

Uniwersytet Warszawski
Wydział Prawa i Administracji

Michał Starczewski
Nr albumu: 214788

Text and data mining a wolność badań naukowych w świetle prawa autorskiego

Praca magisterska
na kierunku prawo

Praca wykonana pod kierunkiem
prof. Jana Bleszyńskiego
Instytut Prawa Cywilnego
WPiA UW

Warszawa, czerwiec 2016

Oświadczenie kierującego pracą

Oświadczam, że niniejsza praca została przygotowana pod moim kierunkiem i stwierdzam, że spełnia ona warunki do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data

Podpis kierującego pracą

Oświadczenie autora pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przez mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data

Podpis autora (autorów) pracy

Streszczenie

Coraz powszechniej w badaniach naukowych stosuje się metody text and data mining (TDM). Rodzi to pytania o granice legalności tych działań. O ile w państwach Unii Europejskiej trwa dyskusja na ten temat i panuje przekonanie o dużej niepewności prawnej, która sprawia, że badania europejskie TDM zaczynają odstawać od amerykańskich i azjatyckich, o tyle temat ten w polskiej doktrynie pojawił się tylko incydentalnie. W pracy przeanalizowane zostały przepisy polskiego prawa autorskiego w zakresie dopuszczalności stosowania metod TDM, w szczególności przepisy o dozwolonym użytku i ochronie integralności utworów. Ze względu na klauzule generalne zderza się zasada ochrony praw twórcy z zasadą wolności badań naukowