming content Direct transactions between creators and consumers	Smart contracts Smart property Cryptocurrency	Transaction fees, commissions Selling original content Platform licensing Services around the open-source platform
SUSTAINING BUSINESS MODELS (OPPORTUNITIES)	**Protecting intellectual property**	Digital content creators	Simplified copyright registration and distribution of digital content	Time-stamping Smart property	Transaction fees, commissions
	Digitizing the music value chain	Existing music value chain players	Reduce transaction costs Speed up revenue distribution	Smart contracts Smart property Blockchain content ledger	Services around an open-source platform
	Playing and trading	Mobile gamers	Full off-game ownership of game assets, tradeable and sellable with cryptocurrency	Smart property Cryptocurrency	In-game asset sales

Źródło: Blockchain is changing how media and entertainment companies compete. FALL 2018 MIT SLOAN MANAGEMENT REVIEW.

W myśleniu o tym, jak Blockchain wpływa na media można dostrzec zarówno zagrożenia, jak i szanse dla graczy branżowych. Dla twórców treści Blockchain oferuje znaczące możliwości - większą kontrolę nad swoją pracą, większą elastyczność modeli licencyjnych, większy udział w przychodach z treści i uproszczenie procesów rozliczeniowych. To są wyraźne możliwości i korzyści, nawet jeśli ich urzeczywistnienie może zająć trochę czasu.

Dla agregatorów, w tym wytwórni płytowych, wydawnictw, organizacji zajmujących się prawami wykonawczymi i innych, zmniejszona rola pośredników i wydajniejsza dystrybucja dochodów w całym łańcuchu może sprawić, że będą mniej istotne, a zatem stanowić potencjalne zagrożenie. Ale włączenie technologii opartej na łańcuchu bloków do istniejących ofert może pomóc agregatorom skoncentrować się na działaniach, gdzie mogą wnieść rzeczywistą wartość dodaną (taką jak odkrywanie i wspieranie nowych talentów, finansowanie złożonych projektów takich jak filmy i programy telewizyjne oraz zapewnienie promocji i siły marketingowej).

Zagrożenia dla modelu biznesowego mediów:

a) Zarabianie na treściach zarówno dla twórców, jak i kuratorów treści. Pierwszy nowy model biznesowy polega na tworzeniu sieć społecznościowej, w której użytkownicy mogą zarabiać pieniądze (w postaci mikropłatności lub płatności w walucie cyfrowej) poprzez zamieszczenie własnych treści lub kurację i promowanie postów innych osób.

b) Budowa kompleksowego sklepu z treścią. Drugi nowy model biznesowy upraszcza łańcuch wartości, zmniejszając lub eliminując potrzebę pośredników między użytkownikami, którzy tworzą treści, a tymi, którzy je konsumują. Model eliminuje wiele z tradycyjnych etapów i warstw, takich jak agregacja i dystrybucja treści, zmniejszając w ten sposób ilość czasu potrzebnego na dostarczenie konsumentom nowych treści i realizacji przychodów. W dużym stopniu opiera się na kryptowalutach i aplikacjach opartych na Blockchain, takich jak inteligentne umowy (*smart contracts*) i inteligentna własność (*smart property*) w celu ułatwienia i przetwarzania bezpośrednich transakcje między twórcami a konsumentami.

Szanse dla modelu biznesowego mediów:

a) Ochrona własności intelektualnej. Ten biznes model wykorzystuje inteligentną własność Blockchain i aplikacje do znakowania czasem, aby pomóc artystom w przystępnej cenie chronić, udostępniać i zarządzać prawami do swoich dzieł cyfrowych.

b) Cyfryzacja muzycznego łańcucha wartości. Głównym celem tego modelu biznesowego jest optymalizacja procesu dystrybucji dochodów z muzyki pomiędzy różne strony łańcucha wartości, tak aby firmy mogły się stać bardziej elastyczne i obniżyć koszty.

c) Granie i handel. Ten model biznesowy pozwala aktywa zarejestrowane w łańcuchu bloków do sprzedaży lub handlu w innych środowiskach.

Metoda, rezultaty

Pix.T jest projektem badawczo – rozwojowym finansowanym ze środków Unii Europejskiej (European Education and Culture Executive Agency). Celem przedsięwzięcia jest wytworzenie wiedzy niezbędnej do upowszednienia w Unii Europejskiej platformy i protokołu Pix.T opartego na technologii blockchain.

W ramach grupy inicjatywnej tworzącej projekt Pix.T składającej się z przedstawicieli Polskiej Agencji Prasowej, Czeskiej Agencji Prasowej, włoskiej agencji Contrasto, holenderskiej agencji NOOR oraz francuskiej firmy WorldCrunch zespół badaczy dąży do

- zmiany nawyków fotoedytorów pracujących w mediach
- identyfikacji nowych usług dla agencji fotograficznych i mediów
- identyfikacji nowych modeli biznesowych dla agencji fotograficznych i prasowych oraz mediów
- specyfikacji połączeń platformy i protokołu Pix.T z systemami i procesami agencji
- zdefiniowania mapy drogowej integracji Pix.T z platformami social media (Facebook, Instagram, Twitter) oraz z wyszukiwarkami (w szczególności Google).

Metodologia badań oparta jest na metodzie eksperckiej wspomaganej badaniem ankietowym wśród firm medialnych w Europie. Celem badania jest uzyskanie zbieżnej wizji co do rozwoju projektu. W panelu ekspertów uczestniczą doświadczeni managerowie z pięciu firm będących

członkami konsorcjum. Badanie ankietowe zaplanowane jest na pierwsze półrocze 2023. W ramach pierwszego panelu eksperckiego, który odbył się w Paryżu we wrześniu 2022 nakreślili najważniejsze założenia strategiczne projektu.

Uczestnicy panelu zidentyfikowali następujące uwarunkowania dla strategii projektu:

- Wszystkie agencje fotograficzne i prasowe ponoszą straty z powodu luk w łańcuchu wykorzystania zdjęć online i braku kontroli nad ich dystrybucją.
- Wszyscy partnerzy współpracują z firmami zajmującymi się zwalczaniem naruszeń, aby lepiej negocjować ze swoimi klientami i próbować generować dodatkowe przychody z kradzieży zdjęć online.
- Platformy do dystrybucji zdjęć muszą być bardzo szybkie i bardzo wydajne / Konkurencja na tym rynku jest już silna.
- Wszyscy partnerzy poświęcają wiele zasobów na działania związane ze sprzedażą i raportowaniem.
- Wszyscy partnerzy obserwują ważny problem związany z fake newsami, edukacją czytelników, ale nikt nie znalazł sposobu na biznesowe rozwiązanie problemu.

W ramach dyskusji o koncepcji wartości projektu (Value Proposition) uczestnicy panelu wyrażali następujące sądy:

- Aby rozwiązać problem zakłóconego rynku (disrupted market), rozwiązanie Pix.T musi zapewniać przejrzystość i zaufanie.
- Aby rozwiązać problem zmiany rynku, który szkodzi twórcom, Pix.T powinien oferować zaszyfrowaną, zabezpieczoną i zaufaną platformę handlu sztuką cyfrową i dziennikarstwem.
- Aby rozwiązać zakłócony rynek (spadek wartości obniżający cenę) – projekt zapewnia przejrzystość i zaufanie (z wykorzystaniem Blockchain).
- Aby rozwiązać problem braku zaufania, rozwiązanie Pix.T musi zapewniać protokół Blockchain i kod embed do certyfikowania obrazu i transakcji.
- Rozwiązanie Pix.T musi zapewniać sposób zapobiegania dystrybucji zdjęć bez sprawdzania poprawności lub wymuszania nowego protokołu do przeglądania obrazów.

- Aby rozwiązać problem złych relacji między kupującymi a artystami oraz spadek przychodów, Pix.T powinien zaoferować wspólną platformę z nową filozofią biznesu fotograficznego.
- Aby rozwiązać problem swobodnego obiegu prac, Pix.T musi zapewnić alternatywę dla pobierania jpeg i nadać wartość fotografii.
- Aby rozwiązać problem braku kontroli, rozwiązanie Pix.T musi zapewnić osadzony protokół Blockchain, który zapewnia pełne śledzenie życia zdjęcia.
- Aby rozwiązać problem z własnością zdjęć w Internecie, Pix.T musi umożliwić graczom umieszczanie obrazów po to, aby móc śledzić, gdzie zdjęcia są wyświetlane i na której stronie internetowej.
- Aby dać Pix.T szansę na zrealizowanie swojego potencjału, projekt powinien pomóc znaleźć ostateczne rozwiązanie dla rewaloryzacji rynku profesjonalnej fotografii, które będzie europejską etyczną alternatywą zmuszania wielkich firm technologicznych (Big Tech) do płacenia za wykorzystanie zdjęć.

W podsumowaniu dyskusji o koncepcji wartości eksperci wspólnie wyartykułowali ją w kształcie „Aby rozwiązać problem spadku wartości w biznesie zdjęciowym, Pix.t powinien zaproponować rozwiązanie Blockchain, które zapewnia zaszyfrowane, śledzone i zaufane zdjęcia".

Eksperci zgodzili się, że problemem jest spadek wartości zdjęć, utrata kontroli nad dystrybucją i spadek zaufania w branży, zaś celem jest zmiana reguł wg których działa rynek i ustanowienie nowej logiki wartości ekonomicznej. Ponadto, konsorcjum będzie dążyć do rozwiązania problemu proponując platformę z protokołem Blockchain, która zapewnia zaszyfrowane, śledzone i zaufane zdjęcia w celu zachowania kontroli i pomiaru wykorzystania

Podsumowanie

Cyfryzacja w branży mediów, podobnie jak w innych branżach, przyniosła wiele pozytywnych zmiana ale i zmaterializowanych zagrożeń. Nowe podmioty na rynku, które zabiegają o uwagę odbiorców proponując konkurencyjne i substytucyjne produkty do tradycyjnych mediów takie jak media społecznościowe, wyszukiwarki i portale stały się głównymi graczami na rynku reklamy znacząco redukując rolę tradycyjnych mediów. Prowadzi to w konsekwencji do znaczącego pogorszenia sytuacji ekonomicznej twórców treści, w tym fotoreporterów.

Ułatwienia związane z dystrybucją informacji poza powszechnie znanymi korzyściami stały się powodem dezinformacji – znaczącego problemu społecznego podważającego procesy demokratyczne.

Technologia Blockchain, której wartość dla biznesu jest związana głównie z transparentnością i brakiem konieczności wzajemnego zaufania (trustless system) gwarantująca pewność obrotu, w tym obrotu treści medialnych ma duży potencjał transformacji dla branży mediów.

W dostępnej literaturze przedmiotu autorzy zwracają uwagę głównie na potencjał wykorzystania Blockchain w walce z dezinformacją związany z możliwością weryfikacji pochodzenia treści, utrzymywania tożsamości i reputacji online oraz właściwymi Blockchain mechanizmami zachęt do tworzenia treści wysokiej jakości. Potencjał nowej technologii pozawala również na efektywne mechanizmy zarzadzania prawami autorskimi online oraz tworzy nowe mechanizmy monetyzacji treści.

Projekt Pix.T dąży do rozwiązania problemu utraty kontroli nad dystrybucją zdjęć reporterskich w internecie poprzez wykorzystanie technologii Blockchain.

Większa kontrola nad obrotem treściami medialnymi jest warunkiem wstępnym odnowienia modelu biznesowego wytwórców treści i przywrócenia rentowności obrotu w branży medialnej. Punktem odniesienia dla twórców projektu Pix.T jest branża muzyczna, w której technologie cyfrowe najpierw przyczyniły się do znaczącego spadku przychodów, żeby po implementacji streamingu zwiększyć przychody jako rezultat odzyskania kontroli nad dystrybucją.

Duże firmy technologiczne (Big Tech) zbudowały swoją pozycję na przewadze konkurencyjne związanej z przetwarzaniem informacji i skali działania. Blockchain będący swoistym internetem wartości zbudowanym na odmiennym paradygmacie do sieci Internet wymusza przetwarzanie zdjęć jako obiektów związanych z wartością a nie zbiorem danych zasilających algorytmy wyszukiwarek i sieci społecznościowych.

Nowa architektura informacji oparta na technologii Blockchain jest szansą na przywrócenie ekonomicznej wartości treściom medialnym oraz przywrócenie zaufania do treści informacyjnych i tym samym ograniczenie zjawiska dezinformacji.

Bibliografia

[1] G. Osbourne, *Strategic Communications: Insights from the commercial sector*, NATO Stratcom CoE, http://www.stratcomcoe.org/strategic-communications-insights-commercial-sector.

[2] Jeffrey Gottfried and Elisa Shearer. 2017. Americans' online news use is closing in on TV news use. Pew Research Center 7 (2017).

[3] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. Science 359, 6380 (2018), 1146–1151.

[4] A. Hulcoop, J. Scott-Railton, P. Tanchak, M. Brooks, and R. Deibert, Tainted leaks. *Disinformation and Phishing With a Russian Nexus*, The CitizenLab, https://citizenlab.ca/2017/05/tainted-leaks-disinformation-phish/.

[5] Puls Medycyny, https://pulsmedycyny.pl/gigantyczne-straty-firmy-produkujacej-insuline-przez-falszywy-wpis-na-twitterze-1169296.

[6] Justin Fox, Google May Employ More People Than the Entire US Newspaper Industry, Bloomberg 17 February 2019.

[7] How Google Abuses Its Position as a Market Dominant Platform TO STRONG-ARM NEWS PUBLISHERS AND HURT JOURNALISM. News Media Alliance, June 2020.

[8] Value of News to Digital Platforms in the U.K. Professor Matthew Elliott, University of Cambridge, published by News Media Association.

[9] https://www.wirtualnemedia.pl/artykul/dyrektywa-w-sprawie-prawa-autorskiego-i-praw-pokrewnych-dziesiec-negocjacyjnych-przykazan-dla-wydawcow-prasy.

[10] Felix Richter, Streaming Drives Global Music Industry Resurgence, Statista, Apr 8, 2022.

[11] Kathryn Harrison and Amelia Leopold. How Blockchain Can Help Combat Disinformation. Harvard Business Review. July 2021.

[12] Mary Lacity, Dan Conway. Authenticating real news with ANSAcheck, a blockchain-enabled solution developed by ANSA and EY. Blockchain Center of Excellence. University of Arcansas.

[13] Andre Dutra, Andranik Tumasjan, Isabell M. Welpe. Blockchain is changing how media and entertainment companies compete. FALL 2018 MIT SLOAN MANAGEMENT REVIEW.

Aleksander Żołnierski, Adam Gulczyński, Jacek Gulczyński

Poszukiwanie ukrytych kompetencji – grafen i jego zastosowania w medycynie

Wprowadzenie

W ostatnich latach w sposób bezprecedensowy wzrasta liczba dostępnych danych, które mogą i są wykorzystywane dla optymalizacji procesów transformacji i dostosowania potencjału organizacji do zmieniających się uwarunkowań zewnętrznych. Zmiany zachodzące w świecie, a szczególnie rozwój technologii materiałowych a pośród nich nowych form terapii medycznych pozwala z nadzieją patrzeć w przyszłość. Instytucje sektora nauki, badań i rozwoju oraz przedsiębiorstwa zaangażowane w badania nad nowymi materiałami stają przed wyzwaniem, jakie niesie identyfikacja specyficznych i unikalnych kompetencji organizacyjnych, które mogą przyczynić się do sukcesu, w tym komercjalizacyjnego. Taka sytuacja dotyczy wielu dziedzin technologii materiałowych, które wytwarzają wiele produktów, coraz to nowszych, bądź też ich kolejnych modyfikacji. Pośród nich poczesną rolę mają w badania nad grafenem.

Sam grafen jest strukturą laminarną, jednowarstwową, o grubości jednego atomu, zbudowaną z atomów węgla w układzie warstwy przestrzennie przypominającej plaster miodu. W zależności od procesu technologicznego uzyskuje się płatki różnej wielkości (co potem wpływa na ich właściwości fizyko-chemiczne, a w konsekwencji bezpośrednio wpływa na obszar użytkowania), bądź też struktury nanorurek.

Pierwsze analizy dotyczące hipotetycznej jednoatomowej warstwy węgla sięgają lat 50-tych dwudziestego wieku. Boehm i wsp. w 1962 roku wprowadzili pojęcie „graphene” łącząc słowo

grafit (graphite) z końcówką pochodzącą z chemicznych związków węglowodorowych. Natomiast fizycznie udało się to dopiero w roku 2004 badaczom z Uniwersytetu w Manchesterze. Dokonali tego Andrei Geim i Konstantin Novoselov (wcześniej współpracujący na Uniwersytecie w Nijmegen), którzy w roku 2010 za wyniki swych prac otrzymali nagrodę Nobla.

Spektrum zastosowań grafenu i jego pochodnych wydaje się nieograniczone. Patrząc choćby na mapę drogową projektu Graphene Flagship widzimy, że obszary badań dotykają praktycznie każdej dziedziny współczesnego życia. Niektóre badania już wykonane bądź prowadzone obecnie, już wyprzedzają założenia tego projektu, choćby transport leków (*drug delivery*) „zaplanowane' na lata 30te XXI wieku. Zastosowania grafenu związane z medycyną wskazują na zidentyfikowane przez nas kompetencje jednostek naukowych zaangażowanych w badania nad tym materiałem. Identyfikacja takich nowych – z punktu widzenia tych jednostek naukowych, kompetencji może wpłynąć na tworzenie nowych specjalizacji strategicznych w tych jednostkach (czego pierwszym potwierdzeniem jest wniosek patentowy złożony przez trzy, pozornie niezwiązane merytorycznie jednostki naukowe – Instytut Technologii Materiałów Elektronicznych [obecnie: Instytut Mikroelektroniki i Fotoniki], Gdański Uniwersytet Medyczny i Instytut Nauk Ekonomicznych Polskiej Akademii Nauk).

Organizacje związane z tworzeniem i badaniami nad tą technologią stoją przed wyzwaniem identyfikacji określonych i pożądanych specyficznych kompetencji i tworzenia warunków dla wykorzystania specyficznej wiedzy, która może zadecydować o możliwościach zastosowań grafenu, w tym omawianym przypadku także w medycynie. W literaturze przedmiotu podkreśla się znaczenie nie tylko kompetencji technicznych, naukowych, czy kompetencji miękkich, ale zdolności do synkretycznego łączenia wiedzy z różnych, często odległych, pozornie niezwiązanych ze sobą, dziedzin i budowania wielospecjalistycznych zespołów badawczo-wdrożeniowych. Nowe wyzwania z tym związane są niezbędne dla skutecznego wykorzystania zasobów ludzkich i zasobów niematerialnych związanych ze wzrostem znaczenia aktywów niematerialnych i prawnych – zatem specjalistycznej wiedzy dziedzinowej. W praktyce organizacji, takie poszukiwanie czy też odkrywanie (często ukrytych) kompetencji polega na stworzeniu potencjału (w zakresie zasobów ludzkich, technologicznych i organizacyjnych) dla wykorzystania dostępnych danych i zamienić je w istotne informacje niezbędne dla optymalizacji podejmowania decyzji zarządczych (Stanton i in., 2016).

W identyfikacji kompetencji ważnym procesem jest rozpoznawanie cech ludzi, które nie są „niezmienne" (Civelli, 1998). Zmienne symboliczne, kulturowe, społeczne i wartości stają się fundamentalne. Zmiana polegająca na odchodzeniu od deterministycznych, czysto mechanistycznych procesach polegających na prostym dostosowaniu oferty do popytu ustąpiło jeszcze w latach '90 XX wieku podejściu „opartym na kompetencjach". Kwestię kompetencji, która przy różnych podejściach jest przedmiotem analiz w organizacji, nie można jednak łatwo powiązać z problemem zatrudnialności. Wynika to z coraz bardziej rozmytych granic między tym, czego organizacja potrzebuje a twardymi i miękkimi kompetencjami dostępnymi po podażowej stronie rynku pracy. Identyfikacja właściwych kompetencji, które pozwalają na sukces organizacji jest coraz bardziej złożone. Organizacje nie tylko nauczyły się już rozpoznawać i rozumieć „język zachowań", ale coraz skuteczniej identyfikują wachlarz niezbędnych kompetencji – zarówno indywidualnych, jak i kompetencji samej organizacji w oparciu o zaawansowane narzędzia analityczne wykorzystujące algorytmu sztucznej inteligencji i narzędzia analiz big data.

Kompetencje są ważnym źródłem przewagi konkurencyjnej, jednak wiele firm napotyka na trudności przy próbie ich identyfikacji i ocenie. Teoria określająca kompetencje jest nieprecyzyjna, a same kompetencje zdają się niejednoznaczne. Wpływa to na zróżnicowane postrzeganie kompetencji przez menedżerów, co jest wynikiem ich indywidualistycznego podejścia. Powoduje to często poważne konsekwencje dla organizacji (King et al., 2001).

Większość opisanych w literaturze metod modelowania kompetencji i wydajności to modele deterministyczne, które pomijają istotny zakres niepewności oraz złożonych i nieliniowych relacji w organizacji. Ten fakt stanowi istotną barierę nie tylko dla prognozowania, ale dla oceny wpływu ukrytych zależności i kompetencji organizacyjnych. Wykorzystanie sztucznych sieci neuronowych w systemach neuro-rozmytych (NFS) czy modelach adaptacyjnych neuro-rozmytych systemów wnioskowania wychodziło naprzeciw tworzeniu ulepszonych modeli predykcyjnych. Jednak i te modele mają pewne ograniczenia. W celu obsługi modelu z wieloma wejściami i wyjściami projektuje się systemy hybrydowe, które łączą ewolucyjną technikę optymalizacji algorytmu genetycznego (GA) z wielowyjściowym adaptacyjnym neuro-rozmytym systemem wnioskowania (MANFIS) (Tiruneh et al., 2022). System taki pozwala na jednoczesną predykcję wielu wskaźników wydajności organizacji przy użyciu kompetencji organizacyjnych. System wiąże kompetencje organizacyjne z wynikami i przewiduje wiele miar efektywności

organizacyjnej oraz – dzięki algorytmom genetycznym zmniejszającym wymiarowość danych – umożliwia identyfikację kompetencji organizacyjnych, które znacząco wpływają na wydajność organizacji.

Jeszcze kilkanaście lat temu, do oceny kompetencji organizacyjnych stosowano równolegle z metodami opartymi na technologiach informacyjnych bardziej tradycyjne rozwiązania. Wykorzystywano metody analizy treści dokumentów organizacyjnych oraz pogłębione wywiady ze specjalistami korporacyjnymi. Takie tradycyjne podejście ułatwiło zrozumienie zjawisk związanych z komunikacją wewnątrz organizacji, z wewnętrznymi sieciami relacji i zarządzaniem dokumentami. Metodologia polega na wyłonieniu kilku badanych komponentów, w tym podstawowych – technologicznych, produktowych oraz procesów, umiejętności funkcjonalnych, technologicznych i zintegrowanych. Badanie obejmowało analizę interakcji między tymi komponentami. Kompetencje organizacyjne są rozwijane w oparciu o wizję organizacji i umożliwiają rozwój miękkich kompetencji oraz wiedzy (Edgar, Lockwood, 2008).

Identyfikacja unikalnych kompetencji organizacyjnych jest zadaniem, które można uznać za kluczowe dla organizacji w dynamicznie zmieniającym się środowisku zewnętrznym. Przykładem mogą być ramy teoretyczne identyfikacji unikalnych kompetencji centrów innowacji technologicznych (Balbinot et al., 2012). Badania empiryczne, które opierały się na „tradycyjnych" metodach badań społecznych obejmowały fazę jakościową i ilościową. Zidentyfikowano w ten sposób unikalne kompetencje organizacyjne takie jak: własność intelektualna, patenty krajowe, usługi doradcze świadczone przez indywidualnych naukowców, obszary doskonałości instytucji naukowych i technologicznych mających istotny wpływ na wyniki działalności naukowej, w tym liczbę publikacji naukowych, deklarowane kierunki badań oraz współpracę badawczą. Nietrudno wykazać, że zidentyfikowane obszary – przede wszystkim za sprawą zastosowanych metod badawczych – nie pozwalają na rzeczywiste zmapowanie ukrytych kompetencji.

Identyfikacja kluczowych z punktu widzenia organizacji kompetencji idzie w parze z praktyką definiowania mierzalnych czynników sprzyjających wprowadzania zmiany organizacyjnej. Czynniki te kształtują „gen zmiany", który zapewnia organizacji przewagę strategiczną, a obejmuje pięć elementów: sformułowanie celu, przywództwo, strukturę, możliwości i kulturę

(Kuzmanova, 2012). Integracja tych obszarów wymaga kompetencji miękkich i zaufania dla tworzenia atmosfery sprzyjającej zmianom w organizacji.

Perspektywa Oparta na Kompetencji (ang. CBP) w zarządzaniu strategicznym ma ograniczone zastosowanie praktyczne. Podejmowano próby włączenia do procedury pomiaru kompetencji metod znanych z procesów zarządzania jakością (ang. QM), co w niewielkim stopniu poprawia jakość zastosowania CBP (Escrig-Tena et. al., 2005).

Wykorzystanie wewnętrznych systemów informacyjnych do określania kompetencji organizacyjnych znajduje zastosowanie od początku wykorzystania IT w organizacjach. Przykładem może być system informacji księgowej (ang. AIS). W zmieniającym się otoczeniu wykorzystanie AIS jest niezbędne nie tylko dla podnoszenia wydajności. Na sprawność organizacyjną wpływ ma synergia trzech kompetencji: elastyczności systemu, komplementarnego systemu business intelligence oraz kompetencji technicznych kadr wykorzystujących systemy informacyjne w organizacji (Prasad, Green, 2015). Połączenie tych kompetencji pozwala na podnoszenie ogólnej wydajności i stanowi samo w sobie wartość dla organizacji.

Sama identyfikacja i zmapowanie ukrytych kompetencji, które mogą stanowić kompetencje kluczowe jest dopiero połową sukcesu. Właściwe włączenie zidentyfikowanych kompetencji ukrytych w procesy biznesowe ujawnia dopiero rzeczywisty potencjał i zdolności organizacji do skutecznego konkurowania. Przykładem na włączanie kompetencji ukrytych (czasami w sposób nieuświadomiony) w procesy biznesowe jest marketing innowacyjnych firm sektora MSP (Chaston, 1997). Istotnymi stają się zdolności do budowania długoterminowych relacji z klientami. W tym kontekście wybór adekwatnych, unikalnych strategii marketingowych pozwala na poprawę ogólnej wydajności małych firm produkcyjnych. Strategia w tym zakresie wpływa korzystnie także na podnoszenie „tradycyjnych" kompetencji w obszarach takich jak HR, produktywność pracowników, zarządzanie jakością i kreatywność.

Już samo zastosowanie podejścia opartego na kompetencjach w organizacjach nie nadąża za postępami w strategicznym zarządzaniu zasobami ludzkimi. Pojawiające się potrzeby zmian uwidaczniają się w coraz bardziej zorientowanym na przyszłość i strategicznym kontekście (Sparrow, 1995). Narzędziem wspierającym identyfikację kompetencji może być system Business Intelligence. Posługiwanie się systemami BI pozwala na grupowanie kompetencji na kategorie, np. związanymi z zarządzaniem danymi, systemami informatycznymi i technologią

informacyjną, zasobami finansowymi, zarządzaniem relacjami i kapitałem ludzkim (Salmasi et al., 2016). Grupowanie kompetencji pozwala ponadto na wykorzystanie ich, jako miary oceny organizacji jako całości.

Okazuje się, że globalnie zintegrowane przedsiębiorstwo wymaga fundamentalnie różnego podejścia do produkcji, dystrybucji i alokacji zasobów ludzkich. Podejście opiera się na aktywnym zarządzaniu operacjami, wiedzą specjalistyczną i możliwościami oraz na bliższych relacjach z partnerami, dostawcami i klientami (za Hewitt, Lesser, 2007). Globalnie zintegrowana organizacja, w tym także jednostka sektora B+R to organizacja skoncentrowana na łączeniu i korzystaniu z różnych źródeł technologii, wiedzy i – generalnie – zasobów. W taki sposób tworzona jest wartość, niezależnie od tego, gdzie „podstawowe" zasoby są zlokalizowane. Integracja wewnętrznych i zewnętrznych interesariuszy kształtuje produkt i tworzoną wiedzę. Można analizować te procesy w rozmaitych modelach, z których popularnym był jeszcze kilka lat temu model potrójnej helisy, gdzie pracownicy, klienci, partnerzy, agencje rządowe i inne strony trzecie mają udział w tworzeniu innowacji. Mimo globalizacji, wykorzystaniu systemów informacyjnych, integracji i dywersyfikacji łańcuchów dostaw, ciągle kluczowymi są zasoby ludzkie, które dynamicznie kształtują kompetencje organizacyjne. Hewitt i Lesser podkreślają, że w globalnie zintegrowanym świecie, organizacja musi przede wszystkim zbudować zestaw podstawowych kompetencji mających na celu optymalizację zasobów ludzkich. Jest to kluczowe dla utrzymania przewagi konkurencyjnej i skuteczności procesów innowacyjnych. Autorzy podkreślają, że budowanie i utrzymywanie zasobów ludzkich dla globalnie zintegrowanej organizacji wymaga współoddziaływania siedmiu kompetencji kluczowych. Identyfikacja zasobów i kompetencji wymaga udzielenia odpowiedzi na kilka pytań: - czy organizacja będzie miała wystarczającą liczbę pracowników, aby wspierać realizację określonej strategii? – czy organizacja posiada odpowiednią wiedzę, umiejętności i kompetencje do realizacji strategii? …wreszcie – czy normy i wartości organizacyjne wspierają opracowanie i wdrożenie strategii?

Wiele badań wskazuje, że zewnętrzne doradztwo i szkolenia może w pewnym tylko stopniu korzystnie wpłynąć na identyfikację i pełniejsze wykorzystanie kompetencji. W przedsiębiorstwach szkolenia wpływają na wyższy zwrot z inwestycji przeznaczonych na rozwój zasobów ludzkich (Vveinhardt, Stonkute, 2015). Aby skuteczność takich zabiegów była

wyższa, niezbędne jest zaistnienie w organizacji adekwatnego środowiska ukierunkowanego na dzielenie się nową wiedzą.

Dla identyfikacji kompetencji stosowane są różnorodne metody. Jedną z nich była metodologia Co-Evolute stworzona na Uniwersytecie Technologicznym w Tampere, która wykorzystywana była do analizowania proaktywnych wizji personelu, co miało służyć rozwojowi osobistemu i organizacyjnemu w firmach technologicznych. Metodologia ta nakierowana była na problematykę uczenia się i tworzenia wiedzy. Metodologia opierała się na wykorzystaniu kilku wyspecjalizowanych aplikacji do oceny kompetencji „fizycznych" i oceny kompetencji ogólnych i szczególnych cech kierownika projektu. Kolejne dwie aplikacje oceniały środowisko organizacyjne z punktu widzenia tworzenia nowej wiedzy i punktu widzenia uczenia się (Paajanen et al., 2009).

Kształtowanie kompetencji – poszukiwanie ich i identyfikacja oraz rozwój są przede wszystkim wynikiem oddziaływań funkcji zarządzania. Szczególną, ale przede wszystkim specyficzną rolę odgrywa przy tym funkcja kontroli. Przykładowo, międzyorganizacyjne kontrole zarządzania opierające się na otwartej księgowości i zarządzaniu kosztami docelowymi oraz analizie funkcjonalnej, stwarzają nowe możliwości interwencji kierownictwa w procesy kształtowania kompetencji organizacyjnych. Kontrole umożliwiają identyfikację i obserwację procesów korporacyjnych, technologii, organizacji i strategii, a tym samym są pomocne w formułowaniu i propagacji informacji o „tożsamości" lub „podstawowych kompetencjach" firm (Mouritsen et al., 2001). Z tego punktu widzenia odgrywają rolę strategiczną i wyznaczają nie tylko ramy organizacji, ale pozwalają na jej jednoznaczną identyfikację w otoczeniu rynkowym.

Kompetencje organizacyjne są z jednej strony efektem zamierzonych działań menedżerów i pracowników, z drugiej – efektem oddziaływań czynników zewnętrznych. Sama orientacja rynkowa organizacji przyczynia się do tworzenia specyficznych kompetencji organizacyjnych, które z kolei prowadzą do lepszych wyników. Dotyczy to takich obszarów jak np. ograniczania kosztów, wzrost przychodów, zdobywanie i utrzymanie klientów, wprowadzanie nowych produktów (Subramanian et al., 2009). Można powiedzieć, że oddziaływania są tu dwukierunkowe i symultaniczne – zestaw kompetencji prowadzi do zorientowania organizacji na rynek, a sama orientacja rynkowa kształtując (czasami wymuszając) pewne specyficzne cechy i kompetencje – prowadzi do sukcesu organizacyjnego.

Formułowanie i rozwój kompetencji organizacyjnych, a zatem formułowanie strategii w globalnych organizacjach ponadnarodowych łączy w swych ramach teoretycznych koncepcje zarządzania strategicznego, zasobowego punktu widzenia firmy i zarządzania operacjami międzynarodowymi. Model analityczny umożliwiający ocenę kompetencji organizacyjnych takich firm wychodzi od analizy konfiguracji produktowej/rynkowej oraz strategii konkurencyjnej. Dowody empiryczne pochodzące z badań w przemyśle telekomunikacyjnym wskazują, nastąpiła radykalna zmiana strategicznego znaczenia kompetencji organizacyjnych. To spowodowało, że zarówno operatorzy sieci, jak i dostawcy technologii w początkach nowego millenium przeszli od strategii opartych na technologii i operacjach do strategii opartych na świadczeniu usług (Fleury, Fleury, 2003).

Identyfikacja nowych, często nieoczywistych kompetencji i lokalizacja ich, jako kluczowe odbywa się praktycznie we wszystkich sektorach. Przykładem może być rynek ubezpieczeń i zadania w zakresie zarządzania finansami w tym obszarze. Modele płatności zorientowane na wartość wymusiły w systemie ochrony zdrowia działania zmierzające do ustabilizowania pozycji organizacji w tym stosunkowo nowym dla nich środowisku. Przykładowo, Medicare Shared Savings Program połączył zakładane zrównoważone cele finansowe z wymogiem tworzenia wartości (Kotzbauer, Weeks, 2015). Podjęcie nowych wyzwań wymaga identyfikacji kluczowych czynników, które definiują strategię transformacji biznesowej. Wydaje się, że wspólnym, dla wszystkich ścieżek transformacji działań zaliczyć można modernizację systemów technologicznych i przyjęcie za warunek funkcjonowania – ograniczony dostęp do kapitału finansowego. Transformacja do nowych modeli lub warunków zewnętrznych wymaga zdolności całego zespołu do zrozumienia i przyjęcia na nowo zdefiniowanej misji. Nowy model biznesowy będzie zaakceptowany i stabilny (co też potencjalnie może przyczynić się do ego sukcesu) tylko wtedy, gdy liderzy będą posiadać kompetencje pozwalające im na zarządzanie transformacją kulturową. Aby odnieść sukces, kompetencje muszą być budowane wokół silniejszych relacji wewnątrzorganizacyjnych, kiedy to menadżerowie skutecznie komunikują cel planowanych zmian i ścieżkę działań potrzebnych do osiągnięcia sukcesu i wdrożenia strategii transformacji. Poszukiwanie, identyfikacja i zarządzanie kompetencjami wymaga wielu działań wspieranych przez wewnętrzny marketing i narzędzia motywujące pracowników. Wykorzystanie marketingu wewnętrznego do realizacji programu transformacji wspiera realizację dowolnej strategii funkcjonalnej (Ahmed et al., 2002).

Właściwa identyfikacja kompetencji organizacyjnych wspiera uzyskanie przewagi konkurencyjnej. Przewaga konkurencyjna to wykorzystanie zasobów i możliwości do rozwijania kompetencji organizacyjnych, które z kolei tworzą wartość dla klientów (Sago, 2003). Przewaga konkurencyjna niekoniecznie musi być budowana na unikalnym zestawie cech – przykładem mogą tu być organizacje skoncentrowane na działalności handlowej. Zwiększenie wartości klienta w efekcie zarządzania ofertą a zatem uzyskanie trwałej przewagi konkurencyjnej jest wzmacniane utrudnionym powieleniem cech organizacji – jej produktów, metod działania i strategii. Przewaga konkurencyjna, a zatem także zestaw charakterystycznych dla organizacji kompetencji jest wyceniane przez rynek, a same kompetencje organizacyjne są podstawą osiągania tej przewagi. Kompetencje organizacyjne są tworzone w oparciu o strategiczne wykorzystanie dostępnych zasobów i zdolności organizacyjnych. Właściwa identyfikacja kompetencji organizacyjnych musi opierać się na ich gradacji w obrębie trzech kategorii: kompetencji wyróżniających, wspólnych i nieistotnych. Identyfikacja każdej z kategorii jest o tyle istotna, że np. kompetencje nieistotne angażując ograniczone lub krytyczne zasoby firmy powodują koncentrację na obszarach, które z punktu widzenia strategii są nieistotne. Z tego punktu widzenia kompetencje takie mogą mieć negatywny wpływ na firmę.

O narzędziach i metodzie

Poszukiwanie kluczowych kompetencji, często może polegać na znalezieniu charakterystycznych dla zdolności organizacji w zakresie transferu wiedzy w oparciu o wykorzystanie łańcucha wartości. Takie podejście może być skuteczne z punktu widzenia jednostki naukowej, na której spoczywają nowe zadania w zakresie komercjalizacji i związanym z tym wymogiem generowania z tego tytułu przychodów. Transfer wiedzy w oparciu o wykorzystanie łańcucha wartości wzmacniany jest obecnością globalnych nabywców na lokalnym rynku, na jakim działa organizacja. Dzieje się tak poprzez tworzenie relacji opartych na wiedzy oraz transferem kompetencji technicznych i organizacyjnych wzdłuż łańcuchów wartości co wynika z procesów dostosowania technologii do „lokalnych kontekstów" (Saliola, Zanfei, 2009). W kontekście organizacji i zarządzania, wiedzę literatura przedmiotu dzieli w oparciu o znane typologie. Założenia dotyczące wiedzy opierają się na analizie wpływu nowych technologii, oddziaływania prądów filozoficznych, teorii społecznej, językoznawstwie, kognitywistyce ale także na pragmatycznie zaplanowanych procesach (Blackler, 1995).

Na wiedzę organizacji, a zatem i na możliwości identyfikacji kompetencji wpływ ma system kulturowy, a jakim organizacja jest zanurzona – kształtujący zarówno sposoby pozyskiwania wiedzy, jak i procesy jej internalizacji oraz wykorzystania.

Pojawienie się narzędzi big data miało znaczący wpływ na analitykę biznesową w obszarach zarówno samych rozwiązań informacyjnych, jak i stosowanych metod ilościowych czy procesów podejmowania (Cegielski et al., 2016). Odnosi się to zarówno do organizacji biznesowych, jak i sektora nauki, które generują podaż specjalistów i mają niebagatelny wpływ na kształtowanie formalnych kompetencji organizacji, do których trafiają absolwenci uczelni.

W naszym badaniu posługujemy się narzędziami analizy big data, analizami statystycznymi i metodyką opartą na zaawansowanych technikach analizy tekstu.

Materiał analityczny stanowił zestaw publikacji naukowych (35.000 artykułów naukowych dotyczących grafenu i technologii z nim związanych), który traktujemy jako bazę do rozważań na temat kompetencji organizacji, której główną (jawną) kompetencją są badania nad grafenem płatkowym. Identyfikując materiał analityczny zdefiniowaliśmy najpierw parametry brzegowe, którymi były publikacje naukowe wybranej jednostki naukowej oraz skategoryzowaliśmy słowa kluczowe. Te działania wyznaczyły strategię badawczą. Posłużyliśmy się dostępnymi narzędziami wyszukiwania informacji pełnotekstowej, w tym przede wszystkim wyszukiwarkami specjalistycznymi dla wyszukań artykułów naukowych. W ten sposób zasilona stworzona na potrzeby naszego badania baza posłużyła do dalszych analiz z wykorzystaniem statystycznych narzędzi analizy tekstu. Opracowaliśmy wyniki analizy słów kluczowych, wykorzystaliśmy analizę sentymentów w oparciu o słowniki. Ponadto przygotowaliśmy analizę n-gram oraz wykorzystaliśmy narzędzie do stworzenia chmury tagów dla zilustrowania wyników naszego badania.

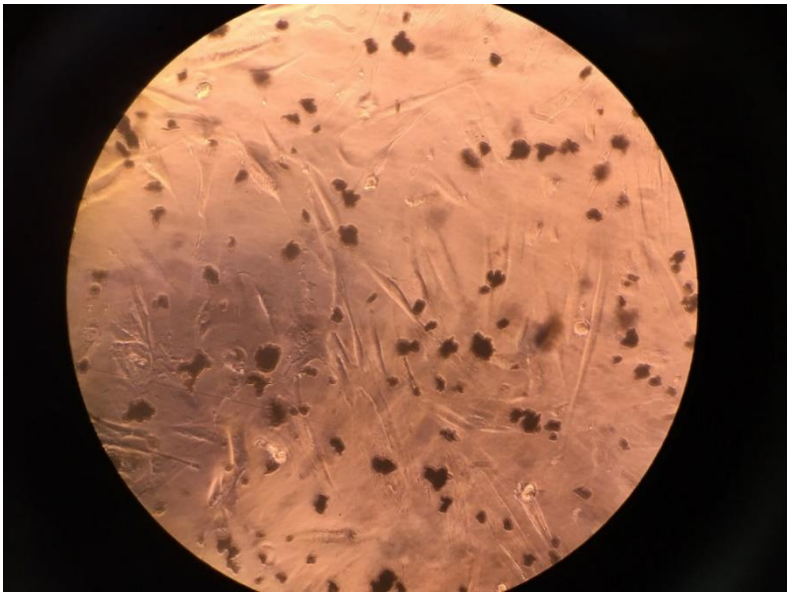
Wykorzystane narzędzia analizy big data dają dobre wyniki dla identyfikacji technologii związanych z rozwijającymi się, nowymi dziedzinami nauki i badań. Podejście oparte na analizie danych niestrukturyzowanych znalazło już swoje zastosowanie w badaniach trendów technologicznych (Cetera et al., 2022). Ponadto, analiza danych strukturyzowanych z publikacji naukowych i innych danych źródłowych (bazy patentowe) czy danych niestrukturyzowanych poszukiwanych i kolekcjonowanych ze źródeł online dają ciekawe i obiecujące wyniki w zakresie zastosowań praktycznych grafenu (Żołnierski et al., 2022).

W naszych badaniach wykorzystaliśmy narzędzia do identyfikacji ukrytych kompetencji, w szczególności analizy semantycznej big data z wykorzystaniem zaawansowanych metod przetwarzania wielkich zbiorów danych. Wynikiem badania jest zestaw danych wynikowych, które posłużyły nam w drodze analiz (m.in. bigram i trigram) do identyfikacji technologii, które znajdują zastosowanie w medycynie, terapii i badaniach związanych i wykorzystywanych w opiece zdrowotnej.

Wyniki badania

Specyficzne właściwości grafenu, jak jego lekkość, elastyczność, wytrzymałość, twardość, ale także przewodnictwo cieplne czy elektryczne definiują ogromny obszar, w którym można się pokusić o jego użycie. Wykorzystanie zmiennych właściwości fizyko-chemicznych grafenu obejmują materiały kompozytowe, energetykę, elektronikę, analizę danych i zastosowania informatyczne oraz, co stanowi główną część tej pracy - medycynę z biotechnologią. To właśnie w medycynie tworzone są modele wykorzystujące bardzo specyficzne właściwości grafenu. Czysto fizyczne wykorzystuje się w budowaniu lekkich struktur podporowych w inżynierii tkankowej. Zdolność przewodnictwa zwłaszcza przy minimalnych potencjałach pozwala na tworzenie ultralekkich, trwałych i ultraczułych sensorów do umieszczenia podskórnie, w okolicy połączeń kostno-mięśniowych, czy chociażby w obrębie ośrodkowego układu nerwowego [Youxiao Ma et al. 2017, Huang et al. 2020].

Rysunek 1. Linia komórkowa zdrowych ludzkich fibroblastów w hodowli in vitro bez negatywnego wpływu będących w otoczeniu cząsteczek grafenu



Źródło: badania własne GUMed.

Wielkość cząsteczek ma ogromy wpływ na właściwości fizyczne i chemiczne, przy zastosowaniu pochodnych (tlenek grafenu GO, zredukowany tlenek grafenu rGO) przekładają się z kolei na zmianę biodostępności oraz przepuszczalności przez poszczególne bariery biologiczne (przekraczania błony śluzowej, czy skóry, ale także przenikanie przez błonę komórkową do wnętrza komórki np. nowotworowej). Udowodniono także toksyczne działanie na kilka najbardziej rozpowszechnionych i zarazem chorobotwórczych bakterii jak gronkowiec złocisty (*Staphylococcus aureus*) czy pałeczka okrężnicy (*Escherichia coli*). Cechy te mają ogromne znaczenie w przypadku tworzenia preparatów aktywnych biologicznie. [Jaworski et al. 2021, Patelis et al. 2016].

Dokładna identyfikacja potrzeb przekłada się w konsekwencji na stworzenie „kompozytu", który będzie pozostawał w miejscu podania, czy też penetrował do głębszych warstw tkanek. Jedne z pierwszych badań nad wpływem pochodnych grafenu na linie komórkowe nowotworów złośliwych (w tym glejaka wielopostaciowego) wykonano na SGGW. Wykorzystano tu między innymi zdolność inicjacji ścieżki apoptozy, czyli „zaprogramowanej" śmierci komórki, w tym przypadku guza nowotworowego. Komórki te podczas procesu odróżnicowywania „deaktywują" tę ścieżkę, przez co rośnie potencjał rozrostowy tkanki

nowotworu. Powtórna inicjacja ścieżki apoptozy powoduje, ze masa guza zmniejsza się, poprzez obumieranie komórek złośliwych. [Jaworski et al. 2021]. Kolejne badania wykorzystywały zdolności adhezyjne grafenu i jego zdolność działania jako nośnika substancji cytotoksycznych w innych typach nowotworów. Na Gdańskim Uniwersytecie Medycznym potwierdzono skuteczność połączenia grafenu z 2-metoksyestradiolem na liniach komórkowych czerniaka (*melanoma malignum*) i mięsaka kostnopochodnego (*osteosarcoma*), a więc nowotworów, z którymi walka należy do najtrudniejszych [Kamm et al. 2022].

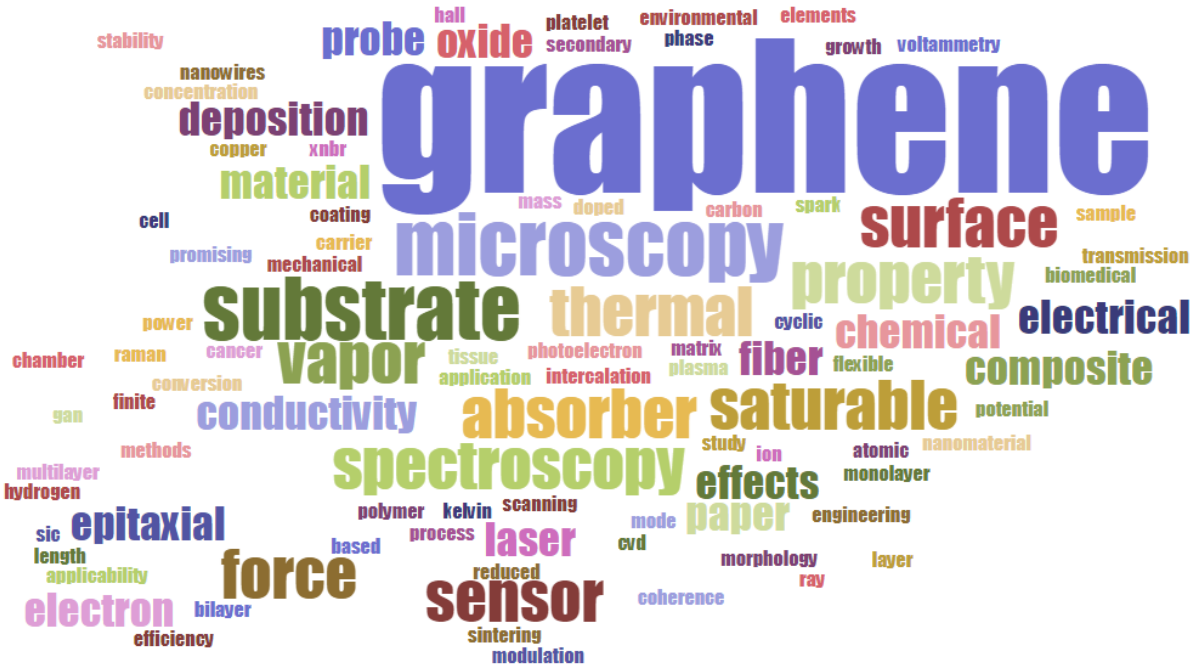
Kolejne specjalizacje medyczne to chociażby chirurgia naczyniowa, gdzie wykorzystano właściwości antyadhezyjne grafenu w protezach naczyniowych znacznie zmniejszając ryzyko choroby zakrzepowej i jej powikłań dla pacjentów. [Patelis et al. 2016].

Wyniki przeprowadzonych badań wskazują jednoznacznie na potencjał grafenu jako materiału, którego wykorzystanie w medycynie, a szczególnie w terapii przeciwnowotworowej sprawiają, że jednostki naukowe zaangażowane w badania nad tym materiałem zyskują nowe kompetencje strategiczne.

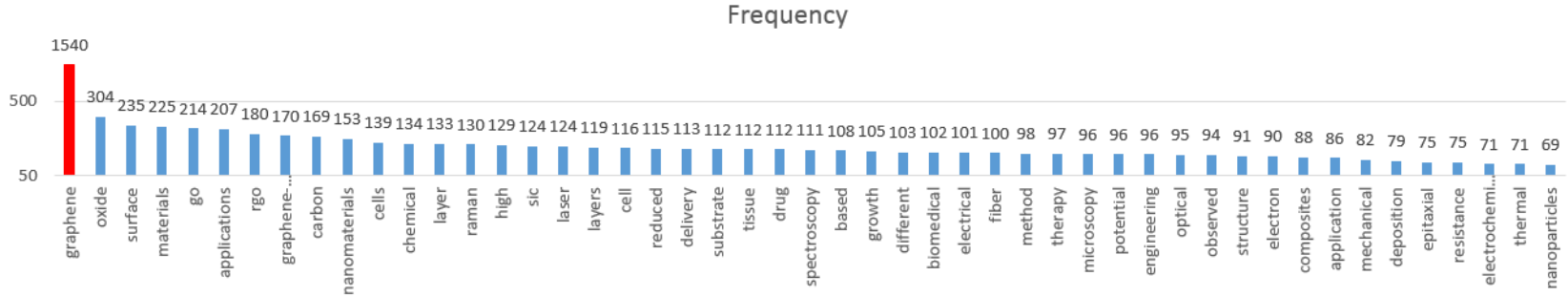
Każda ze wspomnianych specjalizacji, pomimo, że medycznych wiąże się ze współpracą z całymi zespołami badaczy z dziedzin niemedycznych. Przemysł farmaceutyczny, chemiczny, fizyka materiałów i inne są to specjalizacje, którym należy wskazać obszar, a może raczej wspólnie z nimi wypracować obszary, w których wiedza, doświadczenie i kompetencje mogą się uzupełniać w celu osiągnięcia sukcesu.

Na nowe, strategiczne kompetencje wskazuje analiza sentymentów związanych z grafenem i statystyczna analiza bigramów i trigramów wskazująca na nomenklaturę „medyczną" w kontekście wykorzystania grafenu w celach praktycznych. Ilustracją tych analiz są wykresy zamieszczone poniżej.

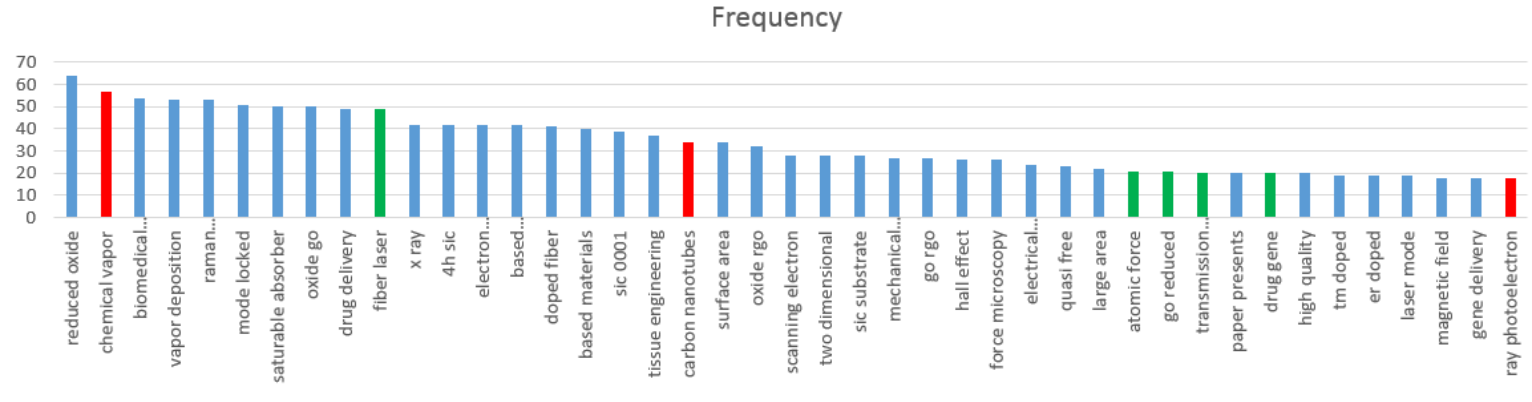
Rysunek 2. Chmura tagów ($\sqrt{n}$) związanych z określeniem „Graphene” w analizowanych źródłach literaturowych



Źródło: opracowanie własne na podstawie badań.


Frequency
1540
304
235 225 214 207 180 170 169 153 139 134 133 130 129 124 124 119 116 115 113 112 112 112 111 108 105 103 102 101 100 98 97 96 96 96 95 94 91 90 88 86 82 79 75 75 71 71 69
500
50
graphene
oxide
surface
materials
go
applications
rgo
graphene-...
carbon
nanomaterials
cells
chemical
layer
raman
high
sic
laser
layers
cell
reduced
delivery
substrate
tissue
drug
spectroscopy
based
growth
different
biomedical
electrical
fiber
method
therapy
microscopy
potential
engineering
optical
observed
structure
electron
composites
application
mechanical
deposition
epitaxial
resistance
electrochemi...
thermal
nanoparticles

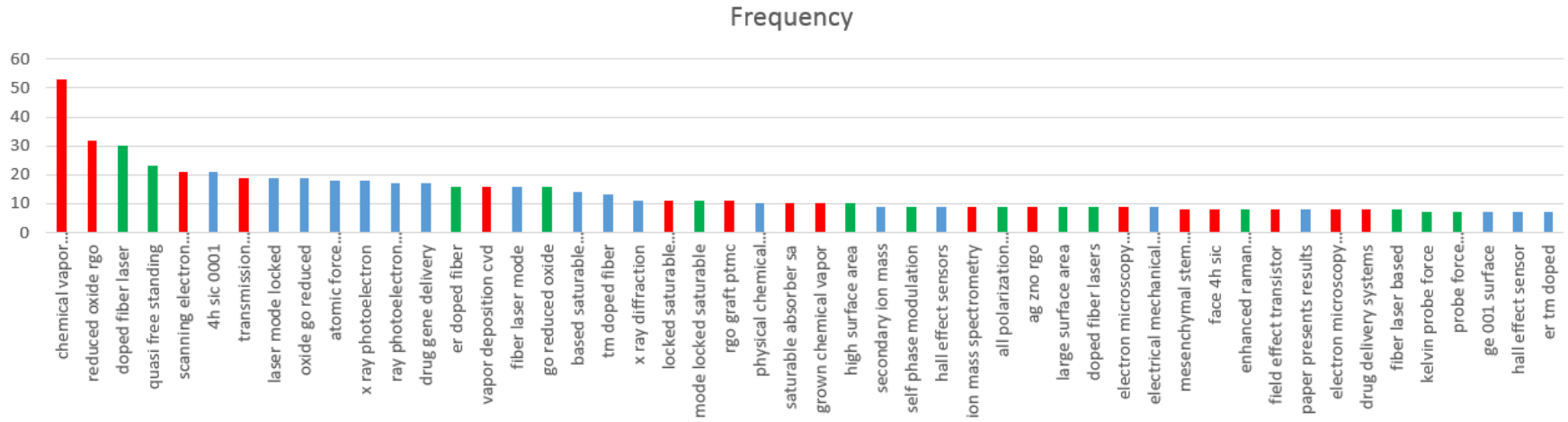
*

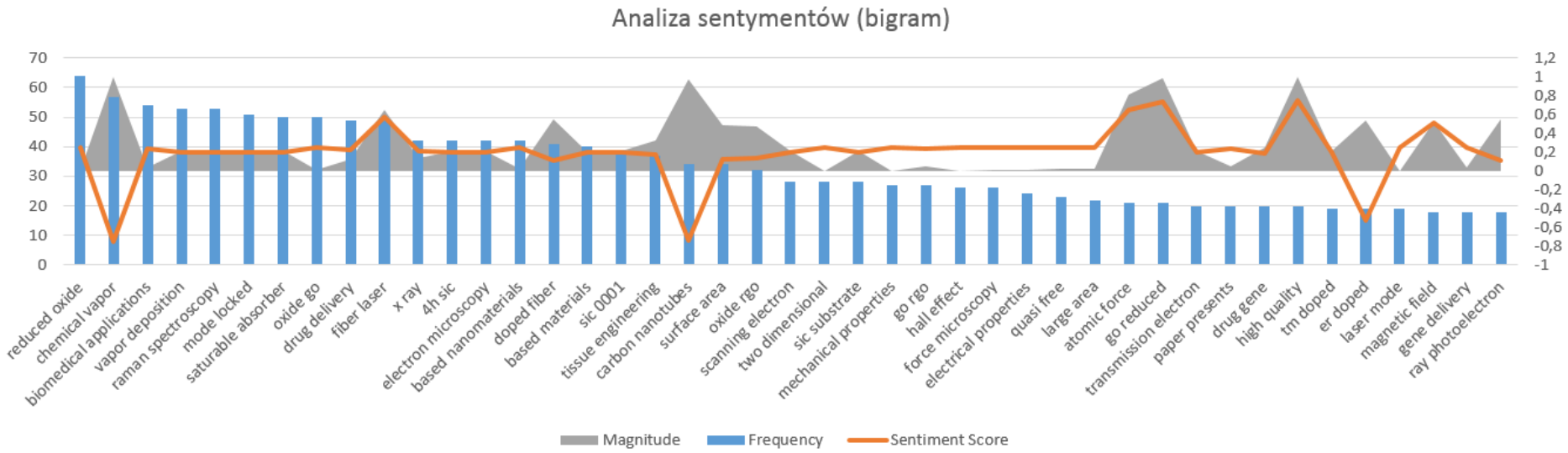

Frequency
70
60
50
40
30
20
10
0
reduced oxide
chemical vapor
biomedical...
vapor deposition
raman...
mode locked
saturable absorber
oxide go
drug delivery
fiber laser
x ray
4h sic
electron...
based...
doped fiber
based materials
sic 0001
tissue engineering
carbon nanotubes
surface area
oxide rgo
scanning electron
two dimensional
sic substrate
mechanical...
go rgo
hall effect
force microscopy
electrical...
quasi free
large area
atomic force
go reduced
transmission...
paper presents
drug gene
high quality
tm doped
er doped
laser mode
magnetic field
gene delivery
ray photoelectron

Wyniki przeprowadzonych badań, a szczególnie wyniki statystycznych analiz bigramów i trigramów wskazują na nowe, dotąd niezidentyfikowane kompetencje jednostek naukowych zaangażowanych w badania i produkcję grafenu oraz tworzenie technologii grafenowych w zakresie przede wszystkim badań podstawowych dotyczących technologii materiałowej.

Już sama analiza częstości dla bi- i trigramów wskazuje na występowanie słów i fraz silnie związanych z medycyną i technologiami stosowanymi w terapii. Uzupełnienie o analizę, w której połączono częstości bigramów i trigramów z analizą sentymentów wskazuje na potencjał grafenu. Częstości występowania takich trigramów, jak na przykład „Er droped fiber", „mode locked saturable", „drug delivery system", „tissue engineering" czy „probe force microsopy" mogą wskazywać na zastosowanie analizowanego materiału w technologiach związanych z medycyną i terapią. Dodatkowo, wziąwszy pod uwagę siłę i kierunek sentymentu, identyfikować można potencjalnie wzrostowe i rozwojowe technologie medyczne, w których grafen może znaleźć zastosowanie. Analizę sentymentów wraz z określeniem siły i kierunku oraz częstości występowania badanych fraz ukazują poniższe ilustracje.

Na potencjał zastosowania proponowanej przez nas technologii identyfikacji ukrytych kompetencji organizacyjnych wskazuje praktyka działalności badawczej jednego z instytutów Sieci Łukasiewicz, GUMed'u oraz INE PAN. Praktyczne zastosowanie wyników działalności naukowej prowadzonej w INE PAN i polegającej na wykorzystaniu narzędzi analiz big data wraz z analizą sentymentów dla wsparcia b. Instytutu Technologii Materiałów Elektronicznych w procesach komercjalizacji technologii tam opracowanej i wytwarzanej zakończyło się złożeniem wniosku patentowego. Przetestowanie i wykorzystanie narzędzi analizy rozproszonych, niestrukturyzowanych zbiorów danych w postaci tekstów publikacji naukowych i innych dokumentów o charakterze naukowym wraz z analizą otoczenia naukowego, także w zakresie innych, niż dominujące w b.ITME, dyscyplin naukowych wskazało na potencjał tych narzędzi dla podnoszenia zdolności do komercjalizacji wyników projektów badawczych i – przede wszystkim – identyfikacji nowych kompetencji naukowych. Analizy potencjału b. ITME wskazały na możliwości współpracy z instytucjami medycznymi. W efekcie, zidentyfikowany potencjał pozwolił na opracowanie metody syntezy oraz zastosowanie nowego biokompozytu Graph-2-ME w leczeniu czerniaka złośliwego i innych nowotworów złośliwych. Efektem tej współpracy jest wspólne zgłoszenie patentowe trzech instytutów (nr zgłoszenia: P.438737).


Frequency
60
50
40
30
20
10
0
chemical vapor...
reduced oxide rgo
doped fiber laser
quasi free standing
scanning electron...
4h sic 0001
transmission...
laser mode locked
oxide go reduced
atomic force...
x ray photoelectron
ray photoelectron...
drug gene delivery
er doped fiber
vapor deposition cvd
fiber laser mode
go reduced oxide
based saturable...
tm doped fiber
x ray diffraction
locked saturable...
mode locked saturable
rgo graft ptmc
physical chemical...
saturable absorber sa
grown chemical vapor
high surface area
secondary ion mass
self phase modulation
hall effect sensors
ion mass spectrometry
all polarization...
ag zno rgo
large surface area
doped fiber lasers
electron microscopy...
electrical mechanical...
mesenchymal stem...
face 4h sic
enhanced raman...
field effect transistor
paper presents results
electron microscopy...
drug delivery systems
fiber laser based
kelvin probe force
probe force...
ge 001 surface
hall effect sensor
er tm doped


Analiza sentymentów (bigram)
70
60
50
40
30
20
10
0
1,2
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
-1
reduced oxide
chemical vapor
biomedical applications
vapor deposition
raman spectroscopy
mode locked
saturable absorber
oxide go
drug delivery
fiber laser
x ray
4h sic
electron microscopy
based nanomaterials
doped fiber
based materials
sic 0001
tissue engineering
carbon nanotubes
surface area
oxide rgo
scanning electron
two dimensional
sic substrate
mechanical properties
go rgo
hall effect
force microscopy
electrical properties
quasi free
large area
atomic force
go reduced
transmission electron
paper presents
drug gene
high quality
tm doped
er doped
laser mode
magnetic field
gene delivery
ray photoelectron
Magnitude
Frequency
Sentiment Score

Podsumowanie

Wyniki badań przeprowadzonych metodami ilościowymi z zastosowaniem technik big data na artykułach naukowych w zakresie badań nad grafenem pokazują szerokie możliwości zastosowania tych metod w zarządzaniu i analizie fundamentalnej kluczowych kompetencji instytutów badawczych. Kompetencje te kształtują zarówno pozycję naukową jednostek badawczych, jak również wpływają na strategiczne możliwości rozwojowe i kondycję ekonomiczną instytutów. Zastosowane narzędzia pozwalają na identyfikację zasobów ekonomicznych jednostek naukowych, które w praktyce mogą okazać się kluczowe dla potencjału komercjalizacyjnego technologii w nich wytwarzanych.

Zakres i możliwości wykorzystania grafenu w medycynie; w chirurgii, transplantologii, neurobiologii, a szczególnie w terapii przeciwnowotworowej wskazują na pojawienie się kompetencji organizacyjnych w instytucjach, które dotąd niezwiązane były z medycyną. Przez pryzmat praktycznych zastosowań technologii materiałowych możliwa jest identyfikacja kluczowych kompetencji i zasobów dla potencjalnej zmiany strategicznej.

Przeprowadzone przez nas badania wskazują, że analizy z zastosowaniem metod big data są ważnym i użytecznym narzędziem w zakresie wsparcia zarządzania, ze szczególnym uwzględnieniem zarządzania informacją i wsparcia procesów zmiany strategicznej, komercjalizacji oraz zarządzania innowacjami.

Poszukiwanie wciąż nowych metod terapeutycznych w chemioterapii, m.in. w oparciu o nanocząstki jako nośniki leków przynosi nowe odkrycia. W ostatnich latach grafen i jego pochodne stały się względnie dostępnym materiałem organicznym o dobrze zdefiniowanej strukturze, stwarzającym obiecujące możliwości zastosowania w kontekście celowanej terapii przeciwnowotworowej. Jedną z szans na poprawę efektywności leczenia onkologicznego jest terapia celowana/chemioterapia regionalna, która jest precyzyjnie adresowana w miejsce tkanki nowotworowej, pozwalając tym samym na optymalizację dawki chemioterapeutyku oraz zmniejszenie toksyczności systemowej. Nanocząstki, w tym nanocząstki grafenu, stanowią obiecującą platformę nośną o unikalnych właściwościach biologicznych i biofizykochemicznych, która umożliwia regulowaną a przez to stosunkowo selektywną dystrybucję związanego z nią chemioterapeutyku zwiększając tym samym efektywność leczenia i minimalizując działania niepożądane.

Zastosowane narzędzia big data i identyfikacja nowych kompetencji pozwoliła nie tylko na złożenie zgłoszenia patentowego, ale także na wykrystalizowanie się nowego, interdyscyplinarnego zespołu badawczego.

Zgłoszenie patentowe obejmuje nowoopracowaną metodę syntezy hybrydy Graph-2-ME oraz biomedyczną ocenę jej potencjalnej efektywności przeciwnowotworowej. Graph-2-ME składa się z tlenku grafenu (GO) oraz jego zredukowanej formy (rGO) w rozmiarze nanopłatków oraz 2-metoksyestradiolu (2-ME). Po raz pierwszy, zsyntezowana została hybryda nanoformy grafenu z 2-metoksyestradiolem (2-ME) i oceniono jej potencjał antynowotworowy w modelu czerniaka złośliwego. Nowopowstały zespół badawczy odkrył, że cytotoksyczne działanie 2-ME, w odpowiednich farmakologicznie stężeniach, względem komórek kostniakomięsaka oraz czerniaka złośliwego, jest również związane z selektywną indukcją neuronalnej syntezy tlenku azotu (nNOS) i indukcją stresu nitro-oksydacyjnego. Zaproponowana przez powstały zespół badawczy innowacyjna hybryda GO-2-ME ma duże szanse na praktyczne zastosowanie w leczeniu nowotworów na całym świecie, zwłaszcza, gdzie możliwe jest miejscowe podanie cytostatyków, m.in. zaawansowany czerniak złośliwy skóry.

Zidentyfikowane metodami analiz big data obszary kompetencji organizacyjnych nie byłyby w praktyce dostrzegalne bez zastosowania zaawansowanych metod obróbki dużych zbiorów danych.

Bibliografia

[1] Ahmed, Pervaiz K., Mohammed Rafiq, and Noorizan Mat Saad. “Internal Marketing, Organizational Competencies, and Business Performance.” AMA Winter Educators’ Conference Proceedings 13 (January 2002): 500. https://search.ebscohost.com/login.aspx?direct=true&db=bsu&AN=9864227&lang=pl&site=ehost-live.

[2] Balbinot, Zandra, Julie Cristini Dias, and Rafael Borim-de-Souza. “Unique Organizational Competencies of Brazilian Technological Innovation Centers.” Journal of Technology Management & Innovation 7, no. 1 (March 2012): 1–16. doi:10.4067/S0718-27242012000100001.

[3] Blackler, Frank. “Knowledge, Knowledge Work and Organizations: An Overview and Interpretation.” Organization Studies 16, no. 6 (November 1995): 1020. doi:10.1177/017084069501600605.

[4] Cegielski, Casey G., and Farmer, L. Allison Jones. “Knowledge, Skills, and Abilities for Entry-Level Business Analytics Positions: A Multi-Method Study.” Decision Sciences Journal of Innovative Education 14, no. 1 (January 2016): 91–118. doi:10.1111/dsji.12086.

[5] Cetera, Wiesław, Włodzimierz Gogołek, Aleksander Żołnierski and Dariusz Jaruga. (2022). Potential for the Use of Large Unstructured Data Resources by Public Innovation Support Institutions. Journal of Big Data. 9. 10.1186/s40537-022-00610-6.

[6] Chaston, Ian. “How Interaction between Relationship and Entrepreneurial Marketing May Affect Organizational Competencies in Small Uk Manufacturing Firms.” Marketing Education Review 7, no. 3 (Fall 1997): 55–65. doi:10.1080/10528008.1997.11488607.

[7] Civelli, Franco. “Personal Competencies, Organizational Competencies, and Employability.” Industrial & Commercial Training 30, no. 2 (February 1998): 48. doi:10.1108/00197859810207652.

[8] Edgar, William B., and Chris A. Lockwood. “Organizational Competencies; Clarifying the Construct. (Cover Story).” Journal of Business Inquiry: Research, Education & Application 7, no. 1 (May 2008): 21–32. https://search.ebscohost.com/login.aspx?direct=true&db=bsu&AN=34433792&lang=pl&site=ehost-live.

[9] Escrig-Tena, Ana Belén, and Juan Carlos Bou-Llusar. “A Model for Evaluating Organizational Competencies: An Application in the Context of a Quality Management Initiative.” Decision Sciences 36, no. 2 (May 2005): 221–57. doi:10.1111/j.1540-5414.2005.00072.x.

[10] Fleury, Afonso, and Maria Teresza Fleury. “The Evolution of Strategies and Organizational Competencies in the Telecommunications Industry.” International Journal of Information Technology & Decision Making 2, no. 4 (December 2003): 577–96. doi:10.1142/S021962200300080X.

[11] Hewitt, Alan, and Eric Lesser. “Critical Organizational Competencies for a Globally Integrated World. (Cover Story).” Chief Learning Officer 6, no. 9 (September 2007): 50–53. https://search.ebscohost.com/login.aspx?direct=true&db=bsu&AN=26375084&lang=pl&site=ehost-live.

[12] Huang J et al. Multi-Arch-Structured All-Carbon Aerogels with Superelasticity and High Fatigue Resistance as Wearable Sensors. ACS Appl Mater Interfaces. 2020 Apr 8;12(14):16822-16830. doi: 10.1021/acsami.0c01794. Epub 2020 Mar 29.

[13] Jagiello J et al. Synthesis and Characterization of Graphene Oxide and Reduced Graphene Oxide Composites with Inorganic Nanoparticles for Biomedical Applications. Nanomaterials (Basel). 2020 Sep 15;10(9):1846. doi: 10.3390/nano10091846.

[14] Jaworski S et al. Comparison of the Toxicity of Pristine Graphene and Graphene Oxide, Using Four Biological Models. 2021. DOI:10.3390/ma14154250.

[15] Jaworski S et al. In vitro and in vivo effects of graphene oxide and reduced graphene oxide on glioblastoma. Int J Nanomedicine. 2015 Feb 25;10:1585-96. doi: 10.2147/IJN.S77591.

[16] Kamm a et al. The cytotoxic properties of graphene derivatives and 2-metoxyestradiol / graphene compound on human melanoma (A375) and osteosarcoma (143B) cell lines. 34th European Congress of Pathology: the art of next generation pathology, 3-7 September 2022 : abstracts ; DOI: 10.1007/s00428-022-03379-4.

[17] King, Adelaide Wilcox, Sally W. Fowler, and Carl P. Zeithaml. "Managing Organizational Competencies for Competitive Advantage: The Middle-Management Edge." Academy of Management Executive 15, no. 2 (May 2001): 95–106. doi:10.5465/AME.2001.4614966.

[18] Kotzbauer, Greg, and William B. Weeks. "Developing the Organizational Competencies Needed in Value-Based Payment Systems." Hfm (Healthcare Financial Management) 69, no. 7 (July 2015): 76–77. https://search.ebscohost.com/login.aspx?direct=true&db=bsu&AN=103738990&lang=pl&site=ehost-live.

[19] Kuzmanova, Mariana. "Creation of Organizational Competencies for Change." Economics & Business 22 (June 2012): 107–11. https://search.ebscohost.com/login.aspx?direct=true&db=bsu&AN=82157072&lang=pl&site=ehost-live.

[20] Mouritsen, J., A. Hansen, and Hansen, C. Ø. "Inter-Organizational Controls and Organizational Competencies: Episodes around Target Cost Management/Functional Analysis and Open Book Accounting." Management Accounting Research 12, no. 2 (June 2001): 221–44. doi:10.1006/mare.2001.0160.

[21] Paajanen, Petri, Pasi Porkka, Henri Paukku, and Hannu Vanharanta. "Development of Personal and Organizational Competencies in a Technology Company." Human Factors & Ergonomics in Manufacturing 19, no. 6 (November 2009): 568–81. doi:10.1002/hfm.20184.

[22] Patelis N et al. The Potential Role of Graphene in Developing the Next Generation of Endomaterials. Biomed Res Int. 2016;2016:3180954. doi: 10.1155/2016/3180954. Epub 2016 Nov 29. PMID: 28025640 PMCID: PMC5153502 DOI: 10.1155/2016/3180954.

[23] Prasad, Acklesh, and Peter Green. "Organizational Competencies and Dynamic Accounting Information System Capability: Impact on AIS Processes and Firm Performance." Journal of Information Systems 29, no. 3 (Fall 2015): 123–49. doi:10.2308/isys-51127.

[24] Sago, Brad. "Building Organizational Competencies for Competitive Advantage." Business Credit 105, no. 2 (February 2003): 16. https://search.ebscohost.com/login.aspx?direct=true&db=bsu&AN=9078151&lang=pl&site=ehost-live.

[25] Saliola, Federica, and Antonello Zanfei. "Multinational Firms, Global Value Chains and the Organization of Knowledge Transfer." Research Policy 38, no. 2 (March 2009): 369–81. doi:10.1016/j.respol.2008.11.003.

[26] Salmasi, Maryam Khalilzadeh, Alireza Talebpour, and Elaheh Homayounvala. "Identification and Classification of Organizational Level Competencies for BI Success." Journal of Intelligence Studies in Business 6, no. 2 (May 2016): 17–33. doi:10.37380/jisib.v6i2.170.

[27] Sparrow, Paul. "Organizational Competencies: A Valid Approach for the Future?" International Journal of Selection & Assessment 3, no. 3 (July 1995): 168–77. doi:10.1111/j.1468-2389.1995.tb00024.x.

[28] Stanton, Angela D'Auria, and Wilbur W. Stanton. "Skills Employers Seek in Analytics-Focused Hires: Implications for Business Schools." Summer Internet Proceedings 18, no. 2 (July 2016): 78–79. https://search.ebscohost.com/login.aspx?direct=true&db=bsu&AN=118491169&lang=pl&site=ehost-live.

[29] Subramanian, Ram, Kamalesh Kumar, and Karen Strandholm. "The Role of Organizational Competencies in the Market-Orientation-Performance Relationship." International Journal of Commerce & Management 19, no. 1 (January 2009): 7–26. doi:10.1108/10569210910939645.

[30] Tiruneh, Getaneh Gezahegne, and Aminah Robinson Fayek. "Hybrid GA-MANFIS Model for Organizational Competencies and Performance in Construction." Journal of Construction Engineering & Management 148, no. 4 (April 2022): 1–14. doi:10.1061/(ASCE)CO.1943-7862.0002250.

[31] Vveinhardt, Jolita, and Egle Stonkute. "Building Organizational Competencies through the Learning Process in Sme: The Benefits of Trainings Organised for Companies' Employees." International Multidisciplinary Scientific Conference on Social Sciences & Arts SGEM, January 2015, 113–20. doi:10.5593/sgemsocial2015/b12/s3.014.

[32] Yuxiao Ma et al. Ultralight Interconnected Graphene-Amorphous Carbon Hierarchical Foam with Mechanical Resiliency for High Sensitivity and Durable Strain Sensors. ACS Appl Mater Interfaces. 2017 Aug 16;9(32):27127-27134. doi: 10.1021/acsami.7b05636. Epub 2017 Aug 2.

[33] Żołnierski, Aleksander, Wiesław Cetera, Dariusz Jaruga, Jan Grzegorek and Grzegorz Sowula. (2022). Graphene and its Applications. Study on The Development Trends in Research and on the Implementation Potential using Big Data and Information Refining Methods. Journal of Nanotechnology Research. 04. 10.26502/jnr.2688-85210034.

Anna Jesionek

Świadomość społeczna rosnącej wulgaryzacji języka - analiza treści serwisów online z wykorzystaniem narzędzi big data

Wstęp

W ostatnich latach obserwowany jest gwałtowny rozwój technologii internetowych. Wraz ze wzrostem liczby użytkowników Internetu rośnie również ilość materiałów w nim zamieszczanych. Nie zawsze jednak są to treści, które wywierają pozytywny wpływ na człowieka. Kontrola treści generowanych każdego dnia przez miliardy osób stała się więc poważnym wyzwaniem dzisiejszych czasów [14]. Jednym z zagrożeń związanych z rozwojem Internetu jest postępująca wulgaryzacja przekazu. W niniejszym rozdziale zakres tego zjawiska zostanie zawężony jedynie do obserwacji języka obecnego w rozrywkowych materiałach dostępnych w sieci. W wielu badaniach naukowych potwierdzono, że wulgaryzacja języka jest zjawiskiem szkodliwym, w szczególności dla dzieci i młodzieży. Powoduje nie tylko ubożenie języka ojczystego, ale także problemy w edukacji, wyrażaniu emocji, agresję, izolację społeczną, depresję oraz zaburzenia w rozwoju psychoseksualnym [16].

W niniejszym rozdziale podjęto próbę zebrania i przeanalizowania najnowszych danych umożliwiających obserwację języka tekstów najpopularniejszych polskich utworów muzycznych docierających do dzieci i młodzieży poprzez serwis internetowy Spotify. Celem badania jest udowodnienie szkodliwego zjawiska wulgaryzacji przekazu w polskiej muzyce, co w konsekwencji może pozwolić na poszerzenie świadomości społeczeństwa w tym temacie oraz

ochronę wielu młodych osób przed zaburzeniami w rozwoju. Sformułowano następującą hipotezę badawczą: większość najpopularniejszych polskich utworów muzycznych w serwisie Spotify w latach 2017–2021 ma wulgarny tekst, a użycie wulgaryzmów w tekście utworu zwiększa jego popularność.

W badaniu wykorzystane zostały dwa rodzaje danych – rankingi najczęściej słuchanych utworów dostępne w serwisie muzycznym Spotify [4] oraz teksty piosenek pochodzące ze stron internetowych: www.genius.com [12] i www.tekstowo.pl [21]. Analizie poddano teksty 552 najpopularniejszych polskich utworów muzycznych słuchanych w latach 2017–2021 w Polsce. Dane zostały pobrane ze stron internetowych dzięki metodzie web scrapingu, służącej do wydobywania informacji z sieci w sposób zautomatyzowany. Narzędzia wykorzystane w badaniu to program Microsoft Excel oraz języki programowania: R i Python. Eksploracja tekstów przeprowadzona została zgodnie z metodyką opisaną przez Welbersa, Van Atteveldta i Benoit, której najważniejszymi elementami są: przygotowanie danych, analiza tekstów wykorzystująca metody statystyczne, słownikowe, nadzorowane uczenie maszynowe, technikę przyporządkowania części mowy wyrazom oraz zaawansowane metody przetwarzania języka naturalnego [25].

Szkodliwe treści w Internecie w świetle badań naukowych

Dynamiczny rozwój Internetu spowodował, że coraz większa część społeczeństwa ma do niego dostęp niemal każdego dnia, przez co w sieci publikowane są ogromne ilości treści. Nie zawsze są to jednak materiały wywierające pozytywny wpływ na społeczeństwo. Na przestrzeni lat przeprowadzonych zostało wiele badań dotyczących internetowych zagrożeń. Jiang, Scheuerman, Fiesler i Brubaker badali różnice w postrzeganiu szkodliwych treści w Internecie w zależności od kraju pochodzenia użytkownika. Analiza wyników ankiet przeprowadzonych w 8 krajach pozwoliła na jednoznaczne stwierdzenie, że jako najpoważniejsze niebezpieczeństwo w Internecie uznawane są treści kierowane do grup szczególnie podatnych na zagrożenia, czyli przede wszystkim do dzieci [14]. Przyczynę tego zjawiska badali Siddiqui i Zeeshan, którzy jako główny powód wskazali brak świadomości młodych osób na temat zagrożeń związanych z korzystaniem z Internetu. Dzieci będące jeszcze na etapie rozwoju i poznawania świata nie zawsze potrafią odróżnić treści wartościowe od szkodliwych. Jednak problem nie dotyczy tylko młodych osób, ponieważ naukowcy podkreślali, że dużym stopniu to

od rodziców zależy, w jaki sposób ich dzieci będą korzystały z Internetu i do jakich treści będą miały dostęp [20]. Rolę rodziców w budowaniu świadomości dzieci na temat zagrożeń w Internecie podkreślali też Arifin, Mokhtar i inni naukowcy. Uznali oni młode osoby za zbyt niedojrzałe i łatwowierne, by mogły zrozumieć wszystkie zagrożenia w sieci, więc kluczowe są wskazówki i wsparcie osób starszych. Dodatkowo, podkreślali też rolę promowania korzystania z nowych technologii wśród rodziców, gdyż umożliwi to zrozumienie i właściwe podejście do tematu zagrożeń w Internecie [1].

W niniejszym rozdziale zakres szkodliwych treści zostanie zawężony jedynie do języka obecnego w materiałach dostępnych przez Internet. Przez ostatnie kilkadziesiąt lat znacznie nasilił się problem wulgaryzacji języka. Wykorzystywanie wulgarnego słownictwa w codziennym życiu może mieć wiele negatywnych skutków. Przede wszystkim jest to ubożenie ojczystego języka, ale poważnym problem związanym z tym zjawiskiem są również zaburzenia w rozwoju człowieka, czego już mniej osób jest świadomych. Pankowska przedstawiła w swoim badaniu konsekwencje wulgaryzacji języka. Jako pierwszy problem wskazała brak chęci korzystania z treści naukowych lub informacyjnych przez społeczeństwo. Powoduje to trudności w edukacji, zanikanie umiejętności czytania ze zrozumieniem, w szczególności tekstów podręcznikowych lub publicystycznych. Powstają przez to kłopoty w codziennym życiu przy rozwijaniu zainteresowań lub sprawach urzędowych. Pankowska określiła konsekwencje wulgaryzacji języka nawet jako zaburzenia w myśleniu ze względu na trudności określaniu pojęć i brak umiejętności nazywania własnych uczuć i emocji przez ograniczony zasób słownictwa. Jednostki nadmiernie używające wulgarnych słów zwykle zastępują wszystkie swoje stany emocjonalne jednym wulgaryzmem, który określa większość negatywnych uczuć i jest możliwy do użycia w wielu sytuacjach [16].

Kolejnym ważnym problemem powodowanym przez zjawisko wulgaryzacji języka, który wskazała Pankowska, jest używanie słów dotyczących seksualności i nazywających narządy rozrodcze jako wyrazy potoczne. Często zauważane jest nawet ich wyśmiewanie lub używanie jako obraźliwe słowa. Zbyt wczesne obcowanie z takim językiem przez młode osoby może spowodować zaburzenia w ich rozwoju psychoseksualnym [16]. Następne zagrożenie wynikające z wulgaryzacji języka związane jest z używaniem wulgaryzmów w celu rozładowania emocji, a w szczególności gniewu. Bushman, Baumeister i Stack udowodnili, że agresywne działania dające upust złości, na przykład uderzenie w worek treningowy, wcale nie uspokajają,

tylko powodują jeszcze większą agresję w przyszłości. Oznacza to, że wyładowanie gniewu poprzez przeklinanie również zwiększa skłonność do zachowań agresywnych. Naukowcy uznali takie działania za zagrożenie dla zdrowia publicznego, pokoju i harmonii społecznej [3]. Dodatkowo, używanie wulgaryzmów może wywołać negatywne reakcje u innych, a przede wszystkim strach i wrogość. Osoba przeklinająca zwykle jest gorzej odbierana przez społeczeństwo i w dłuższej perspektywie naraża się na utratę statusu społecznego i zmniejszenie wsparcia emocjonalnego [23]. Robbins i inni naukowcy potwierdzili to w swoich badaniach, określając nawet, że osoba używająca wulgarnego słownictwa w pewien sposób odstrasza innych ludzi, przez co może zostać wyizolowana społecznie. W konsekwencji może prowadzić to do poczucia odrzucenia i depresji. Są to zjawiska coraz częściej obserwowane w dzisiejszych czasach, jednak niewiele osób poszukuje przyczyn w wulgarnym zachowaniu lub słownictwie [18].

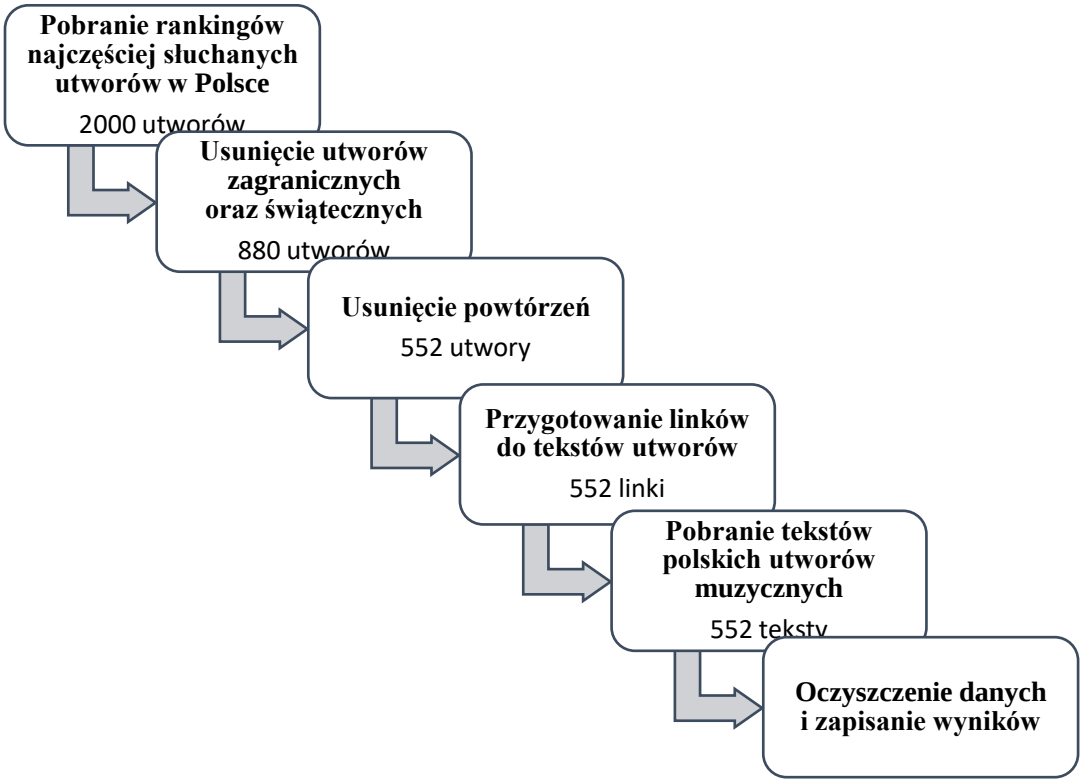
Przygotowanie zbioru danych

W Internecie jest bardzo wiele miejsc, w których można natknąć się na wulgarne treści. W niniejszym rozdziale uwaga zostanie skupiona na rozrywkowej jego części, a przede wszystkim na utworach muzycznych i ich tekstach. Przeanalizowane zostały teksty najpopularniejszych polskich piosenek, aby możliwe było stwierdzenie, czy są one wulgarne. Pierwszy rodzaj danych wykorzystany w badaniu to zestawienia najczęściej słuchanych utworów pochodzące z serwisu Spotify, który jest najpopularniejszym na świecie pod względem liczby subskrybentów serwisem strumieniowym oferujących dostęp do muzyki [13]. Udostępniane są w nim również rankingi najczęściej słuchanych utworów w podziale na kraje, dni oraz tygodnie. W celu przeprowadzenia badania zebrano zestawienia najpopularniejszych utworów muzycznych w Polsce z każdego pierwszego tygodnia stycznia i czerwca z pięciu lat – od 2017 do 2021 roku. Otrzymano w ten sposób 10 rankingów, każdy z nich zawierał po 200 utworów. Z zestawień odrzucono zagraniczne utwory, aby analizie poddać jedynie piosenki z polskim tekstem. Następnym krokiem było odrzucenie utworów świątecznych oraz takich, które występowały w rankingach kilkukrotnie. Zbiorcza tabela zawierała 552 wiersze i składała się z dwóch kolumn – wykonawcy oraz tytułu utworu.

W celu odnalezienia tekstów najpopularniejszych utworów muzycznych w Polsce wykorzystano serwis internetowy Genius, który uznawany jest za największy na świecie zbiór tekstów

piosenek oraz wiedzy muzycznej [12]. Zaobserwowano, że linki do tekstów pochodzące z tej strony zawierają podobną strukturę, która wygląda w następujący sposób: https://genius.com/Wykonawca-tytuł-lyrics, więc są możliwe do stworzenia na podstawie przygotowanej już tabeli z tytułami oraz wykonawcami utworów. Opracowano program komputerowy, który miał za zadanie połączyć zebrane już informacje o piosenkach w linki do ich tekstów. Finalnie otrzymano 545 linków do tekstów pochodzących ze strony internetowej www.genius.com [12] oraz siedem ze strony internetowej www.tekstowo.pl [21].

Rysunek 1. Schemat przygotowania zbioru danych do eksploracji tekstów



Źródło: Opracowanie własne na podstawie danych ze stron internetowych: www.charts.spotify.com, data odczytu: 10.07.2021; www.genius.com, www.tekstowo.pl, data odczytu: 20.07.2021.

Następnym etapem przygotowania danych było pobranie tekstów wszystkich zebranych utworów muzycznych na podstawie adresów internetowych opracowanych w poprzednim kroku. Postanowiono, że zostanie wykorzystana metoda web scrapingu, która pozwoli na wydobycie informacji ze stron internetowych w sposób zautomatyzowany. Przygotowano skrypt w języku programowania R, który bazował wcześniej opracowanym pliku zawierającym

linki do tekstów piosenek. Biblioteki, które wykorzystano w skrypcie to: *httr*, *dplyr*, *rvest* oraz *readxl*. Zainstalowano je za pomocą polecenia *install.packages()* i załadowano do programu poprzez użycie funkcji *library()*. Dzięki pakietowi *readxl* możliwe było zaimportowanie danych z pliku Excel, który zawierał linki do tekstów piosenek [8]. Biblioteka *httr* pozwoliła na korzystanie z narzędzi do pracy z adresami URL i HTTP [6], a *dplyr* – z obiektami typu *data frame* [5]. Pakiet *rvest* umożliwił wydobycie informacji ze stron internetowych [9]. Dzięki odpowiednio przygotowanym i oczyszczonym linkom pobrano teksty 552 utworów muzycznych. Usunięto z nich zawartość, która nie była tekstem piosenki, na przykład „[Refren]" lub „[Zwrotka 1]". Otrzymano w ten sposób dane gotowe do dalszej analizy, które zapisano w pliku CSV za pomocą polecenia *write.csv()*. Proces przygotowania zbioru danych podsumowano na Rysunku 1.

Metoda badań

Przygotowane teksty najpopularniejszych polskich utworów muzycznych możliwe były do analizy dzięki metodom eksploracji tekstu (z ang. text mining), nazywanej też inteligentną analizą tekstu. Termin ten opisuje proces wydobywania interesujących informacji z tekstu nieustrukturyzowanego za pomocą zestawu technik językowych, statystycznych oraz uczenia maszynowego (z ang. machine learning) [19]. Welbers, Van Atteveldt oraz Benoit przedstawili czynności, które należy wykonać, aby przeanalizować dane tekstowe w języku programowania R, który wybrali ze względu na możliwość wykonania w nim analizy i wizualizacji danych, a także dostępność wielu pakietów i narzędzi aktualizowanych na bieżąco przez programistów. Działania podzielili na trzy grupy: przygotowanie danych, analiza oraz zaawansowane tematy. Pięć elementów przygotowania danych do analizy określili jako: importowanie danych, operacje na ciągach znaków, przetwarzanie wstępne tekstu (w tym tokenizacja, usunięcie stopwords oraz stemming lub lematyzacja), stworzenie oraz filtrowanie macierzy DTM. Welbers, Van Atteveldt oraz Benoit jako drugą grupę operacji na danych tekstowych wskazali analizę. Odwołali się do podziału przedstawionego przez Boumansa i Trillinga, wyróżniając cztery metody analizy tekstu: słownikowe, statystyczne oraz nadzorowane i nienadzorowane uczenie maszynowe [2]. Naukowcy jako zaawansowane tematy wskazali dodatkowe metody przetwarzania języka naturalnego (z ang. Natural Language Processing, NLP) oparte na zewnętrznych modułach oprogramowania. Jedną z nich jest lematyzacja, dzięki której słowa sprowadzane są do ich form podstawowych. Podczas analizy tekstu często wykorzystuje się też

technikę przyporządkowania części mowy wyrazom (z ang. Part-of-speech tagging, POS), a także określania relacji między słowami, co umożliwia odkrycie nowych zależności w tekście [25].

W niniejszym rozdziale, dzięki metodom eksploracji tekstu przeanalizowana została treść najczęściej słuchanych polskich utworów muzycznych w latach 2017–2021 w serwisie Spotify. Zgodnie z metodyką opisaną przez Welbersa, Van Atteveldta oraz Benoit wykonano następujące czynności:

- Przygotowanie danych:
    - Import danych do środowiska R,
    - Operacje na ciągach znaków – oczyszczenie tekstów,
    - Przetwarzanie wstępne tekstu:
        - Tokenizacja,
        - Usunięcie stopwords,
        - Normalizacja,
    - Stworzenie macierzy terminów dokumentu (DTM),
    - Odfiltrowanie nieistotnych terminów z macierzy DTM,
- Analiza tekstów:
    - Metody statystyczne:
        - Analiza korelacji,
        - Obliczenie częstotliwości występowania słów,
    - Metody słownikowe:
        - Analiza sentymentu,
        - Występowanie wulgaryzmów tekstach,
    - Nadzorowane uczenie maszynowe:
        - Model logitowy badający najczęściej występujące słowa w tekstach,
- Zaawansowane działania:
    - Metody przetwarzania języka naturalnego (NLP):
        - Lematyzacja,
    - Technika przyporządkowania części mowy wyrazom (POS):
        - Analiza najczęściej występujących rzeczowników w tekstach polskich utworów muzycznych [25].

Eksplorację tekstów przeprowadzono z wykorzystaniem narzędzi: R, Python oraz Microsoft Excel. Język programowania R zastosowano do większości wymienionych wyżej czynności oprócz utworzenia modelu logitowego, który opracowano w języku programowania Python. Program Microsoft Excel posłużył do wykonania pomocniczych analiz i obliczeń.

W statystyce do analizy współzależności zjawisk wykorzystywane są modele, które umożliwiają określenie powiązań między zmienną zależną (objaśnianą) a zmiennymi niezależnymi (objaśniającymi) na podstawie odpowiedniej funkcji matematycznej – regresji [26]. Jednym z jej rodzajów jest regresja logistyczna, która opisuje zależność między zmiennymi ilościowymi lub jakościowymi oraz dychotomiczną zmienną zależną, czyli przyjmującą jedynie dwie wartości (0 i 1). Model logitowy umożliwia przewidywanie prawdopodobieństwa, że zmienna zależna przyjmie wartość 1, dlatego opisywany jest zgodnie z Równaniem 1. [17].

Równanie 1. Równanie opisujące model regresji logistycznej

$$P(Y = 1 \mid x_1, x_2, \ldots, x_k) = \frac{e^{(\alpha_0 + \sum_{i=1}^{k} \alpha_i x_i)}}{1 + e^{(\alpha_0 + \sum_{i=1}^{k} \alpha_i x_i)}}$$

gdzie:

α_i – współczynniki regresji (i = 0, …, k);

x_1, $x_{2, \ldots,}$ x_k – zmienne niezależne.

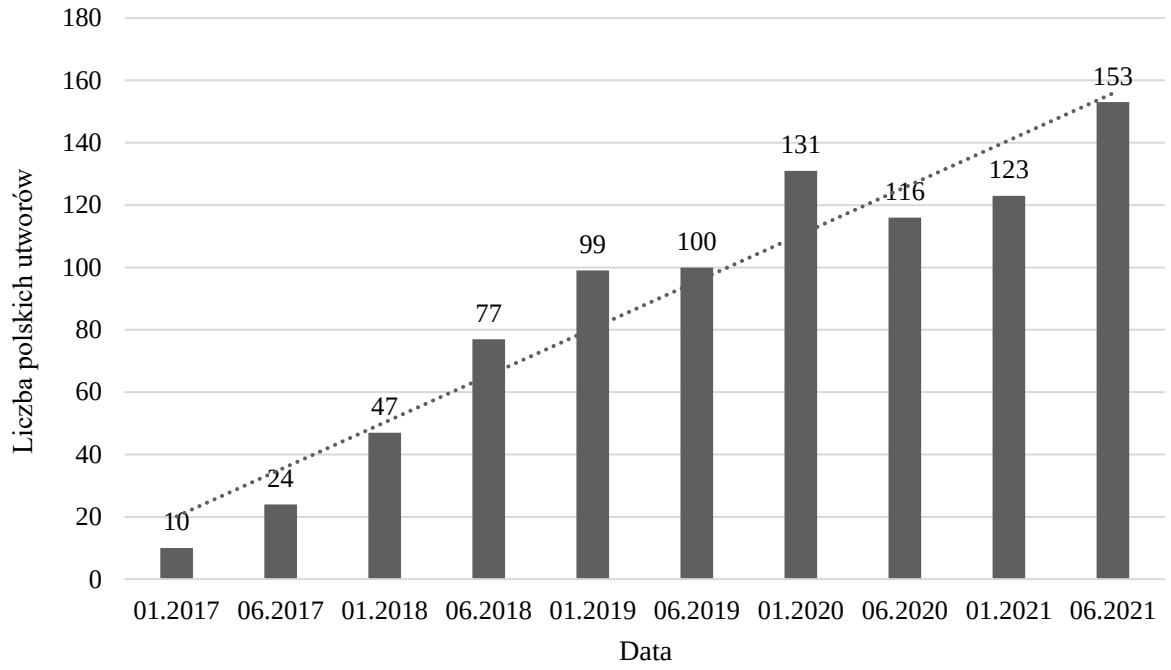
Źródło: M. Rabiej, Analizy statystyczne z programami Statistica i Excel, Gliwice 2018, s. 278.

Wyniki i wnioski

Badanie zostało podzielone na cztery części. Na początku przeprowadzono analizę rankingów najpopularniejszych utworów muzycznych w Polsce w latach 2017–2021. Następnie przeanalizowano teksty zebranych utworów. Trzecim etapem była analiza wulgaryzmów w nich występujących. Na zakończenie opracowano model logitowy badający najczęściej występujące słowa w tekstach oraz przedstawiono wnioski.

Na początku przeanalizowano zależności występujące w zebranych danych. Pierwszym etapem badania było przygotowanie rankingów najpopularniejszych utworów muzycznych w Polsce w latach 2017–2021. Występowanie utworów z polskim tekstem w rankingach przedstawiono na Wykresie 1.

Wykres 1. Występowanie utworów muzycznych z polskim tekstem w rankingach najpopularniejszych utworów w latach 2017–2021 w Polsce



Źródło: Opracowanie własne na podstawie danych ze strony internetowej: www.charts.spotify.com, data odczytu: 10.07.2021.

W latach 2017–2021 zaobserwowano gwałtowny wzrost liczby polskich utworów muzycznych w rankingach Spotify. Oznacza to, że polskie utwory muzyczne stają się coraz bardziej popularne w serwisie Spotify, a w ostatnich latach słuchane są nawet częściej niż utwory zagraniczne. W celu lepszego zrozumienia przedmiotu badania przeanalizowano również, którzy polscy artyści byli najczęściej słuchani w latach 2017–2021. W Tabeli 1. przedstawiono częstotliwość występowania utworów danego artysty w rankingach Spotify z lat 2017–2021.

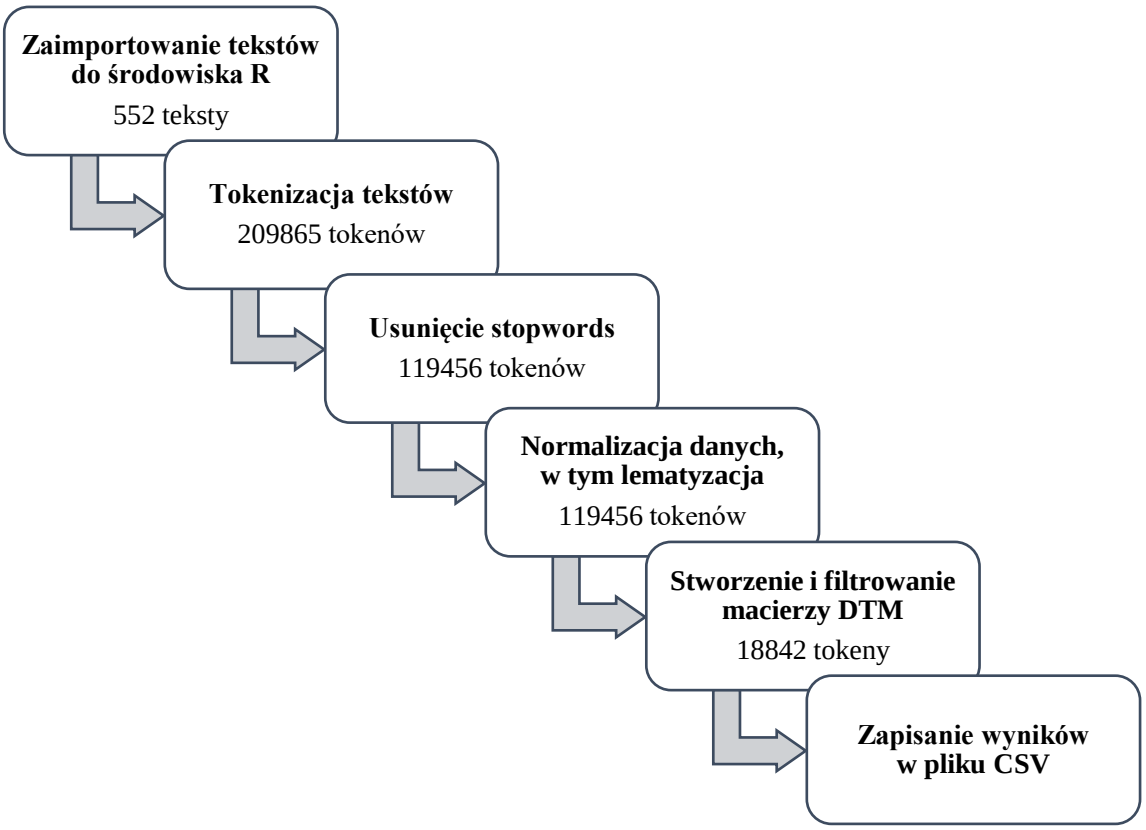
Tabela 1. Częstotliwość występowania utworów danego artysty w rankingach najpopularniejszych utworów muzycznych w latach 2017–2021

Lp.	Wykonawca	Liczba utworów w rankingu	Rok
1.	Bedoes	67	2017-2021
2.	Taco Hemingway	46	
3.	Quebonafide	44	
4.	Dawid Podsiadło	35	
5.	Taconafide	35	
6.	Tymek	28	
7.	White 2115	28	
8.	sanah	27	
9.	Sobel	25	
10.	Mata	24	

Źródło: Opracowanie własne na podstawie danych ze strony internetowej: www.charts.spotify.com, data odczytu: 10.07.2021.

W rankingach najpopularniejszych polskich utworów muzycznych z pięciu lat znalazło się najwięcej utworów Bedoesa. Warto dodać, że wykonawca ten jest raperem, tak samo jak większość artystów przedstawionych w Tabeli 1. i 2. Jest to ciekawe zjawisko, ponieważ platforma Spotify nie gromadzi jedynie utworów w tym gatunku [15]. Oznacza to, że dużą część twórców polskiej sceny muzycznej stanowią obecnie raperzy. Drugą częścią badania wykonanego w niniejszym rozdziale, była eksploracja tekstów polskich utworów muzycznych. Przeprowadzono ją zgodnie z metodyką opisaną przez Welbersa, Van Atteveldta i Benoit, której pierwszym etapem było przygotowanie danych [25]. Zaimportowano oczyszczone teksty 552 najpopularniejszych polskich utworów muzycznych w latach 2017–2021 do środowiska R i przeprowadzono na nich tokenizację, dzięki której teksty zostały podzielone na słowa. Wykorzystano w tym celu bibliotekę *tokenizers* oraz polecenie *tokenize_words()* [11]. Otrzymano w ten sposób 209865 tokenów. Następnie usunięto stopwords, czyli słowa o małym znaczeniu, które nie informowały o treści tekstów.

Rysunek 2. Schemat przygotowania danych do analizy tekstów



Źródło: Opracowanie własne na podstawie danych ze stron internetowych: www.genius.com, www.tekstowo.pl, data odczytu: 20.07.2021.

Zastosowano bibliotekę *stopwords*, a jako źródło danych wybrano słownik *stopwords-iso* zawierający 356 wyrazów [10]. Uzyskano w ten sposób 119456 tokenów, więc prawie 2 razy mniej niż przed wykonaniem tego działania. Kolejnym krokiem była normalizacja danych, w tym lematyzacja. Do jej przeprowadzenia opracowano specjalną funkcję o nazwie *lemmatize*, która bazowała na zewnętrznym słowniku. Wykorzystano w niej między innymi bibliotekę *hunspell* oraz polecenie *hunspell_stem*, dzięki któremu możliwe było zwrócenie form podstawowych zebranych słów [7]. Ostatnim etapem przygotowania danych do analizy tekstów było stworzenie macierzy terminów DTM i odfiltrowanie z niej nieistotnych słów. Finalnie otrzymano 18842 różne tokeny. Wyniki zapisano w pliku CSV za pomocą polecenia *write.csv()*, wykorzystując kodowanie UTF-8. Schemat przeprowadzonych działań przedstawiono na Rysunku 2.

Tabela 2. Zestawienie terminów oraz częstotliwości ich występowania w tekstach najpopularniejszych polskich utworów muzycznych przed i po lematyzacji

Lp.	Przed lematyzacją		Po lematyzacji	
	Termin	Częstotliwość występowania	Termin	Częstotliwość występowania
1.	chcę	836	chcieć	2040
2.	wiem	515	wiedzieć	944
3.	robię	346	mówić	840
4.	życie	342	robić	708
5.	k*rwa	336	móc	523
6.	świat	286	życie	514
7.	ciągle	283	noc	427
8.	będę	275	k*rwa	426
9.	jesteś	253	widzieć	411
10.	chyba	240	żyć	332
11.	mogę	228	lubić	330
12.	masz	224	czas	319
13.	czas	220	dać	318
14.	lubię	216	czuć	314
15.	wiesz	215	dzień	312
16.	czuję	215	mieć	311
17.	hej	211	ciągły	310
18.	miałem	209	tańczyć	300
19.	widzę	204	znać	293
20.	noc	200	powiedzieć	291

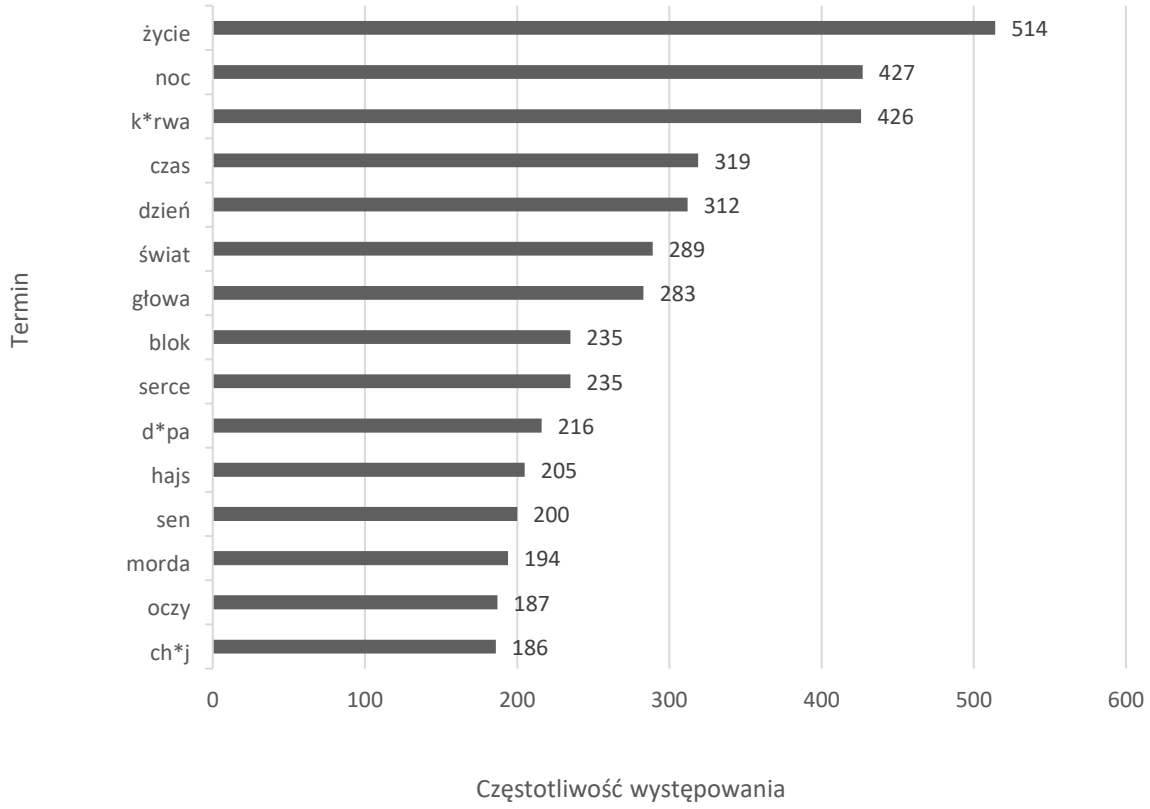
Źródło: Opracowanie własne na podstawie danych ze stron internetowych: www.genius.com, www.tekstowo.pl, data odczytu: 20.07.2021.

Jako drugi etap eksploracji danych przeprowadzono analizę tekstów. Z metod opisanych przez Boumansa i Trillinga wybrano trzy – statystyczne, słownikowe oraz nadzorowane uczenie maszynowe [2]. W celu zobrazowania tematyki tekstów policzono częstotliwość występowania każdego słowa we wszystkich zebranych tekstach piosenek. Uszeregowano je malejąco według tej wartości i wyniki dla pierwszych dwudziestu terminów przedstawiono w Tabeli 2. W Tabeli 2. przedstawiono dwa typy słów – przed i po lematyzacji, aby zobrazować cel sprowadzania słów do form podstawowych. W kolumnie z terminami przed lematyzacją występuje kilka odmian danego słowa, na przykład: *wiem*, *wiesz*. Są to wyrazy utworzone od słowa *wiedzieć*, które występuje w drugiej części tabeli. Częstotliwość jego występowania obrazuje sumę tej wartości dla wszystkich wyrazów pochodnych, dzięki czemu tematyka tekstu może zostać właściwie zobrazowana. Poprzez uszeregowanie słów według częstotliwości ich występowania w tekstach zaobserwowano, że wulgaryzm *k*rwa* obecny jest w Tabeli 2. na piątym miejscu wśród terminów przed lematyzacją oraz na ósmym miejscu w drugiej części zestawienia. Warto dodać, że różnych słów przed lematyzacją było 28260, a po niej 18857, więc pełne wersje tabeli częstotliwości występowania terminów składają się z tylu wierszy. Wulgaryzm występujący na tak wysokiej pozycji można uznać więc za niepokojące zjawisko, ponieważ oznacza to, że termin ten jest jednym z najczęściej występujących słów w tekstach najpopularniejszych polskich utworów muzycznych.

Proces lematyzacji określony został przez Welbersa, Van Atteveldta i Benoit jako zaawansowane tematy wśród metod przetwarzania języka naturalnego. Drugim z nich była technika przyporządkowania części mowy wyrazom [25]. Następną częścią badania w niniejszym rozdziale była więc analiza najczęściej występujących rzeczowników w tekstach najpopularniejszych polskich utworów muzycznych. Wyniki przedstawiono na Wykresie 2.

Najczęściej występujące rzeczowniki pozwoliły na określenie tematyki najpopularniejszych polskich utworów muzycznych. Słowa te wskazują, że badane teksty mogą mieć niewłaściwy przekaz, w szczególności dla młodych osób. Potwierdza to fakt, że wśród piętnastu najpopularniejszych rzeczowników wystąpiły trzy słowa wulgarne: *k*rwa*, *d*pa* oraz *ch*j*. Terminy te mają kilka znaczeń, mogą być na przykład używane jako wyzwiska lub określenia narządów płciowych.

Wykres 2. Zestawienie piętnastu najczęściej występujących rzeczowników w tekstach najpopularniejszych polskich utworów muzycznych w latach 2017–2021

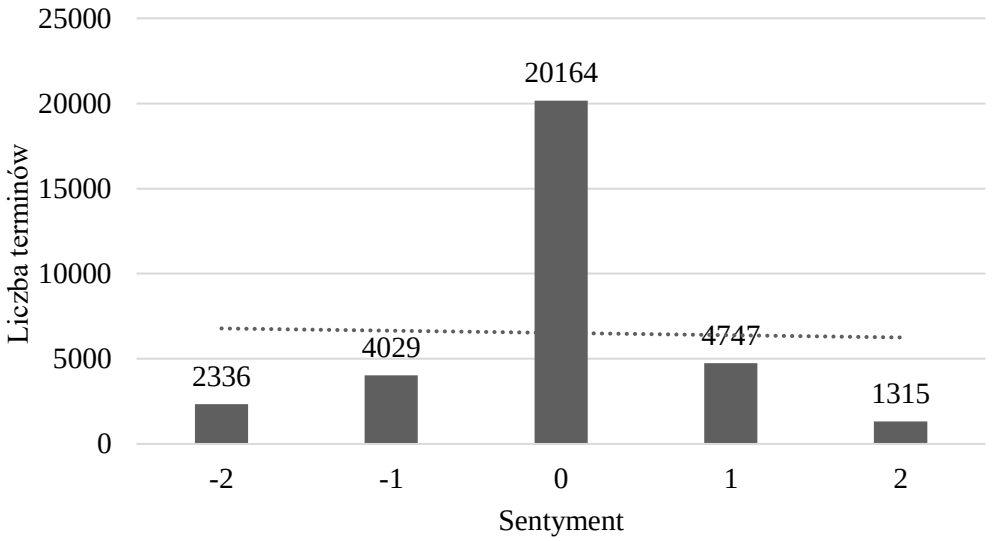


Źródło: Opracowanie własne na podstawie danych ze stron internetowych: www.genius.com, www.tekstowo.pl, data odczytu: 20.07.2021.

Następną techniką analizy danych tekstowych wykorzystaną w niniejszym badaniu była analiza sentymentu. Metoda ta polega na klasyfikacji tekstów ze względu na występujące w nich nacechowane emocjonalne słowa [22]. Do przyporządkowania wartości danym słowom wykorzystywane są słowniki wydźwięku. Wybrano jeden z nich, który klasyfikował terminy według pięciu kategorii:

- Bardzo negatywny (-2),
- Negatywny (-1),
- Neutralny (0),
- Pozytywny (1),
- Bardzo pozytywny (2) [24].

Wykres 3. Liczba terminów o danym sentymencie



Źródło: Opracowanie własne na podstawie danych ze stron internetowych: www.genius.com, www.tekstowo.pl, data odczytu: 20.07.2021.

Liczby w nawiasach oznaczają wartości przyporządkowane słowom należącym do danej kategorii. Przeprowadzono analizę sentymentu dla wszystkich zebranych słów z tekstów najpopularniejszych polskich utworów muzycznych w latach 2017–2021. Na Wykresie 3. przedstawiono liczbę terminów należących do każdej kategorii.

Najwięcej terminów miało wydźwięk neutralny, a najmniej – bardzo pozytywny. Suma wartości sentymentu wszystkich słów wynosiła -1324, więc w tekstach występowało więcej terminów negatywnie nacechowanych emocjonalnie niż pozytywnie. Analiza sentymentu wykazała, że wydźwięk tekstów najpopularniejszych polskich utworów muzycznych w latach 2017–2021 był negatywny.

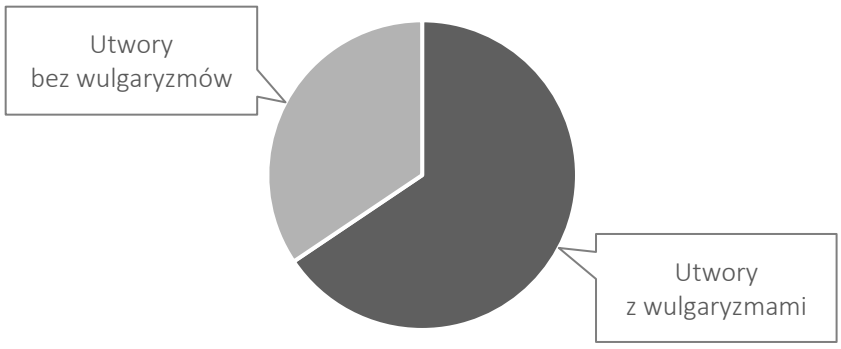
Wulgaryzmy w tekstach polskich utworów muzycznych

Trzecim etapem badania przeprowadzonego w niniejszym rozdziale była analiza wulgaryzmów występujących w tekstach polskich utworów muzycznych. Zaimportowano słownik wulgaryzmów, który zawierał 621 terminów. Policzono wulgaryzmy występujące w każdym z 552 badanych utworów. Okazało się, że 362 piosenki w swoim tekście zawierały co najmniej jeden wulgaryzm, co stanowiło 65,58% wszystkich zebranych utworów. Opisaną zależność

przedstawiono na wykresie kołowym (patrz Wykres 4.). Obserwacja ta stanowi potwierdzenie pierwszej części hipotezy badawczej, ponieważ wykazano, że większość najpopularniejszych polskich utworów muzycznych w serwisie Spotify w latach 2017–2021 ma wulgarny tekst.

Liczba wulgarnych utworów muzycznych, która w przybliżeniu stanowi 66% wszystkich najpopularniejszych piosenek, to bardzo wysoki wynik. Warto podkreślić również fakt, że są to utwory nieznane przez część społeczeństwa, która do słuchania muzyki wykorzystuje jedynie radio lub telewizję, ponieważ stosowana jest tam cenzura. Zgodnie z ustawą o radiofonii i telewizji nadawcy mają obowiązek przestrzegania poprawności językowej i przeciwdziałania wulgaryzacji [27]. Wyniki przeprowadzonej analizy wskazują na to, że większość najpopularniejszych polskich piosenek nie mogłaby być transmitowana przez wspomniane media.

Wykres 4. Występowanie wulgaryzmów w tekstach najpopularniejszych polskich utworów muzycznych w latach 2017–2021

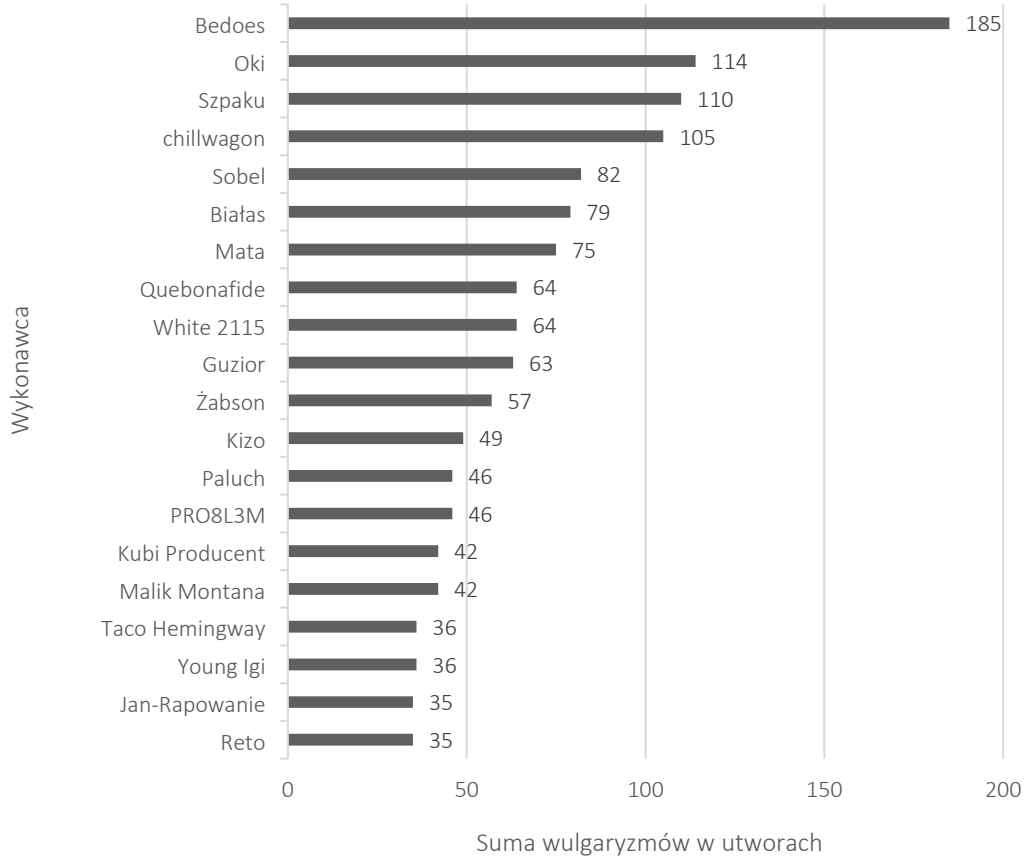


Źródło: Opracowanie własne na podstawie danych ze stron internetowych: www.genius.com, www.tekstowo.pl, data odczytu: 20.07.2021.

Są to jednak utwory bardzo popularne wśród części społeczeństwa korzystającej z internetowych serwisów muzycznych, ponieważ odtwarzane są setki tysięcy razy w tygodniu. Jest to niepokojące zjawisko, ponieważ w grupie tej jest dużo młodych osób, których rodzice mogą nie być świadomi, że tak szkodliwe treści docierają do ich dzieci.

W celu sprawdzenia, którzy artyści są najbardziej wulgarni, policzono również występowanie wulgaryzmów we wszystkich badanych utworach poszczególnych wykonawców. Wyniki przedstawiono na Wykresie 5.

Wykres 5. Zestawienie dwudziestu najbardziej wulgarnych artystów



Źródło: Opracowanie własne na podstawie danych ze stron internetowych: www.charts.spotify.com, data odczytu: 10.07.2021; www.genius.com, www.tekstowo.pl, data odczytu: 20.07.2021.

Najbardziej wulgarnym artystą okazał się Bedoes. Warto dodać, że wykonawca ten posiadał najwięcej utworów w rankingach najpopularniejszych polskich utworów muzycznych w latach 2017–2021, co przedstawione było w Tabeli 1. Jest to niepokojące zjawisko, ponieważ artyści często uznawani są za wzór do naśladowania, więc nie powinni przekazywać swoim fanom szkodliwych treści. Sobel oraz Oki również znaleźli się wysoko w rankingu. Są to artyści, których utwory osiągnęły rekordowe liczby odtworzeń w ciągu tygodnia – ponad milion [4]. Można więc uznać, że duża część społeczeństwa zna ich twórczość. Wulgaryzacja języka stanowi zagrożenie

dla wielu młodych osób oraz ma wiele negatywnych skutków, a na podstawie przeprowadzonej analizy można stwierdzić, że najpopularniejsi polscy artyści przyczyniają się do pogłębiania tego zjawiska.

Model logitowy badający słowa wpływające na popularność utworu muzycznego

Czwartym etapem badania przeprowadzonego w niniejszym rozdziale było opracowanie modelu logitowego badającego słowa wpływające na popularność utworu muzycznego. Budowę modelu rozpoczęto od przygotowania zbioru danych. 552 najpopularniejsze polskie utwory muzyczne z lat 2017–2021 podzielono na połowę według liczby odtworzeń w serwisie Spotify, tworząc zbalansowany zbiór danych. Utworom przyporządkowano odpowiednio wartości 1 i 0, tworząc w ten sposób zmienną zależną *hit*. Wartość zmiennej hit=0 oznaczała więc utwór mniej popularny, a wartość hit=1 – utwór znajdujący się wyżej w rankingu, czyli częściej słuchany. W celu wybrania zmiennych niezależnych (objaśniających) wyznaczono korelację ze zmienną objaśnianą dla 700 najczęściej występujących słów w tekstach polskich utworów muzycznych. Na podstawie wcześniej przeprowadzonej analizy zdecydowano, że jako zmienne niezależne zostaną wybrane rzeczowniki, ponieważ ze wszystkich części mowy najlepiej przedstawiają tematykę tekstów. Dwadzieścia rzeczowników o najwyższej wartości korelacji ze zmienną zależną zostało umieszczonych w modelu jako zmienne niezależne. Zakres przyjmowanych wartości oraz znaczenie zmiennych przedstawiono w Tabeli 3.

Między zmiennymi objaśniającymi nie zaobserwowano silnych korelacji, ponieważ wszystkie wartości współczynnika Pearsona były mniejsze niż 0,6. Można więc było przejść do kolejnego kroku, czyli estymacji modelu regresji logistycznej. W języku programowania Python wyestymowano model logitowy oparty na zmiennej zależnej *hit* oraz dwudziestu zmiennych niezależnych, które oznaczały słowa: *życie, k*rwa, głowa, d*pa, ch*j, rap, plan, kraj, imię, cash, usta, wstyd, piekło, tekst, wina, bluza, cel, gaz, fan, narkotyk*. Zaobserwowano, że wiele zmiennych jest nieistotnych w modelu przez wysoki poziom p-value. Poziom istotności przyjęty w badaniu to 0,05. Po przeprowadzeniu testów istotności usunięto z modelu osiem zmiennych, więc w modelu ostatecznym pozostało dwanaście istotnych statystycznie zmiennych oznaczających słowa: *życie, k*rwa, głowa, ch*j, plan, kraj, imię, cash, tekst, bluza, fan, narkotyk*. W Tabeli 4. przedstawiono wersję ostateczną modelu logitowego badającego słowa wpływające na popularność utworu muzycznego.

Tabela 3. Zmienne niezależne w modelu logitowym

Lp.	Termin	Nazwa zmiennej	Przyjmowane wartości	Typ zmiennej
1.	życie	X6	<0; 514>	Zmienna niezależna
2.	k*rwa	X8	<0; 426>	
3.	głowa	X22	<0; 283>	
4.	d*pa	X37	<0; 216>	
5.	ch*j	X49	<0; 186>	
6.	rap	X67	<0; 158>	
7.	plan	X100	<0; 123>	
8.	kraj	X298	<0; 57>	
9.	imię	X309	<0; 55>	
10.	cash	X358	<0; 50>	
11.	usta	X362	<0; 49>	
12.	wstyd	X376	<0; 47>	
13.	piekło	X378	<0; 47>	
14.	tekst	X396	<0; 45>	
15.	wina	X458	<0; 40>	
16.	bluza	X469	<0; 40>	
17.	cel	X478	<0; 39>	
18.	gaz	X516	<0; 37>	
19.	fan	X541	<0; 36>	
20.	narkotyk	X556	<0; 35>	

Źródło: Opracowanie własne na podstawie danych ze stron internetowych: www.genius.com, www.tekstowo.pl, data odczytu: 20.07.2021.

Największy wpływ na zmienną zależną miały zmienne oznaczające terminy: *imię, fan, życie, narkotyk*. Oznacza to, że użycie tych słów w tekście utworu może zwiększyć lub zmniejszyć prawdopodobieństwo, że dana piosenka będzie częściej słuchana, a więc znajdzie się wyżej w rankingu. Współczynnik przy zmiennych X6, X309, X358, X396, X469 i X541 jest ujemny, więc wykorzystanie słów *życie, imię, cash, tekst, bluza i fan*, a także ich odmian, wpływa negatywnie na popularność utworu. Natomiast dzięki użyciu słów *k*rwa, głowa, ch*j, plan, kraj i narkotyk* (zmienne X8, X22, X49, X100, X298 i X556) zwiększa się prawdopodobieństwo, że dana piosenka

będzie częściej odtwarzana, a więc znajdzie się też wyżej w rankingu najczęściej słuchanych utworów.

Tabela 4. Wersja ostateczna modelu logitowego badającego słowa wpływające na popularność utworu muzycznego

Lp.	Nazwa zmiennej	Współczynnik	Błąd standardowy	P-value
1.	X6	-7,5408	1,8739	0,0001
2.	X8	3,4989	1,4341	0,0147
3.	X22	2,7844	0,8753	0,0015
4.	X49	3,0536	1,5191	0,0444
5.	X100	4,7512	1,8291	0,0094
6.	X298	4,8698	2,3678	0,0397
7.	X309	-14,505	6,3593	0,0226
8.	X358	-2,8435	1,1668	0,0148
9.	X396	-5,9055	2,8784	0,0402
10.	X469	-5,4497	2,4691	0,0273
11.	X541	-9,3337	3,0619	0,0023
12.	X556	7,2032	3,2861	0,0284
Liczba obserwacji			552	
Pseudo R^2			0,128	
AIC			691,1412	
BIC			742,9038	

Źródło: Opracowanie własne na podstawie danych ze stron internetowych: www.genius.com, www.tekstowo.pl, data odczytu: 20.07.2021.

Zaskakujący jest fakt, że termin jednoznacznie wskazujący na tematykę związaną z narkotykami oraz dwa wulgaryzmy znalazły się wśród słów, które zwiększają popularność polskich utworów muzycznych. Oznacza to, że dla artystów korzystne jest używanie tych terminów w tekstach

piosenek, ponieważ dzięki temu przyciągają one większą liczbę odbiorców. Na podstawie wyników opracowanego modelu regresji logistycznej potwierdzono drugą część hipotezy, ponieważ udowodniono, że wykorzystanie wulgaryzmu w tekście utworu zwiększa jego popularność.

Podsumowanie

Jednym z problemów wynikających z bardzo szybkiego rozwoju Internetu jest brak wystarczającej kontroli treści z nim zamieszczanych. Powoduje to trudności w ocenie wiarygodności informacji, promowanie niewłaściwych zachowań, a także postępującą wulgaryzację przekazu. W niniejszym rozdziale przeanalizowane zostały teksty najpopularniejszych polskich utworów muzycznych w latach 2017–2021 w celu udowodnienia szkodliwego zjawiska wulgaryzacji języka. Utwory wybrano na podstawie rankingów pochodzących z serwisu muzycznego Spotify [4], natomiast teksty piosenek pobrano ze stron internetowych: www.genius.com [12] oraz www.tekstowo.pl [21]. Do badań wykorzystano program Microsoft Excel oraz języki programowania: R i Python. Opracowanie odpowiednich skryptów umożliwiło pobranie danych ze stron internetowych, ich przetworzenie, analizę oraz wizualizację wyników. Najważniejsze metody wykorzystane w badaniu to web scraping oraz eksploracja tekstów obejmująca: oczyszczenie, tokenizację i lematyzację tekstów, a także ich analizę, w tym badanie: częstotliwości występowania słów, wulgaryzmów, sentymentu, korelacji oraz model regresji logistycznej.

Na podstawie analizy przeprowadzonej w niniejszym rozdziale udowodniono szkodliwe zjawisko wulgaryzacji przekazu w polskiej muzyce popularnej w serwisie Spotify. Wykazano, że w przybliżeniu 66% badanych utworów zawiera w swoim tekście co najmniej jeden wulgaryzm, co stanowi potwierdzenie pierwszej części hipotezy, że większość najpopularniejszych polskich utworów muzycznych w serwisie Spotify w latach 2017–2021 jest wulgarna. Analiza najczęściej występujących słów oraz wydźwięku tekstów umożliwiła obserwację, że tematyka najczęściej słuchanych utworów jest szkodliwa, ponieważ ich teksty zawierają wiele słów wulgarnych, obraźliwych, negatywnie nacechowanych emocjonalnie, związanych z narkotykami lub wskazujących na tematykę seksualną. Wykazano również, że w ostatnich latach w Polsce popularnych staje się coraz więcej raperów, a najczęściej słuchani z nich używają w swoich utworach najwięcej wulgaryzmów. Można więc stwierdzić, że najpopularniejsi polscy artyści

przyczyniają się do pogłębiania zjawiska wulgaryzacji przekazu. Dodatkowo, wyniki opracowanego modelu logitowego wskazały, że wykorzystanie w tekście piosenki słów: *k*rwa*, *ch*j* oraz *narkotyk* zwiększa prawdopodobieństwo, że utwór będzie częściej słuchany, co stanowi potwierdzenie drugiej części hipotezy badawczej.

Wyniki otrzymane w niniejszym badaniu jednoznacznie wskazują, że najpopularniejsze polskie utwory muzyczne w serwisie internetowym Spotify nie powinny docierać do dzieci i młodzieży. Warto podkreślić, że duża część dostępnych tam utworów nie mogłaby być transmitowana w radiu lub telewizji ze względu na stosowaną cenzurę. Jest to niepokojące zjawisko, ponieważ wielu rodziców może nie być świadomych, jakich utworów słuchają ich dzieci, a to właśnie młode osoby są najbardziej narażone na negatywne konsekwencje wynikające z wulgaryzacji języka. Kluczowe jest więc uświadamianie społeczeństwa zarówno na temat szkodliwych materiałów dostępnych w Internecie, jak i zaburzeń w rozwoju, jakie może powodować postępujące zjawisko wulgaryzacji przekazu.

Bibliografia

[1] Arifin, U.A. Mokhtar, Z. Hood, S. Tiun, D. Indrayani Jambari, *Parental Awareness on Cyber Threats Using Social Media*, „Jurnal Komunikasi: Malaysian Journal of Communication" 2019, nr 35 (2), s. 485-498;

[2] J.W. Boumans, D. Trilling, *Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars*, „Digital Journalism" 2016, nr 4(1), s. 8-23;

[3] B.J. Bushman, R.F. Baumeister, A.D. Stack, *Catharsis, aggression, and persuasive influence: Self-fulfilling or self-defeating prophecies?* „Journal of Personality and Social Psychology" 1999, nr 76(3), s. 367-376;

[4] www.charts.spotify.com, [data odczytu: 10.07.2021];

[5] www.cran.r-project.org/web/packages/dplyr, [data odczytu: 25.07.2021];

[6] www.cran.r-project.org/web/packages/httr, [data odczytu: 25.07.2021];

[7] www.cran.r-project.org/web/packages/hunspell, [data odczytu: 25.07.2021];

[8] www.cran.r-project.org/web/packages/readxl, [data odczytu: 25.07.2021];

[9] www.cran.r-project.org/web/packages/rvest, [data odczytu: 25.07.2021];

[10] www.cran.r-project.org/web/packages/stopwords, [data odczytu: 25.07.2021];

[11] www.cran.r-project.org/web/packages/tokenizers, [data odczytu: 25.07.2021];

[12] www.genius.com, [data odczytu: 20.07.2021];

[13] M. Iqbal, *SpotifyRevenue and UsageStatistics* (2022), www.businessofapps.com/data/spotify-statistics/, [data odczytu: 10.05.2022];

[14] J.A. Jiang, M.K. Scheuerman, C. Fiesler, J.R. Brubaker, *Understanding international perceptions of the severity of harmful content online*, „PLoS ONE" 2021, nr 16 (8), s. 1-22;

[15] www.open.spotify.com, [data odczytu: 10.07.2021];

[16] D. Pankowska, A. Bieganowska-Skóra, *Wulgaryzacja języka jako kontekst socjalizacji dzieci i młodzieży*, „Przegląd Badań Edukacyjnych" 2018, nr 27 (2), s. 183-201;

[17] M. Rabiej, *Analizy statystyczne z programami Statistica i Excel*, Gliwice 2018, s. 278-283;

[18] M.L. Robbins, E.S. Focella, S. Kasle, A.M. López, K.L. Weihs, M.R. Mehl, *Naturalistically observed swearing, emotional support, and depressive symptoms in women coping with illness*, „Health Psychology" 2011, nr 30(6), s. 789-792;

[19] Z. Rybchak, O. Basystiuk, *Analysis of methods and means of text mining*, „Econtechmod. An International Quarterly Journal on Economics of Technology and Modelling Processes" 2017, nr 6(2), s. 73-78;

[20] Z. Siddiqui, N. Zeeshan, *A Survey on Cybersecurity Challenges and Awareness for Children of all Ages, [w:] International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, 2020, s. 131-136;

[21] www.tekstowo.pl, [data odczytu: 20.07.2021];

[22] K. Tomanek, *Analiza sentymentu - metoda analizy danych jakościowych. Przykład zastosowania oraz ewaluacja słownika RID i metody klasyfikacji Bayesa w analizie danych jakościowych*, „Przegląd Socjologii Jakościowej" 2014, nr 10(2), s. 118-136;

[23] A.J.J.M. Vingerhoets, L.M. Bylsma, C. De Vlam, *Swearing: A biopsychosocial perspective*, „Psihologijske teme" 2013, nr 22(2), s. 287-304;

[24] A. Wawer, *Polish sentiment dictionary* (2013), www.zil.ipipan.waw.pl/SlownikWydzwieku, [data odczytu: 21.07.2021];

[25] K. Welbers, W. Van Atteveldt, K. Benoit, *Text analysis in R*, „Communication Methods and Measures" 2017, nr 11(4), s. 245-265;

[26] A. Zimny, *Statystyka Opisowa. Materiały pomocnicze do ćwiczeń*, Konin 2010, s. 22-86;

[27] Dz.U. z 1993 roku Nr 7, poz. 34.

Małgorzata Stochmal

Using Critical Realism to Analyze Big Data: Ontic, Epistemic and Ethical Assumptions

Introduction to the issues

The expansion of human activity in the digital sphere is becoming more and more dynamic and there are no indications that it will lose its importance in the future. We started from accumulating kilobytes on floppy disks, to megabytes stored on hard drives, to terabytes stored on matrix disks, to perabytes stored in the cloud [13]. This development includes even larger volumes of data in terms of zettabytes, yotabytes, or exabytes. Large data archives are created to meet the increasing need to store this digital data. Many facts or reports have happened in the digital space that were previously reserved exclusively for real reality. As innovative technologies develop, algorithms, due to the scale of the analysis and the complexity of the decisions made, not only mediate in many relationships implemented in the digital world but are also able to define or modify the decision-making process [8, p.3]. It is also astonishing that the technological progress that has taken place from typical social studies to large database analyzes is astonishing. This progress is impressive.

The techniques currently used in social research are referred to as obsolete, which is associated with some kind of crisis in empirical sociology [16]. Is this really how this situation should be perceived? Definitely not, each type of research creates its own challenges and is conducted in a specific scientific paradigm with the use of adequate analytical techniques. The view of reality

changes due to the complexity and changeability of its accompanying conditions. We are currently witnessing the fourth paradigm; the individual stages are characterized as follows:

1. Experimental science. Empiricism; describing natural phenomena (pre-Renaissance).
2. Theoretical science. Modelling and generalization (pre-computers).
3. Computational science. Simulation of complex phenomena (pre-Big Data).
4. Exploratory science. Data-intensive; statistical exploration and data mining (Now) [15, p. 3].

Critical realism can also provide the foundation for proper research design, from theoretical assumptions to the construction of the initial model, to the selection of data inputted for analysis, and by how to analyse it and determine the meanings of the results obtained.

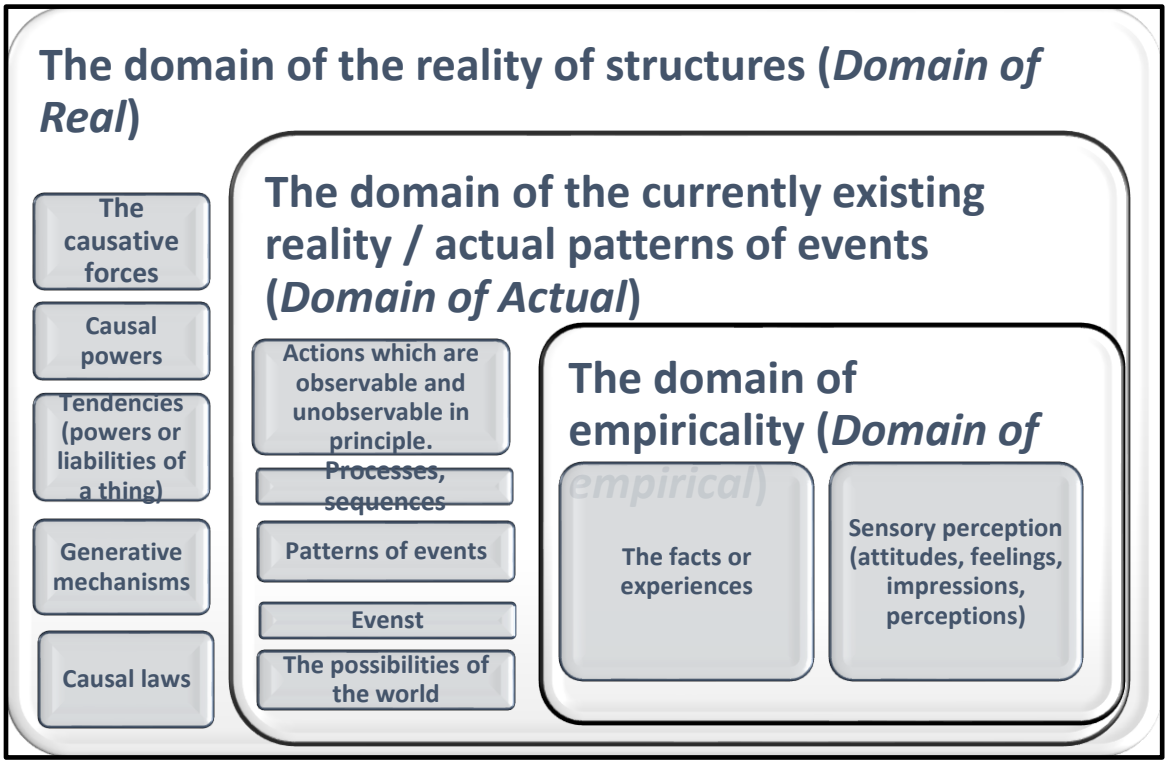
The stratified depth of ontology

In critical realism, priority is given to ontology, which ranks above epistemology. Ontology deals with the existence of beings and strong assumptions about their nature. The form of a strongly rooted realistic ontology restores the proper place not only to being, but also to its absence. The fact that we do not perceive a being at some point is not a proof of its non-existence, but it may result from its absence conditioned by the external context. Ontology as a theory of being emphasizes that all beings located in social, cultural, natural, biological or psychological reality exist independently of our any — complete or incomplete, fallible or true — awareness of them. The fact that they exist forces the tacit assumption that they operate in a strictly defined manner (e.g. the law of gravity, gravity, social inequalities, etc.).

The ontology of being has not been reduced to one dimension, but analyses its coherent structure in depth. Each being should be analysed in a deathly way, revealing more and more precise characteristics located on its individual layers.

A stratified ontology contains three elements placed in relation to each other: *Domain of Empirical*, d*omain of Actual* and d*omain of Real*) [5, p. 6-7]. A visualization of this concept is presented in Diagram 1.

Diagram 1. Ontological in-depth structures of reality in critical realism



Own study based on [6, s. 231].

Roy Bhaskar proposed an order that deepens us the structured and diversified ontology of reality. The proposed domains emerge from one another, they are relatively related to each other, although they are shifted in relation to each other in time. Shifting them in time at the stage of scientific analysis is significant because scientific research is usually not carried out in real time, but after some time from events related to the empirically studied experience. Bhaskar pointed to the "emergent power of the emergence" of being in three specific domains. The stratified ontology, although it naturally reflects the natural order of things, in a very insightful way shows us the complexity of the connection and coexistence (in the understanding of natural necessity) of individual results revealed in every area of reality. The concept of ontological depth of reality seems to be a structure consistent with the nature of the observed entities.

The empirical domain remains the innermost layer of being that emerges from experiencing the outside world. These experiences can be described as facts, sensations, practices or experiences resulting from sense perception. Everything that can be realized in action can also

be observed, and it is this observation that is of interest to the researcher in the empirical domain. Experiences that are part of scientific diagnosis should not constitute the basis for making generalizations, but constitute a plane for searching for the regularity of the events that cause them. The fact that you can experience something does not force you to experience it. There can be a world without the experience of certain events. An example would be rainfall. An event like rainfall is entirely real, but due to space-time conditions, some people will experience it and get wet, while others will not.

The domain of the currently existing reality / actual patterns of events (*Domain of Actual*). The name of this field seems to be significant, as it refers to the currently existing reality in which actual patterns of events are noticed. Bhaskar himself states that "(t)he intelligibility of scientific change (and criticism) and scientific education thus presupposes the ontological independence of the objects of experience from the objects of which they are the experiences" [3, p.21]. These events are located on the middle level of the stratified depth of reality, but we can easily emerge this analytical layer. These are all kinds of circumstances that determine the existence of resources, concepts, practices, and relationships that enable experience in the domain of empiricism.

So we naturally come to the presentation of the last layer - the domain of the reality of structures, which remains the area of the emergence of generative mechanisms and causal laws. Generative mechanisms generate phenomena that are noticeable in the event domain for real experiences. Bhaskar defines the functions of generative mechanisms as follows:

> "The real basis of causal laws are provided by the generative mechanisms of nature. Such generative mechanisms are, it is argued, nothing other than the ways of acting of things. And causal laws must be analysed as their tendencies. Tendencies may be regarded as powers or liabilities of a thing which may be exercised without being manifest in any particular outcome" [3, p. 3].

In the domain of the reality of structures the causal forces are revealed that activate the powers of generative mechanisms that make the subject, under the influence of current events, behave in the empirical domain in one way or another. Thus, generative mechanisms have the potential to generate events, and these in turn have an impact on their disclosure in the empirical domain

[17, p. 797] through perceived experience. Mechanisms generate streams of events that can satisfy the postulate of causal laws that generalize certain regularities that occur in specific circumstances.

While maintaining the rigour of the ontological recognition of entities embedded in social reality, such a bottom-up and iterative process of determining their form should be carried out, however, these findings should not be burdened with the subjectivism of the man who formulates these truths. The ontological in-depth structures of reality in critical realism can be defined as its peculiar ontological morphology.

Epistemic meanders

The area of the emergence of scientific laws is epistemology understood as the area of generating or producing scientific knowledge. The basic assumption of critical realism is epistemic relativism, which expresses a strong assumption of the natural fallibility of man in terms of the truths he formulates.

> "Epistemic realism means that all our claims are socially and historically conditioned. Our judgments are determined by circumstances, by what we know at the time and by binding criteria of judgment. For this reason, among other things, our judgments are always error prone. Epistemic relativism then means that each of us is in a situation from which we see the world in a slightly different way. Our experiences of the world are different" [1, p. 55].

This fallibility results from the spatio-temporal context of the production of knowledge, so it is generated at a specific historical moment, and in the course of ongoing changes, it is updated, distorted, or falsified. Social production of knowledge carries many burdens that can distort the truth inadvertently or unintentionally. Bhaskar defines the creation of science as "a process-in-motion", meaning that "knowledge must be viewed as the produced means of production and science as a constant social activity in a continuous process of transformation." Moreover, knowledge formulated by researchers is always mediated by concepts, language, history, or social constructs that emerge in a given context. The characteristic circumstances of the production of knowledge are the conditions of the social system, which is perceived as an open system, and thus constantly subject to changes in its structure.

In addition to epistemic relativism and the awareness of ontological depth, one should take into account the rationality of judgments. The rationality of judgments is related to the ability to evaluate competitively existing theoretical structures explaining reality to us and to choose the one that does so to the fullest extent. Scientists present the arguments of their evidence and reach a consensus in the course of rational discussion. The position that most accurately explains the given phenomenon of cognition wins. It should be realized that the production of knowledge is always mediated in some way by the researcher, his perception of reality, and his knowledge.

Bhaskar distinguishes between two key dimensions of knowledge: transitive and intransitive. Objects of transitive knowledge come first and serve as means of production to define intransitive knowledge. The transitive dimension includes objects of knowledge that are "raw materials of science - artificial creations shaped as objects of knowledge by the science in force at a given time". They are changeable due to the awareness of being critical in the sense of revealing breakthrough knowledge in the achievements of mankind. “Scientists try to discover the reasons for things and events, patterns and processes, sequences and structures. To understand how they do so one needs both a concept of the transitive process of knowledge-production and a concept of the intransitive objects of the knowledge they produce: the real mechanisms that generate the actual phenomena of the world, including as a special case our perceptions of them” [3, p. 52].

Objects of transitory knowledge connect the adopted “facts and theories, paradigms and models, methods and techniques of inquiry available to a particular scientific school or worker [6, p. 237]. This is the natural sequence of things, before any content is included in the field of science, it reveals itself in social spaces and constitutes the basis for formulated generalizations. “In this way social products, antecedently established knowledges capable of functioning as the transitive objects of new knowledges, are used to explore the unknown (but knowable) intransitive structure of the world” [6, p. 239]. Knowledge in the transitive dimension differs from knowledge embedded in the intransitive dimension.

Defining intransitive objects of knowledge, Bhaskar states that they are “the intransitive objects of knowledge are in general invariant to our knowledge of them: they are the real things and structures, mechanisms and processes, events and possibilities of the world; and for the most part they are quite independent of us. They are not unknowable, because as a matter of fact

quite a bit is known about them [3, p.12]. This citation is crucial in the epistemic realm. First, the language used to describe the objects of intransitive knowledge that a person learns is relatively unchanging, it can be said that they are truths ultimately expressed in the context of a given paradigm (an example of quantum physics and discoveries nominated for the Nobel Prize. It is knowledge that extends and deepens classical physics). Second, these truths are expressed in terms of causal laws. The intransitive dimension of science allows us to have a coherent understanding of reality. The law of gravity belongs to the realm of intransitive knowledge, knowing it and realizing that falling objects is an observable event, no reasonable person jumps out of the window from the tenth floor.

Formulated in terms of true knowledge or scientific laws, the findings relate to the operation of generative mechanisms. Bhaskar himself puts it this way: "The goal of science, however, is to generate knowledge of the mechanisms of creating phenomena in nature that come together to generate a real and continuous variation of world phenomena" [6, p. 231]. The mechanisms revealed in the course of in-depth reflection function independently of what people do in the world. Adopting a strong assumption about the existence of beings independent of people is a transcendental condition conducive to the development of science.

In the context of critical realism, causal laws operate even when they are not experienced as a result of factual events. They are of an intransitive nature. "By saying that the objects of discovery and scientific research are 'intransitive' I mean the indication that they exist independently of any human activity; and by saying that they are 'structured' I mean that they are separate from the patterns of events that occur" [6, p. 249].

The goal of epistemology is to identify the ways in which the world works, and so to identify hidden forces - processes or mechanisms - that produce observable effects or events in the empirical domain. Critical realism in the layer of assumptions radically separates ontology from epistemology, but their analytical interdependence and mutual co-constitution should be recognized [14, p. 358]. Epistemic complexities also enable knowledge to be formulated independently of its ontological experience. The ontological and epistemic assumptions are necessary to define social reality in order for scientific knowledge to be possible.

The emergence of generative mechanisms and their causal forces takes place in the course of the retroductive logic of discovery along with abductive inference. Bhaskar defines these issues as follows:

> "Abduction involves redescription or recontextualization, most usually (in CR research) in terms of a characteristic causal mechanism or process which serves to explain it. Retroduction involves imagining a model of a mechanism, which, if it were real, would account for the phenomenon in question. (These two can often shade into each other: there is only a relative difference between them.)" [4, p. VII].

The abductive logic of inference consists in a permutation, in other words, a reconfiguration of premises that lead us to conclusions. Abductive reasoning overcomes the shortcomings resulting from inductive or deductive reasoning. Abduction is a way to establish an approach to a problem that is logically justified due to the existing knowledge about the analyzed phenomenon. In turn, the retroductive logic of model discovery consists in imagining the necessary conditions for the occurrence of its individual elements. "A thought operation involving a reconstruction of the basic conditions for anything to be what it is, or, to put it differently, it is by reasoning we can obtain knowledge of what properties are required for a phenomenon to exist. Transfactual or transcendental argument is a form of retroduction implying that one seeks these qualities beyond what is immediately given" [10, p. 206].

Ethical assumption

The metatheory of critical realism allows us to consider ethical questions from the perspective of moral realism. Bhaskar formulates the position of moral realism in the Dialectical Critical Realism stage, where the pursuit of aletic truth remains a value, as does the pursuit of freedom or justice. „Bhaskar's argument for the universality of morality is a component of his dialectical critical realist ethics; this is a moral realist and ethical naturalist position that seeks to ground moral theory in an understanding of reality" [2, p. 30]. This ontologically transcendental ethical dimension exists regardless of whether individual values are recognized by people in their daily agency. The existence of a social reality, and therefore ontologically real, which is an area of subjective agency for people [11, p. 274] make this perpetration immanently embedded in the moral agency. Morality saturates with its being the real reality, regardless of the sources of this message, and activates the powers of moral provenance that make changes in the social area. Morality is the basis of the ontoaxiological agency of every subject.

Morality is embedded in the intransitive dimension of knowledge. As Stive Ash points out: “This can be understood as stating that moralities are transitive, but they have an intransitive object – intrinsic value [2, p. 42]. Thus, the dogma of Weber's axiological neutrality was questioned. Values can resonate in the subjective agency of people and constitute not only a declaration of what should be done but constitute the transformative force of their agency.

Critical and realistic data analysis in the area of Big Data

The Big Data environment is a collective term expressing large volumes of data, along with the possibilities of their storage and processing, visualization and ways of formulating conclusions about them. There is a strict relational relationship between the elements of the *Big Data* environment. In the literature on the subject, it is referred to as a socio-technological phenomenon [9, p. 663]. The first context is all kinds of communities, collectives or groups, while the second is the technological context.

Technologies included in the group of Big Data facilitate taking actions using the data-driven approach. Data-driven is a management approach based on data, i.e. a constant response to the results provided by data in real time and adapting practical actions to them [18, p. 72]. The resulting "rapid" data streams, known as Big Data, are a very good material for research, regardless of whether they pursue a business, scientific or any other goal.

When carrying out research in the Big Data environment, it is worth bearing in mind the abundance of benefits resulting from the use of a critical-realistic approach. At the ontological level, we deal with data nested in a relatively broad social context. This data is extremely complex. Before we move on to contextualizing them, it is important to properly recognize their essence. Understanding data begins with disclosing the sources of origin and the properties assigned to them. The emerged entities are defined in natural language [20, p. 54]. This should be done by deepening the reflection on their identity in three separate layers of reality. One should move from the level of data (*Domain of Empirical*), through the domain of reality (*Domain of Actual*) and end the search by revealing the causal mechanisms generating general regularities (*Domain of Real*). We are dealing here with the emergence of the identity of ontological beings and the opportunity to grasp them connected with the epistemological dimension [19, p. 100 and next].

The epistemic dimension, second only to the ontic dimension, is important due to the scientific progress already made in our interest. When undertaking activities aimed at disclosing new knowledge, we are obliged to analyse the achievements so far. A thorough analysis of the scientific achievements determines for us the intransitive dimension of knowledge, which is used to produce new knowledge, or to modify or improve the existing knowledge. The researcher should fully understand the issues in his area of exploration. It is worth paying attention to the foreign achievements. In the era of globalization, access to knowledge in every field is "just a mouse clicks away".

An example is the issue of fear, willingly undertaken by researchers in the field of Big Data, especially taking into account such techniques as sentiment analysis. From an ontological point of view, any existential uncertainty plays a role in creating social change. These changes can take place in a positive, negative and neutral dimension. Elemer Hankiss claims that existential security, i.e. the sense of, inter alia, fear of threats was the main factor of civilization changes. On the one hand, this fear raised the structural dimension and embedded in institutions responsible for ensuring broadly understood security. Secondly, with the protective sphere of symbols: myths and religions, values and belief systems, ideas and scientific theories, moral and practical rules of behaviour, and a wide range of everyday rituals and trivialities [12, p. 1-2]. Fear is undoubtedly an emotion that resonates with people's actions or causes them to be abandoned. These activities may be part of the generalized three main lines of action, although in reality there may be more of them[24]. Positive actions to overcome fear. Not taking action, delegating responsibility for creating safe living spaces to designated institutions. And neutral actions, showing people's reluctance to take any action.

An interesting characteristic of big data analysis is the involvement of representatives of various scientific disciplines and professions in the implementation of such projects. Members of these teams must develop a range of terms that are equally understood. Another important characteristic is the fact that all data on an observable phenomenon is collected, hence it is possible to observe them and identify patterns that characterize them.

> „The data are not subject to every ontological framing possible, or every form of data-mining technique in the hope that they reveal some hidden truth. Rather,

[24] Interesting doctoral dissertation describing the culture of fear towards local threats [7].

> theoretically informed decisions are made as to how best to tackle a data set such that it will reveal information which will be of potential interest and is worthy of further research" [15, p.6].

When looking for generative mechanisms, one should define their components and the strength of interaction between them. Then the identity of the trends revealed from the data can be identified. It seems important to consider many competing causal forces and powers and to select the most appropriate based on the logic of retroduction.

Critical realism formed ontic and epistemic assumptions, which with time were supplemented with axiomatic assumptions or those relating to the transcendental and even theological realm of reality. The problem of research in the digital sphere is maintaining ethical standards, for example those that are already a standard for the reality actually experienced. Among the ethical problems we can find issues related to the anonymization of data, the method of obtaining it, tracking the surveyed people without their informed consent, etc. Merely determining the potential and actual impact of an ethical algorithm is difficult for a number of reasons. Identifying the impact of human subjectivity in an algorithm design and configuration often requires the study of long-term, multi-user development processes [8, p. 2].

Final conclusions

People involved in large database analytics can successfully embed their analytical projects in the mainstream of critical realism. It is a perspective that alleviates the shortcomings posed by the positivist and post-positivist paradigms. The concept of critical realism becomes a popular perspective for viewing social reality, also in its digital layer. The assumptions of critical realism can be successfully applied to virtual reality research, regardless of the subject carried out. Critical realism remains an interesting research program to explore social reality in an in-depth way. A reflection on its creation is intended to alleviate the shortcomings of positivist and post-positivist approaches.

References

[1] Archer, Margaret A., Andrew Collier, Douglas V. Porpora, “Transcendencja. Realizm krytyczny i Bóg”, przedł. i wstęp Artur Wysocki, posłowie Krzysztof Wielecki, UKSW, Warszawa, 2021, ISBN: 978-83-8090-833-8.

[2] Ash, Steve, “Explaining Morality: Critical Realism and Moral Questions”, Routledge, London and New York, 2022, ISBN: 9781003080442.

[3] Bhaskar, Roy, “A Realist Theory of Science”, with a Introduction by Mervyn Hartwig, London and New York, Routledge, 2008, ISBN: 978-0-203-89263-3.

[4] Bhaskar, Roy, “Foreword”, [in:] “Studying Organizations Using Critical Realism. A Practical Guide”, Paul K. Edwards, Joe O’Mahoney, Steve Vincent, eds. Oxford University Press, Oxford, 2014, p. VII.

[5] Bhaskar, Roy, “Enlightened Common Sense: The Philosophy of Critical Realism”, edited with a preface by Mervyn Hartwig, Abingdon, Routledge, 2016. ISBN: 978-1-315-54294-2.

[6] Bhaskar, Roy, „Realistyczna teoria nauki”, trans. Katarzyna Zahorodna [in:] „Realizm wobec wyzwań antyrealizmu. Multidyscyplinarny przegląd stanowisk”, Marek Sikora, ed., Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 2011, ISBN: 978-83-7493-614-9.

[7] Biłobran, Czesław, *Społeczne konstruowanie operatorów dominacji nad zagrożeniami występującymi lokalnie na przykładzie powiatu nyskiego*, Uniwersytet Wrocławski, Wrocław, 2022.

[8] Brent, Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, Luciano Floridi (2016). “The ethics of algorithms: Mapping the debate”, Big Data & Society, 3(2) 2022, p.2. DOI: 10.1177/2053951716679679.

[9] Danah Boyd, Kate Crawford, “Critical questions for big data”, Information, Communication & Society, 15:5, 2012 p. 663, DOI: 10.1080/1369118X.2012.678878.

[10] Danermark Berth, Mats Ekstrom, Liselotte Jakobsen, Jan Ch. Karlsson (2001), “Explaining Society: An Introduction to Critical Realism in the Social Sciences”. London and New York, Routledge, 2001, ISBN: 0–415–22183–8.

[11] Porpora, Douglas V., “A reflection on critical realism and ethics”, Journal of Critical Realism, 18:3, 2019, p. 274, DOI: 10.1080/14767430.2019.1618064.

[12] Elemer, Hankiss, “Fears and Symbols: An Introduction to the Study of Western Civilization”, CEU Press, Budapest, 2001, ISBN: 9780585439389.

[13] https://www.wired.com/2008/06/pb-theory/ (Pobrano 2.10.2022).

[14] Katelin Albert, Jonah Stuart Brundage, Paige Sweet, Frédéric Vandenberghe, “Towards a critical realist epistemology?”, Journal for the Theory of Social Behaviour, 50(3), 2020, p. 358. DOI: 10.1111/jtsb.12248.

[15] Kitchin, Rob, “Big Data, new epistemologies and paradigm shifts”, Big Data & Society, 1(1), 2014, p. 3,6, DOI: 10.1177/2053951714528481.

[16] Savage, Mike, Roger Burrows, “The coming crisis of empirical sociology”, Sociology, vol. 41, no. 5, 2007, pp. 885–899. DOI: 10.1177/0038038507080443.

[17] Mingers, John, Mutch Alistair, Willcocks Leslie, “Critical Realism in information systems research”, IS Quarterly, Vol. 37, No. 3, 2013, p. 797. http://irep.ntu.ac.uk/id/eprint/20232/1/216160_300.pdf (Pobrano 2.10.2022).

[18] Stephenson, David (2020), BIG DATA NAUKA O DANYCH I AI BEZ TAJEMNIC, tłum. Wojciech Bombik, HELION, Gliwice, 2020. ISBN: 978-83-283-5796-9.

[19] Stochmal, Małgorzata, Relacyjna moc darów troski i ofiarność druhów Ochotniczych Straży Pożarnych: Perspektywa krytycznego realizmu i ontologii społecznej, PWN, Warszawa, ISBN: 978-83-01-21925-3.

[20] Lytvyn, Vasyl, Victoria Vysotska, Oleh Veres (2018). *Ontology of Big Data Analytics*, „MEST Journal”, 6(1), 2018, DOI: 10.12709/mest.06.06.01.06.

Justyna Komorowska

Technologie zabezpieczające bazy danych w przedsiębiorstwie

Streszczenie

Ostatnich 25 lat upłynęło na kształtowaniu nowej gospodarki, której bazą stały się technologie informatyczne. Postęp zatrzymał się na tak zaawansowanym poziomie, że obecnie programiści skupiają się na doskonaleniu stworzonych już narzędzi informatycznych, zwłaszcza systemów mających na celu ochronę informacji. Bezpieczeństwo danych ma szczególne znaczenie dla przedsiębiorstw, które na co dzień przetwarzają bardzo dużo ważnych informacji. Dane, które są użytkowane przez przedsiębiorstwa są wyjątkowe ze względu na wysoki poziom zróżnicowania informacji, dlatego efektywny i bezpieczny przepływ tych danych jest tak bardzo istotnym elementem w funkcjonowaniu firmy. Utworzenie odpowiedniej struktury informatycznej oraz zastosowanie odpowiedniego oprogramowania ochronnego są jednymi z najważniejszych działań dotyczących zapewnieniu bezpieczeństwa, jakie powinny zastosować przedsiębiorstwa. Obecnie oprogramowania są nie tylko w stanie wykryć zagrożenie, ale także całkowicie zablokować złośliwego wirusa, chroniąc ważne informacje. Narzędzia są tak szczegółowo dopracowane, że są w stanie korelować różne ataki, tym samym ucząc się na jakie zdarzenia reagować w sposób optymalny, tworzący spersonalizowanych priorytet alertów. Różne oprogramowania ochronne są tak zaprojektowane, że mogą płynnie wymieniać się informacjami, co daje możliwość jeszcze dokładniejszej ochronny i sprawniejszej reakcji na zagrożenia oraz bezustannego monitorowania sieci.

Podział technologii ochronnych

Pomimo zróżnicowania tych technologii, stosuje się prosty podział decydujący o miejscu ich stosowania. W pierwszej kolejności wyróżnia się oprogramowania mające na celu wykrycie zagrożenia i informowanie o znalezionych nieprawidłowościach. Głównym zadaniem tych technologii jest kontrola i blokada przed wirusem. Przykłady takich systemów to DLP czy DAM Database Activity Monitoring[25]. Drugim rodzajem technologii są narzędzia kontrolujące i w razie ataku, od razu reagujące na zagrożenie. Do tego rodzaju technologii należą przede wszystkim najbardziej nowoczesne technologie EDR, które często wykorzystywane są jako substytut dla klasycznego oprogramowanie antywirusowego czy zaawansowane systemy SIEM.

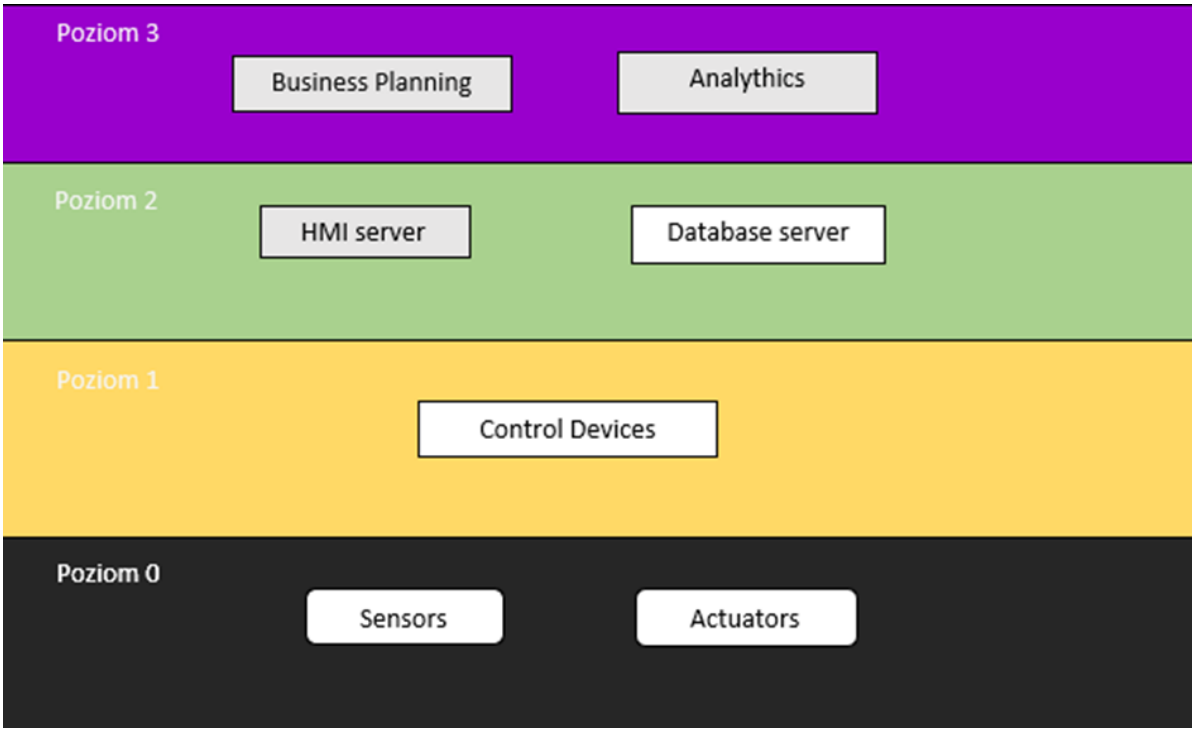
Architektura systemowa

Odpowiednio dobrana i zabezpieczona architektura systemowa jest istotna do sprawowania kontroli i zapewnienia bezpieczeństwa informacji dla firm, jednak szczególnie znaczenia ma ona dla przedsiębiorstw produkcyjnych. Dane zbierane z urządzeń produkcyjnych najczęściej potrzebne są kadrze zarządzającej, która znajduje się w obszarze sieci biurowej w celu analizy pracy zakładu. Taka analiza jest możliwa dzięki wbudowaniu specjalnych czujników, które umożliwiają monitorowanie wydajności maszyn oraz odpowiednie ich serwisowanie bez przerywania ich pracy. Urządzenia zbierające informacje ze względu na swoją strukturę są trudniejsze do zabezpieczenia niż sieci i urządzenia biurowe. Dlatego należy zapewnić odpowiednią strukturę minimalizując ryzyko zagrażające bezpieczeństwu związane z nierozważnym przesyłaniem danych z bazy do sieci przemysłowej. Aby spełnić te założenie dokonuje się segmentacji sieci, czyli podziału jeden głównej sieci na kilka mniejszych. Dzięki temu łatwiej jest zarządzać podzielonymi fragmentami, ale także zwiększa się tym samym również szybkość przysłania informacji. Bardzo często w celu dodatkowej ochrony sieci instaluje się firewalle, które pełnią funkcje zapory, filtrującej ruch pomiędzy poszczególnymi strefami oraz blokowanie informacji, które mogłyby zaszkodzić urządzeniom o delikatniejszej strukturze

[25] Niedźwiecki Zawiła J, Rostek Katarzyna, Gąsiorkiewicz Artur, Informatyka Gospodarcza, Wydawnictwo CH Beck, 2010.

podpiętym do sieci[26]. Architektura zabezpieczająca składa się z kilku poziomów, różniących się rodzajem zabezpieczeń stosowanych do obecnych na nich urządzeń informatycznych.

Rysunek 1. Poziomy architektury zabezpieczającej



Źródło: Opracowanie własne.

Na poziome 0 zazwyczaj znajdują się urządzenia wykonawcze bądź pomiarowe, na przykład pompy czy turbiny. Następie na urządzeniu umieszczane są specjalne czujniki, które zbierają informacje dotyczące ich parametrów pracy. Urządzenia z niższych warstw są zdecydowanie mniej zaawansowane technologicznie jest to spowodowane zazwyczaj intensywnym trybem pracy, można ich na bieżąco aktualizować, bo wtedy ograniczana byłaby produkcja. Maszyny produkcyjne są kupowane na dłuższy okres, dlatego odpowiednie monitorowanie ich zużycia i serwisowanie jest tak bardzo istotne. Zapewnia to efektywność danej produkcji. Na kolejnych poziomach znajdują się urządzenia do programowania i zarządzania wszystkimi maszynami z niższych stadium. Najbardziej znaczącym dla przedsiębiorstwa poziomem jest 3 dlatego, że następuje na nim połączenie danych przemysłowych z siecią biurową, niestety też zaczyna

[26] Kisielnicki Jerzy Henryk Sroka Systemy informatyczne biznesu Warszawa, Wydawnictwo Placet, str. 155-173,

się pojawiać coraz więcej zagrożeń. Jest to spowodowane głównie przez dostępu do Internetu oraz nieostrożność pracowników, którzy przez nieuwagę bądź niezastosowanie się do zasad polityki bezpieczeństwa mogą doprowadzić do zainfekowania systemu złośliwym oprogramowaniem. Wirusy, które dostałyby się na tym poziomie z komputerów firmowych mogłyby bez zabezpieczeń z łatwością przeniknąć do niższych poziomów infrastruktury. Dlatego aby uniknięcia zagrożenia atakiem przedsiębiorstwach stosują politykę prewencyjną, obowiązkowe back-upy zarówno w chmurze jak i na urządzeniach zewnętrznych[27]. Te zapobiegawcze działania umożliwiają powrót do danych, jednak mimo wszystko konieczna jest separacja poszczególnych elementów architektury systemowej zwłaszcza niższych poziomów.

Zasada ograniczonego dostępu

Zasada ograniczonego dostępu jest rodzajem ochrony polegającym na dokładnym określeniu jaki pracownik ma uprawnienia do korzystania z określonych informacji. W przypadku tej zasady korzysta się z firewalli, ale także z polityk zarządzaniami uprawnieniami oraz oprogramowań monitorujących i aktywność w sieci biurowej. Istnieją też specjalne programy, które ułatwiają stosowanie tej zasady w praktyce. Pracownik z danego działu loguje się do bazy danych przedsiębiorstwa, podając swój ustalony login a następnie podaje kod, który jest generowany wyłącznie dla jego osoby i dzięki temu może pracować na określonych plikach, którą są mu udostępnione. Taki system zabezpieczenia często jest wykorzystywany w poszczególnych działach w przedsiębiorstwie[28]. Oczywiście jeżeli pracownicy biurowi mają w jakiś sposób mieć dostęp do aplikacji przemysłowych stosuje się serwery przesiadkowe – dodatkowe pośredniczące serwery pomiędzy siecią biurową, a przemysłową oraz dostęp do aplikacji ustawianych w trybie read-only, czyli bez możliwości zmieniania treści, czyli oferujących wyłącznie sam podgląd na dane.

[27] Waldemar Zadworny, Marcin Kłak, Cloud Computing w nowoczesnym w modelowaniu biznesu,2019
[28] Dupont Stephen, Taming Big Data, The Strategist, 2014

Narzędzia Endpoint Threat Detection and Response

Narzędzia EDR Endpoint Threat Detection and Response należą do technologii umożliwiających namierzanie i reagowanie na nieprawidłową, nagłą zmianę pojawiającą się w systemie. Dzięki temu programowi jakiekolwiek niepożądane działanie zostaje wykryte i następnie zdezaktywowane. Technologia EDR jest często stosowana ze względu na to, że posiada umiejętność zapamiętywanie i uczenia się na jaki rodzaj nieprawidłowości reagować. Korelacja zaistniałych zdarzeń oraz zdolność do ich zapamiętywania powoduje ciągłe doskonalenie się tego systemu oraz coraz sprawniejsze reagowanie na zagrożenia[29]. Narzędzia EDR są stosowane w przedsiębiorstwach, które są szczególnie narażone na częste, trudno wykrywalne ataki. Systemy EDR zapamiętują nieprawidłowe zachowania i tworzą szczegółowy priorytet alertów. Zwykłe oprogramowania antywirusowe działają na bazie sygnatur, czyli fragmentu kodu wirusa, który umożliwia jego identyfikacje. Kiedy w systemie zostaje napotkany wirus zdaniem oprogramowania jest szybka detekcja i zablokowanie złośliwego kodu na podstawie sygnatury, którą rozpoznaje szkodliwe oprogramowanie. Obecnie niestety taka ochrona nie zawsze się sprawdza, ponieważ codziennie tworzą się miliony nowych wirusów, dlatego systemy antywirusowe nie są zawsze w stanie wykryć zmutowaną nową sygnaturę. Wystarczy zmienić jedną sekwencje nieważną z punktu działania programu i pojawia się inny znacznik, a nawet już tak minimalna zmiana nie może być wykryta przez standardowy program ochronny. Jest zatem bardzo łatwo oszukać system antywirusowy, który opiera się tylko na samych fragmentach wirusów, bez analizy ich zachowania w systemie. Narzędzia EDR, specjalizują się w analizie zachowania plików, programów i jeśli uznają zachowania za niebezpieczne to je blokują, dlatego też są nazywane antywirusami nowej generacji. EDR-y stosuje się głównie na komputerach związanych z ochroną danych klienta, stąd nazwa końcówki klienckie, ale także te narzędzia mogą być stosowane do innych rodzajów danych.

Narzędzia Data Loss Prevention

Systemy DLP (Data Loss Prevention) należą do technologii monitorujących, których zadaniem jest informowanie o wycieku informacji. Bazy danych, które mają bardzo dużo rekordów potrzebują odpowiednio zaawansowanego oprogramowania, aby chronić wszystkie miejsca,

[29] Cristian Mihai.The importance of databases in economy, Revista Economică ,2017

w których mogłoby dojść do utraty danych. Systemy DLP służą temu, żeby wychwytywać przypadkowe jak i celowe próby przechwycenia danych, dlatego są bardzo często wybieranym oprogramowaniem ze względu na swoją uniwersalność. DLP kontroluje sposób pracy, między innymi wychwytuje czy dany pracownik wysyła informacje, których nie powinien udostępnić, czy podpina USB do komputera stacjonarnego oraz innych działań, które mogą zagrozić bezpieczeństwu danych. System zauważa nieprawidłowe szyfrowanie oraz koduje wrażliwe dokumenty w taki sposób, by nigdy nie opuszczały one organizacji w formie nieochronionej, zaś wewnątrz były możliwe do oczytania tylko dla osób uprawnionych. Sposób w jaki system zareaguje, gdy pojawi się zagrożenie zależy od ustawionego rozwiązania przez użytkownika docelowego. Może zwyczajnie przerwać prace i wyświetlić komunikat albo zatrzymać transmisję danych i powiadomić dział IT. Niektóre z systemów DLP pozwalają również a zablokowanie takich możliwości jak robienie screenshotów czy drukowanie.

Systemy klasy SIEM

SIEM jest bardzo zaawansowanym systemem, który łączy wiele cech innych wymienionych technologii bezpieczeństwa. Służy zarówno do monitorowania oraz do dokładnej analizy zagrożeń jak i również pomaga aktywnie reagować na próbę przechwycenia danych. Wraz z postępującym rozwojem technologii informatycznych zwłaszcza w przypadku przedsiębiorstw, następuje duży przepływ danych co skutkuje fuzją danych napływających z wielu źródeł[30]. Tym samym pojawia się coraz więcej możliwości ataku, dlatego prawidłowa korelacja między różnymi zdarzeniami stała się konieczna. Wzrost liczby ataków spowodował, że wymagania dotyczące zgodności z przepisami i normami są coraz bardziej rygorystyczne. Stąd coraz większe zainteresowanie kompleksowymi systemami kontroli bezpieczeństwa, które oferują możliwości nadzorowania, raportowania i natychmiastowej odpowiedzi na atak. Wszystkie te czynności oferują systemy SIEM, dlatego są tak często wybierane przed przedsiębiorstwa, które operują dużymi bazami danych. Nie każda jednak organizacja może sobie pozwolić na zakup takiego oprogramowanie. Jest to spowodowane wysokim skomplikowaniem obsługi dlatego że do odpowiedniego działania systemu potrzebny jest doświadczony zespół odpowiednio przeszkolonych osób reagujących w przypadku wykrytego

[30] Susan Davidson, Philip Bernstein, Challenges and Opportunities with Big Data,2020

zagrożenia. System klasy SIEM nie jest autonomicznym narzędziem, potrzebnym do jego obsługi jest wyspecjalizowana grupa pracowników, która analizuje działanie oprogramowania 24 godziny na dobę. Kadra SIEM składa się z trzech linii. Pierwsza lina zajmuje się nadzorowaniem czy dane zdarzenie jest incydentem, jeżeli zostanie wykryte to zagrożenie to wprowadzana jest określona procedura i następnie przekazywana dalej do grupy drugiej, która zajmuje się analizowanie złożonych alertów przypadku, gdy alarm okazuje się proponowane są rekomendacje zmiany konfiguracji firewalli i usprawnienia innych narzędzi zabezpieczeń. Trzecią lina to linia dynamicznej interwencji wobec ataku na dane. System SIEM jest technologią wymagającą pracy zespołu, aby w pełni zachować bezpieczeństwo i na bieżąco analizować zagrożenia, dzięki czemu jest tak bardzo skuteczny oraz posiada możliwość integracji innych systemów bezpieczeństwa. Z tego powodu bardzo często przedsiębiorstwa korzystają z outsourcingu, wynajmują firmy zewnętrzne, które sprawują kontrolę nad bezpieczeństwem. Jeśli pojawia się poważne zagrożenie eskalują alert i wzywają specjalistów od ochronny danych. Natomiast jeśli wszystko jest w porządku to zazwyczaj co miesiąc wysyłają raport z systemu i jeśli firma chciałaby ulepszyć system bezpieczeństwa to na jej specjalną prośbę przeprowadzana zostaje symulacja ataku a następnie proponowane są nowe innowacje poprawiające obecne zabezpieczeń.

Database Activity Monitoring

DAM Database Activity Monitoring, jest rodzajem technologii służącej do monitorowania oraz analizowania aktywności w bazach danych, a w przypadku ataku również do blokowania nieautoryzowanych działań zagrażających bezpieczeństwu. Wnikając bardziej w jego szczegółowe zastosowanie wyróżnia się przede wszystkim: kontrolę działań użytkowników, identyfikowanie wszystkich nieprawidłowych działań oraz alarmowanie w przypadku złamania polityk bezpieczeństwa. DAM pozwala również na sprawniejsze wykrywanie oszustów i innych nadużyć, które przeprowadzane są za pośrednictwem innych aplikacji, a nie poprzez bezpośredni dostęp do baz danych. Aplikacje, które połączone są z zbiorem informacji, zbudowane są zazwyczaj z warstwy pośredniej, znajdującej się pomiędzy użytkownikiem końcowym a bazą danych. Tożsamość osób, które mają dostęp do informacji jest maskowana, a wszelkie połączenia i operacje na danych przeprowadzane są na wspólnym koncie usługi. System DAM ma możliwość powiązania konkretnego działania z określoną użytkownikiem

danych co pomaga zidentyfikować nieautoryzowane lub podejrzane działania. DAM może zapobiegać takim działaniom dzięki stworzeniu linii bazowej aktywności na serwerze oraz monitorowaniu wszystkich operacji. Istotną cechą tego oprogramowania jest możliwość integracji z innymi systemami ochronnymi na przykład oprogramowaniem SIEM czy DLP, dzięki czemu przedsiębiorstwo może zwiększyć efektywność działania w przypadku zagrożenia.

Podsumowanie

Wraz z tym gwałtownym rozwojem technologii informatycznych, pojawiły się także nowe metody kradzieży, dlatego obecnie wszystkie organizacje zarówno publiczne jak i prywatne podejmują wiele działań mających na celu zapewnienie maksymalnej ochrony. Zaostrzanie polityki bezpieczeństwa w przedsiębiorstwach przejawia się przede wszystkim poprzez zastosowanie nowych szybko reagujących na atak systemów, ale także przez poprawienie zabezpieczeń sieci architektury czy monitorowanie oraz wprowadzania ograniczonego dostępu do danych. Bardzo istotna jest też kwestia doboru nowoczesnych rodzajów zabezpieczeń do danych, które są najbardziej eksploatowane wewnątrz organizacji. Obecnie nadal są stosowane klasyczne metody ochrony informacji takie jak systemy antywirusowe, ale także pojawiły się zaawansowane oprogramowania dostępne dla przedsiębiorstw takie jakie EDR czy SIEM oraz inne programy monitorujące i reagujące na zagrożenie spowodowane podejrzaną aktywnością w bazach danych sposób szybszy i efektywniejszy od klasycznych oprogramowań. Widoczny wzrost świadomości w kwestii ochrony danych przeciętnych użytkowników jak i przedsiębiorstw, będzie skutkował coraz bardziej innowacyjnymi oraz zaawansowanymi systemami ochronnymi.

Bibliografia

[1] Cristian Mihai.The importance of databases in economy, Revista Economică ,2017.

[2] D. Dziembek, Cloud Computing – charakterystyka i obszary zastosowań w przedsiębiorstwach.

[3] Dupont Stephen, Taming Big Data, The Strategist, 2014.

[4] Innowacje w zarządzaniu i inżynierii produkcji, R. Konosala (red.), tom 2, Oficyna Wydawnicza PTZP,2016.

[5] Kędziora Mariusz Chmura (Cloud Computing) (online)[dostęp 26.01.2013].

[6] Kisielnicki Jerzy Henryk Sroka Systemy informatyczne biznesu Warszawa, Wydawnictwo Placet, str. 155-173,215-236,2013.

[7] Niedźwiecki Zawiła J, Rostek Katarzyna, Gąsiorkiewicz Artur, Informatyka Gospodarcza, Wydawnictwo CH Beck, ,2010.

[8] Susan Davidson, Philip Bernstein, Challenges and Opportunities with Big Data,2020.

[9] Terence Serendipity in the Data-Driven Era, 2017.

[10] Waldemar Zadworny, Marcin Kłak, Cloud Computing w nowoczesnym w modelowaniu biznesu,2019.

Łukasz Piętak

Wykorzystanie modeli autokorelacji przestrzennej do analizy występowania efektu dyspersji (spillover effect)

Wstęp

Występowanie efektu dyfuzji jest częstym tematem badań w ekonomii. Teorie rozwoju regionalnego autorstwa Perroux (1955;1964) czy Myrdala (1957) zakładają możliwość kumulowania się procesów ekonomicznych w przestrzeni geograficznej, co później daje podstawy do dyfuzji rozwoju na obszary przyległe. W erze globalizacji i swobodnego przepływu dóbr i usług zachodzących procesów nie można ograniczyć jedynie do jednego kraju lub regionu. Problematyka dyspersji (spillover effect) jest wykorzystywana w przypadku badań dotyczących rozprzestrzeniania się polityki spójności czy samego wzrostu gospodarczego (Antunes et al., 2020; Rodríguez-Pose et al., 2012).

W analizie ekonometrycznej do badania efektu dyfuzji wykorzystuje się modele autokorelacji przestrzennej, które uwzględniają aspekty sąsiedztwa pomiędzy obszarami. W przypadku danych przekrojowych nie istnieją poważniejsze problemy podczas estymacji. Oprogramowanie R oferuje pełen zakres funkcji pozwalający na wykonanie obliczeń. Problemy mogą pojawić się w przypadku danych panelowych. Wówczas pakiet R charakteryzuje się pewnymi brakami, którym jednak można zaradzić. Celem opracowania jest z jednej strony przedstawienie szerokich możliwości stosowania modeli autokorelacji przestrzennej w ramach analiz ekonometrycznych. Z drugiej strony opracowanie opisuje sposoby przygotowania danych

i modeli w przypadku modeli panelowych w programie R i wskazuje, jak poradzić sobie z ograniczeniami oprogramowania.

Artkuł składa się z trzech rozdziałów. Po wstępie, w rozdziale pierwszym scharakteryzowano budowę modeli panelowych oraz macierzy sąsiedztwa wykorzystywanych w modelach autokorelacji przestrzennej. W rozdziale drugim przedstawiono budowę modeli wykorzystywanych do badania efektu dyfuzji. Trzeci rozdział ma charakter empiryczny i charakteryzuje efekt dyfuzji pomiędzy regionami NUTS 3 w Polsce.

Modele panelowe i macierz sąsiedztwa

Estymacja modeli panelowych jest bardziej efektywna niż modeli przekrojowych. Zwiększa ona liczbę stopni swobody oraz ogranicza korelację pomiędzy zmiennymi objaśniającymi (Hsiao, 2003). W przypadku danych panelowych analizę rozpoczyna się wykorzystując estymator klasycznej metody najmniejszych kwadratów (KMNK). Następnie stosuje się testy specyfikacyjne, a mianowicie testy Breuscha-Pagana (1980) i Hausmana (1978). Test Breuscha-Pagana weryfikuje, czy estymator efektów stałych (ES) powinien zastąpić specyfikację KMNK. Natomiast test Hausmana sprawdza, czy estymator efektów losowych (EL) stanowi alternatywę dla estymatora ES. Zaletą estymatora ES jest uwzględnienie efektów specyficznych dla każdej jednostki (regionu, kraju, przedsiębiorstwa etc.) nieobjętych zmiennymi objaśniającymi, co poprawia dokładność estymacji (Baltagi, 2005). Z drugiej strony, większa zmienność danych w poszczególnych jednostkach (obszarach, regionach) zmniejsza dokładność oszacowań estymatora ES (Partridge, 2005). Ponieważ specyfikacja ES uwzględnia zmienność szeregów czasowych w każdym regionie (kraju), jej wyniki są interpretowane jako efekty krótkookresowe, podczas gdy specyfikacja KMNK wyraża efekty długookresowe (Partrige, 2005).

W przypadku estymacji modeli panelowych dane w panelu uporządkowane są najpierw według jednostek a następnie według czasu. Z kolei w przypadku paneli przestrzennych dane należy uporządkować najpierw według jednostek czasu a następnie według jednostek wchodzących w skład panelu. Rysunki 1a i 1b przedstawiają różnice pomiędzy klasycznym panelem oraz panelem wykorzystywanym w analizach przestrzennych. Brak przygotowania panelu na wzór rysunku 1b prowadzi do błędnych oszacowań modeli.

Rys. 1a Przykład danych panelowych

Region	Rok	Dane 1	Dane 2	Dane 3
A	2000	1	3	4
A	2001	7	6	4
A	2002	12	7	6
B	2000	2	12	8
B	2001	5	7	2
B	2002	7	4	1
C	2000	4	1	5
C	2001	6	5	3
C	2002	8	4	6

Źródło: opracowanie własne.

Rys. 1b. Przestrzenne dane panelowe

Region	Rok	Dane 1	Dane 2	Dane 3
A	2000	1	3	4
B	2000	2	12	8
C	2000	4	1	5
A	2001	7	6	4
B	2001	5	7	2
C	2001	6	5	3
A	2002	12	7	6
B	2002	7	4	1
C	2002	8	4	6

Źródło: opracowanie własne.

W przestrzennych modelach panelowych oprócz danych ilościowych występują także dane wyrażające położenie geograficzne jednostek. Najczęściej relacje pomiędzy regionami opisane

są przy pomocy macierzy sąsiedztwa. Wówczas element macierzy dla regionów (krajów) posiadających wspólną granicę wynosi 1, zaś w przeciwnym przypadku jest równy 0. Należy zaznaczyć, że macierz sąsiedztwa może być budowana na wiele sposobów uwzględniając przepływy handlowe pomiędzy jednostkami czy też ograniczając się do *k* najbliższych jednostek danego podmiotu. Rysunek 2 przedstawia przykładową macierz sąsiedztwa dla czterech jednostek (macierz a). Macierz ta następnie jest sumowana według wierszy (macierz b) i standaryzowana do jedności (macierz c). Tak przygotowaną macierz można wykorzystać w programie R. Jednakże nie jest to konieczne, wykorzystując pliki o rozszerzeniu .shp oprogramowanie R samo stworzy macierz sąsiedztwa bazującą na wspólnej granicy. W załączniku przedstawiono procedurę tworzenia macierzy dla przykładowych 17 jednostek bazującej na wspólnej granicy w R (zał. 1).

Rysunek 2. Macierz sąsiedztwa-wspólna granica

cztery jednostki

A	B
C	D

a

	A	B	C	D
A	0	1	1	0
B	1	0	0	1
C	1	0	0	1
D	0	1	1	0

b

	A	B	C	D	suma
A	0	1	1	0	2
B	1	0	0	1	2
C	1	0	0	1	2
D	0	1	1	0	2

c

	A	B	C	D	suma
A	0	0.5	0.5	0	1
B	0.5	0	0	0.5	1
C	0.5	0	0	0.5	1
D	0	0.5	0.5	0	1

Źródło: opracowanie własne.

Przygotowanie plików .shp, które później zostaną wykorzystane w programie R do stworzenia macierzy sąsiedztwa wymaga realizacji kilku kroków. Przede wszystkim należy posiadać plik graficzny .shp jednostek wykorzystanych w analizie ekonometrycznej. Jednakże nie zawsze można odszukać w Internecie mapy z podziałem na dane regiony czy kraje, czasem badacz zainteresowany jest kilkoma specyficznymi krajami jak Kraje Bałtyckie, czy też Europa Środkowo-Wschodnia, wówczas taką mapę musimy stworzyć samemu. Przydatnym do tego jest program Q-Gis, który pozwala na wyodrębnienie pewnych jednostek (regionów, krajów) z ogólnej ich ilości (np. wszystkie kraje lub regiony UE). Następnie przygotowaną mapę należy

zapisać w formacie .shp. Stworzony plik można otworzyć w programie Geoda, który pozwala na obszerną wizualizację danych czy też wstępną analizę autokorelacji przestrzennej danych wykorzystanych w analizie ekonometrycznej (np. inwestycje czy kapitał ludzki). Należy także pamiętać, że kolejność jednostek z pliku .shp powinna pokrywać się z kolejnością tychże jednostek w panelu. Zatem należy panel danych posortować zgodnie z pozycją danej jednostki w pliku shp.

Modele autokorelacji przestrzennej wykorzystywane do badania efektu dyfuzji

Model regresji liniowej przedstawiający wpływ zmiennych niezależnych na zmienną zależną wyraża się wzorem

$$y_{i,t} = +\beta X_{i,t} + e_{i,t} \quad (1)$$

gdzie $y_{i,t}$ jest zmienną zależną uzyskaną w jednostce i w czasie t, $X_{i,t}$ jest zbiorem zmiennych egzogenicznych, zaś $e_{i,t}$ wyznacza błąd oszacowań. Jednakże w modelach autokorelacji przestrzennej klasyczny model wyrażony równaniem 1 zyskuje nowe elementy pozwalające śledzić zależności geograficzne pomiędzy jednostkami i ich wpływ na zmienną objaśnianą.

W modelach przestrzennych możemy kontrolować korelację przestrzenną dla trzech elementów, tj. autokorelacji przestrzennej: zmiennej zależnej (ρ), zmiennych niezależnych (ϑ) oraz reszt (λ). Ponieważ głównym przedmiotem rozważań jest istnienie efektu dyfuzji (spillover effect), zostaną omówione modele, które z jednej strony zawierają wyżej wymienione elementy przestrzenne, zaś z drugiej strony pozwalają na śledzenie właśnie efektów dyfuzyjnych. W przypadku efektu dyfuzji wskazuje się na jego globalny i lokalny charakter (Aselin, 2003). Ze względu na efekt sprzężenia zwrotnego, globalny charakter dyfuzji sprawia, że zmiany zachodzące w jednej jednostce są przekazywane do wszystkich jednostek objętych analizą, pomimo braku wspólnej granicy. Podczas gdy lokalny charakter dyfuzji występuje tylko pomiędzy regionami posiadającymi wspólną granicę.

Modele (2) i (3) przedstawiają dwie specyfikacje służące do śledzenia globalnych efektów dyfuzyjnych, tj. odpowiednio model autoregresji przestrzennej (Spatial Autoregresive Model - SAR) i model przestrzenny Durbina (Spatial Durbin Model - SDM)

$$y_{i,t} = \rho W y_{i,t} + \beta X_{i,t} + e_{i,t} \quad (2)$$

$$y_{i,t} = \rho W y_{i,t} + \beta X_{i,t} + \theta W X_{i,t} + e_{i,t} \quad (3)$$

gdzie ρ oznacza parametr autoregresyjny zmiennej zależnej, X jest zbiorem regresorów, W oznacza macierz sąsiedztwa, zaś e jest błędem estymacji modelu. Dodatnia i istotna statystycznie wartość ρ wskazuje na istnienie klastrów jednostek podobnych. Ujemna wartość współczynnika wyraża odmienność, która może mieć odzwierciedlenie w konkurencji regionalnej lub efekcie płukania (Kao i Bera, 2013). Macierz wag przestrzennych opisuje sąsiedztwo między jednostkami, gdzie w_{ij} jest elementem macierzy W, takim, że $w_{ij} = 1$ jeśli jednostki i oraz j są sąsiadami, zaś $w_{ij} = 0\ w$ przeciwnym razie, niezależnie od długość granicy. Różnica między dwoma modelami polega na tym, że w modelu SAR pominięto składnik Durbina (ϑWX). Ponadto model SAR ma kilka ograniczeń. Elhorst (2010) podkreśla, że stosunek efektów bezpośrednich do pośrednich jest taki sam dla wszystkich regresorów, co nie zostało potwierdzone w wielu badaniach empirycznych. Z kolei Pinkse i Slade (2010) wskazują, że cała przestrzenna struktura zależności w modelu SAR sprowadza się do jednego nieznanego parametru. Zgodnie z LeSage (2014), aby ocenić globalne efekty dyfuzyjne, analiza ekonometryczna jest ograniczona do modelu SDM reprezentowanego przez model 3.

Zmienna zależna (y) pojawia się po obu stronach modelu 3, po przepisaniu otrzymujemy

$$y_{i,t} = (I - \rho W)^{-1}\beta X_{i,t} + (I - \rho W)^{-1}\theta W X_{i,t} + (I - \rho W)^{-1} e_{i,t} \quad (4)$$

Parametr $(I - \rho W)^{-1}$ wyznacza siłę sprzężenia zwrotnego i oznacza, że zmiany w zmiennej egzogenicznej w jednej jednostce wpływają nie tylko na obszary przygraniczne, ale na wszystkie jednostki w próbie. Porównując modele lokalnych efektów zewnętrznych (model 5 i 6) oraz globalnych efektów zewnętrznych (model 3) dostrzegamy, że główna różnica leży w strukturze macierzy W. W przypadku specyfikacji SLX (Spatial Lag X) i SEDM (Spatial Error Durbin Model) macierz W zawiera elementy równe zero, z wyjątkiem sytuacji, gdy regiony mają wspólną granicę. W przypadku modelu SDM macierz $(I - \rho W)^{-1}$ nie zawiera elementów równych zero, a zmiany zmiennych objaśniających przynoszą efekty we wszystkich jednostkach. Ponadto w przypadku modelu SDM współczynniki nie odzwierciedlają bezpośrednio marginalnych efektów zmiennych objaśniających. Jak zaznacza Le Sage i Pace (2008) efekt całkowity jest sumą efektów w wierszu macierzy $(I - \rho W)^{-1}$. Efekt bezpośredni sumuje elementy diagonalne macierzy $(I - \rho W)^{-1}$, natomiast efekt pośredni (efekt dyfuzji) jest wyrażony przez różnicę między efektem całkowitym a efektem bezpośrednim. Ponieważ przestrzennie opóźniona

zmienna zależna w modelu SDM stwarza problem endogeniczności, szacunki opierają się na estymatorze największego prawdopodobieństwa (Elhorst, 2014).

W przypadku oprogramowania R niestety nie pozwala ono na otrzymanie efektów bezpośrednich, pośrednich i całkowitych. Taka opcja jest możliwa jedynie w przypadku danych przekrojowych. Istnieją dwie metody aby zaradzić tej niedogodności. W pierwszym przypadku należy estymować model panelowy podobnie jak w przypadku danych przekrojowych wykorzystując tzw. dużą macierz sąsiedztwa (załącznik 1). W drugim przypadku należy zainstalować starsze wersje bibliotek (packages): *spdep 0.5-56* oraz *splm 1.4-11*. Wówczas wykorzystując starsze wersje *packages* jest możliwe wyodrębnienie wszystkich trzech rodzajów efektów. Należy także pamiętać o nieinstalowaniu pakietu *spatialreg*, ponieważ automatycznie zablokuje on funkcje pakietu *spdep* i nie będzie możliwe wyodrębnienie efektów. Procedura wygląda następująco:

```
time=length(unique(mydata3$Year))

sparse.W=listw2dgCMatrix(matW)

s.lwstates=kronecker(Diagonal(time), sparse.W)

set.seed(123456)

trMatc=trW(s.lwstates, type="mult")

imp=impacts(eq6, tr=trMatc,R=200)

summary(imp, zstats=TRUE, short=T)

summary(imp, zstats=TRUE)
```

W przypadku analizy lokalnych efektów dyfuzji rozważane są dwie specyfikacje ekonometryczne. Modele 5 i 6 przedstawiają odpowiednio model z opóźnieniem przestrzennym (Spatial Lag X Model - SLX) i model błędu przestrzennego Durbina (Spatial Durbin Error Model - SDEM),

$$growth_{i,t} = \beta X_{i,t} + \theta W X_{i,t} + e_{i,t} \quad (5)$$

$$growth_{i,t} = \beta X_{i,t} + \theta W X_{i,t} + u_{i,t} \quad (6)$$

$$u_{i,t} = \lambda W u_{i,t} + e_{i,t}$$

gdzie *X* jest zbiorem zmiennych objaśniających, *W* reprezentuje macierz wag przestrzennych, $u_{i,t}$ jest błędu estymacji, a λ jest przestrzennym parametrem autoregresji, który wyraża cechy, które są trudne do zmierzenia i nie zostały objęte przez zmienne egzogeniczne. Dodatnia wartość parametru λ wskazuje na istnienie nieobserwowanych klastrów przestrzennych, pomijanych przez specyfikację ekonometryczną tj. język, kulturę czy kapitał społeczny. Obydwa modele śledzą lokalne efekty dyfuzyjne, poprzez komponent Durbina (ϑWX). Zmienne egzogeniczne mają dwa różne znaczenia w modelach ze składnikiem Durbina. Zmienna ze zbioru *X* wyraża wpływ zmiennej z jednostki i na zmienną zależną *y* w obszarze i (efekt bezpośredni). Natomiast zmienna zawarta w składniku Durbina (ϑWX) opisuje efekt pośredni (efekt dyfuzyjny), wpływ zmiennej *X* w jednostce i na zmienną zależną w jednostce *j*. Podsumowując, w modelach SLX i SDEM efekt bezpośredni wyraża się parametrem β, a efekt pośredni θ.

Dyfuzja wzrostu gospodarczego pomiędzy regionami NUTS 3 w Polsce

Dane wykorzystane w badaniu zostały pobrane z następujących źródeł. Roczne dane dotyczące polskich podregionów NUTS 3 zaczerpnięto z bazy danych Banku Danych Lokalnych GUS. Fundusze strukturalne przeznaczone dla polskich podregionów uzyskano z bazy danych Ministerstwa Funduszy i Polityki Regionalnej. Agencja Restrukturyzacji i Modernizacji Rolnictwa (ARiMR) udostępniła dane dotyczące wartości dotacji w ramach wspólnej polityki rolnej. Plik graficzny podregionów NUTS 3 w UE został pobrany ze strony internetowej Eurostatu. Następnie mapa posłużyła do stworzenia mapy polskich podregionów NUTS 3. Analizę ekonometryczną wykonano przy użyciu oprogramowania R. Do tworzenia map i figur wykorzystano oprogramowanie Geoda i QGIS.

Wszystkie zmienne wykorzystane w badaniach zostały przekształcone na wartości logarytmiczne. Zmienną zależną jest roczne tempo wzrostu PKB per capita ($growth_{i,t}$) podregionu i w czasie *t*. Zmienna *gdppc* jest początkową wartością PKB per capita. Jego ujemna wartość potwierdza proces konwergencji, co oznacza, że biedniejsze obszary rozwijają się szybciej niż bogate. Zmienna funduszy strukturalnych (*fund*) wyraża średnią roczną wartość polityki spójności na mieszkańca. Zmienna *agri_fund* obejmuje dotacje polityki spójności i wspólnej polityki rolnej łącznie. Możliwe są pozytywne i negatywne znaki tych zmiennych. Przegląd literatury podkreśla brak konsensusu w tej dziedzinie badań (Di Caro, Fratesi, 2021; Di

Cataldo, 2017; Rodríguez-Pose, Novak, 2013) . Inną istotną zmienną jest gęstość zaludnienia (*dens*), która nawiązuje do koncepcji aglomeracji. Średnia wielkość miasta i wzrost liczby ludności zwiększają produktywność i wzrost gospodarczy (Frick i Rodríguez-Pose, 2016; Fujita i Thisse, 2002; OECD, 2016; Castells-Quintana i Royuela, 2011). Ponadto Sala-i-Martin i in. (2004) wyliczają gęstość zaludnienia jako zmienną egzogeniczną pozytywnie wpływającą na wzrost gospodarczy. Wartość inwestycji (nakłady brutto na środki trwałe) nie jest przedstawiana w statystykach krajowych na poziomie podregionów, dlatego też w badaniach wykorzystano zmienną *proxy*, która wyraża inwestycje (*invest*) jako udział nakładów inwestycyjnych w przedsiębiorstwach w PKB. Oczekuje się dodatniej korelacji między inwestycjami a wzrostem gospodarczym. Według badań empirycznych inwestycje mają pozytywny wpływ na regionalny wzrost gospodarczy (Mas et al., 1996; Pereira, Roca-Sagalés, 2003; Rodríguez-Pose et al., 2012). Analiza ekonometryczna obejmuje także zmienną bezrobocia (*unemploy*). Oczekuje się negatywnego znaku tej zmiennej. Osoby bezrobotne nie mogą uczestniczyć w procesie produkcyjnym.

Analiza ekonometryczna

Tabela 1 przedstawia oszacowania modelu SDM. Specyfikacja umożliwia śledzenie globalnych efektów zewnętrznych. Opóźniony współczynnik zmiennej zależnej (ρ) jest dodatni i istotny statystycznie we wszystkich modelach. Oznacza to, że model wzrostu w sąsiednich podregionach pozytywnie wpływa na lokalny wzrost gospodarczy. Zmienna *gdppc* jest ujemna i istotna statystycznie, co świadczy o istniejącej konwergencji pomiędzy subregionami. Ponadto szacunki wskazują na pozytywną i statystycznie istotną zależność między inwestycjami a wzrostem oraz niekorzystny wpływ bezrobocia na wzrost. Zmienne *fund* i *agri_fund* mają pozytywne znaki, ale są one nieistotne statystycznie. W Tabeli 1 przedstawiono również wartości statystyki Loglik dla specyfikacji SAR i SEM. Ich wartości są zbliżone do wartości modelu SDM, co oznacza, że model SDM może być zredukowany do prostszej wersji, takiej jak SAR lub SEM.

Współczynniki modelu SDM nie wyrażają bezpośredniego wpływu regresorów na zmienną zależną. Tabela 2 zawiera rezultaty odnoszące się do specyfikacji ze zmiennymi odpowiednio *fund* i *agri_fund*. Wyniki zostały podzielone na efekty bezpośrednie, pośrednie i całkowite. Rezultaty potwierdzają, że globalny efekt dyspersji w przypadku tych regresorów nie występuje,

efekt pośredni tych zmiennych jest statystycznie nieistotny. Zjawisko dyspersji ma miejsce w przypadku bezrobocia oraz inwestycji. Negatywny i istotny statystycznie wpływ pośredni potwierdza, że bezrobocie w jednym podregionie negatywnie wpływa na wzrost we wszystkich podregionach próby.

Kolejny punkt analizy empirycznej weryfikuje występowanie lokalnych efektów dyfuzyjnych. W tabeli 3 przedstawiono wyniki estymacji modelu (5). Lokalne efekty dyfuzyjne są modelowane za pomocą opóźnień regresorów (ϑWX) modelu SLX. Szacunki ponownie potwierdzają konwergencję, zmienna *gdppc* jest ujemna i statystycznie istotna. Wyniki potwierdzają pozytywny wpływ inwestycji oraz negatywny wpływ bezrobocia na wzrost. Ponadto ujemna i istotna statystycznie zmienna *W×unemploy* potwierdza istnienie dyspersji bezrobocia w ujęciu lokalnym, co oznacza, że jego wzrost w danym regionie negatywnie wpływa na wzrost gospodarczy w regionach granicznych.

Tabela 11 przedstawia wyniki modelu SDER, który również bada procesy dyfuzyjne w wymiarze lokalnym. Otrzymane rezultaty potwierdzają oszacowania otrzymane w ramach modelu SLX. Bezrobocie jest jedyną zmienną potwierdzającą lokalną dyspersję bezrobocia, pozostałe zmienne pozostały statystycznie nieistotne.

Podsumowanie

Modele autokorelacji przestrzennej stanowią alternatywę wobec modeli klasycznych. Poprzez uwzględnienie czynnika geograficznego w bardziej realny sposób opisują gospodarczą rzeczywistość. Jednakże bardziej skomplikowana budowa modeli sprawia, że ich estymacja jest trudniejsza. W przypadku modeli wykorzystujących dane przekrojowe programy ekonometryczne jak R dostarczają pełne pakiety funkcji pozwalających na przeprowadzenie procesu estymacji i jego weryfikacji. Jednakże w przypadku danych panelowych pojawiają się problemy, które po części utrudniają analizę ekonometryczną. Rozważania przedstawione w rozdziale opisują jak owe trudności można zniwelować. Należy pamiętać, że w przypadku przestrzennych modeli panelowych nie ma ustalonej procedury postępowania podczas procesu estymacji, który ma raczej cechy „sztuki" i wymaga od badacza doświadczenia i rozumienia zachodzących procesów ekonomicznych.

Bibliografia

[1] Antunes, A., Viegas, M., Varum, C., & Pinho, C. (2020). The impact of Structural Funds on Regional Growth: A Panel Data Spatial Analysis. *Intereconomics*, 55(5), 312–319.

[2] Anselin, L. (2003). Spatial Externalities, Spatial Multipliers, and Spatial Econometrics. *International Regional Science Review,* 26 (2), 153–166.

[3] Baltagi, B. H. (2005). *Econometric analysis of panel data*. John Wiley, Chichester.

[4] Breusch, T. S., Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies*, 47(1), 239–253.

[5] Castells-Quintana, D., & Royuela, V. (2011). *Agglomeration, Inequality and Economic Growth*. IREA-WP series, no. 2011/14.

[6] Di Caro, P., & Fratesi, U. (2021). One policy, different effects: Estimating the region-specific impacts EU cohesion policy. *Journal of Regional Science*, doi.org/10.1111/jors.12566.

[7] Di Cataldo, M. (2017). The impact of EU objective 1 funds on regional development: Evidence from the U.K. and the prospect of Brexit. *Journal of Regional Science*, 57(5), 120–141.

[8] Elhorst, J. P. (2010). Applied Spatial Econometrics: Raising the Bar. *Spatial Economic Analysis*, 5(1), 9–28.

[9] Elhorst, J. P. (2014). *Spatial econometrics*. Springer.

[10] Frick, A. F., & Rodríguez-Pose A. (2016). Average city size and economic growth. *Cambridge Journal of Regions, Economy and Society*, 9(2), 301–318.

[11] Fujita, M., & Thisse, J. F. (2002). *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge University Press.

[12] Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251–1271.

[13] Hsiao, C. (2003). *Analysis of panel data*. Cambridge University Press, Cambridge.

[14] LeSage, J. P. (2014). What Regional Scientists Need to Know About Spatial Econometrics. Available at SSRN: https://ssrn.com/abstract=2420725.

[15] LeSage, J. P., & Pace, R. K. (2008). *Introduction to spatial econometrics*. CRC press.

[16] Mas, M., Maudos, J., Pérez F. J., & Uriel, E. (1996). Infrastructures and productivity in the Spanish regions. *Regional Studies*, 30(7), 641–649.

[17] Myrdal, G. (1957). *Economic Theory and Underdeveloped Regions*. Gerald Duckworth and Co. Ltd, London.

[18] OECD (2016). *OECD Regional Outlook 2016: Productive Regions for Inclusive Societies*. (Paris: OECD Publishing).

[19] Partridge, M. D. (2005). Does income distribution affect US state economic growth? *Journal of Regional Science*, 45 (2), 363–394.

[20] Pereira, A. M., Roca-Sagalés, O. (2003). Spillover effects of public capital formation: Evidence from the Spanish regions. *Journal of Urban Economics*, 53(2), 238–256.

[21] Perroux, F. (1964). *L'Économie du XX^{e} siècle*. Presses universitaires de France, Paris.

[22] Perroux, F. (1955). Note sur la nation de pôle de criossance. *Economie Appliquée*, 7(1/2), 307–320.

[23] Pinkse, J., & Slade M. A. (2010). The Future of Spatial Econometrics. *Journal of Regional Science*, 50(1), 103–117.

[24] Rodríguez-Pose, A., Novak, K. (2013). Learning processes and economic returns in European Cohesion Policy. *Investigaciones Regionales*, 25, 7–26.

[25] Rodríguez-Pose, A., Psycharis, Y., & Tselios, V. (2012). Public investment and regional growth and convergence: Evidence from Greece. *Papers in Regional Science*, 91(3), 543–568.

[26] Sala-i-Martin, X., Doppelhofer, G., & Miller, R. (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review*, 94(4), 813–835.

Tabele

Tabela 1. Wyniki estymacji modelu SDM

Zmienna zależna: growth	1	2
Rho (ρ)	0.4904***	0.4884***
	(0.0303)	(0.0303)
Const	0.4352***	0.3689**
	(0.0749)	(0.0804)
gdppc_{t-1}	–0.0185**	–0.0178**
	(0.0059)	(0.0058)
fund	0.0037	
	(0.0035)	
agri_fund		0.0058
		(0.0046)
dens	0.0020	0.0024
	(0.0017)	(0.0017)
invest	0.0089*	0.0088*

	(0.0034)	(0.0034)
unemploy	–0.0093*	–0.0085*
	(0.0043)	(0.0042)
$W \times gdppc_{t-1}$	–0.0183*	–0.0191**
	(0.0072)	(0.0072)
W×fund	–0.0039	
	(0.0072)	
W×agri_fund		0.0025
		(0.0082)
W×dens	–0.0004	0.0022
	(0.0028)	(0.0039)
W×invest	–0.0024	–0.0014
	(0.0054)	(0.0054)
W×unemploy	–0.0080	–0.0085
	(0.0057)	(0.0058)
Observations	1022	1022
R^2	0.13	0.12
Log lik	2118.2	2118.0
AIC	–3922.7	–3921.6
Log lik SAR	2114.2	2116.0
Log lik SEM	2105.9	2106.7

*Źródło: opracowanie własne. Zmienna zależna: growth. Błędy standardowe umieszczono w nawiasach. *q < 0.1 oznacza poziom istotności na poziomie 10%. **q < 0.05 oznacza poziom istotności na poziomie 5%. ***q < 0.01 oznacza poziom istotności na poziomie 1%.*

Tabela 2. Efekty w modelu SDM

	Bezpośrednie	Pośrednia	Całkowite		Bezpośrednie	Pośrednie	Całkowite
lny	–0.0198***	–0.0166**	–0.0364***	lny	–0.0189**	–0.0158**	–0.0348**
	(0.0058)	(0.0054)	(0.0111)		(0.0062)	(0.0054)	(0.0114)
fund	0.0039	0.0033	0.0072	agri_fund	0.0062	0.0051	0.0113
	(0.0039)	(0.0034)	(0.0073)		(0.0049)	(0.0043)	(0.0092)
dens	0.0021	0.0018	0.0039	dens	0.0026	0.0021	0.0048
	(0.0018)	(0.0015)	(0.0033)		(0.0019)	(0.0016)	(0.0036)
invest	0.0094*	0.0079*	0.0173*	invest	0.0093*	0.0078*	0.0172*
	(0.0036)	(0.0033)	(0.0068)		(0.0037)	(0.0033)	(0.0071)
unemploy	–0.0099*	–0.0083*	–0.0183*	unemploy	–0.0090*	–0.0075*	–0.0166*
	(0.0044)	(0.0039)	(0.0083)		(0.0046)	(0.0041)	(0.0087)

Źródło: opracowanie własne.

Tabela 3. Rezultaty modelu SLX

Zmienna zależna: growth	1	2
Constant	0.7300***	0.7249***
	(0.0921)	(0.0920)
$\mathbf{gdppc}_{t-1}$	–0.0220***	–0.0217**
	(0.0067)	(0.0067)
fund	0.0005	
	(0.0041)	
agri_fund		0.0059
		(0.0052)
dens	0.0026	0.0032*
	(0.0019)	(0.0012)
invest	0.0104**	0.0097*

	(0.0039)	(0.0039)
unemploy	–0.0110*	–0.0097*
	(0.0049)	(0.0046)
$W \times gdppc_{t-1}$	–0.0497***	–0.0501***
	(0.0082)	(0.0082)
W×fund	0.0137	
	(0.0087)	
W×agri_fund		0.0088
		(0.0094)
W×dens	0.0054	0.0047
	(0.0044)	(0.0044)
W×invest	0.0051	0.0064
	(0.0062)	(0.0062)
W×unemploy	–0.0211**	–0.0222***
	(0.0066)	(0.0067)
Observations	1022	1022
R^2	0.15	0.15
F-statistic	17.502	17.649
	(0.0000)	(0.0000)

Źródło: opracowanie własne.

Tabela 4. Rezultaty modelu SDEM

Zmienna zależna: growth	1	2
Lambda (λ)	0.4881***	0.5076***
	(0.0308)	(0.0290)
Constant	0.6510***	0.6316***
	(0.1219)	(0.1231)
$gdppc_{t-1}$	–0.0242***	–0.0238***

	(0.0055)	(0.0054)
fund	0.0034 (0.0035)	
agri_fund		0.0062 (0.0043)
dens	0.0023 (0.0015)	0.0032* (0.0012)
invest	0.0095** (0.0036)	0.0093* (0.0035)
unemploy	–0.0112** (0.0041)	–0.0105*** (0.0041)
$W \times gdppc_{t-1}$	–0.0381*** (0.0085)	–0.0386*** (0.0085)
W×fund	0.0072 (0.0092)	
W×agri_fund		0.0062 (0.0092)
W×dens	0.0054 (0.0043)	0.0056 (0.0043)
W×invest	0.0026 (0.0066)	0.0035 (0.0067)
W×unemploy	–0.0159* (0.0067)	–0.0161* (0.0067)
Observations	1022	1022
R^2	0.14	0.15
Log lik Log lik SEM	2115.7 2105.9	2116.3 2106.7

Źródło: opracowanie własne.

Z recenzji dra hab. Zbigniewa Matyjasa

Dzielenie się własnym doświadczeniem badawczym świadczy o dużej dojrzałości piszących poszczególne rozdziały, co nieczęsto stanowi wartość wieloautorskich monografii naukowych [...]. Fragmenty rozdziałów, które traktują o metodach i technice prowadzenia badań mogą z powodzeniem być wykorzystane przez badaczy, którzy dopiero planują wykorzystanie AI i BD w swoich wysiłkach naukowych.

Monografia powinna stanowić lekturę obowiązkową każdego badacza nie tylko w obszarze nauk społecznych, ale także w humanistyce i naukach ścisłych. Szczególnie środowisko naukowe zajmujące się naukami społecznymi, przywiązane do „tradycyjnych" ilościowych metod badawczych i niechętnie akceptujące metody zaczerpnięte z nauk technicznych i nauk o informacji [powinno być zainteresowane tą monografią]. Siłą monografii jest jej bogactwo i różnorodność [...]. Stanowi ona ważny i warty polecenia wkład w branżową literaturę przedmiotu.

Z recenzji dra hab. Leszka Bohdanowicza

Redaktorzy naukowi zatroszczyli się o spójność monografii – pomimo jej wieloaspektowości i wielodyscyplinarności. Autorzy [...] opisują analizowane problemy badawcze w kontekście wojny hybrydowej, [zatem[monografia nabrała bardzo istotnej wartości poznawczej – szczególnie dla przedstawicieli nauk o bezpieczeństwie, ekonomistów i strategów.

Poprawność merytoryczna, spójność koncepcji i rzetelność metodologiczna to mocne strony recenzowanej monografii.

Największą zaletą monografii jest fakt, że nie ogranicza się ona jedynie do opisu ogólnych, teoretycznych rozważań nad istotą wykorzystania nowoczesnych narzędzi analitycznych, ale przybliża czytelnikowi praktykę „technologii" analiz.

Partner publikacji:

CRI CENTRUM RAFINACJI INFORMACJI

ISBN 978-83-68039-08-5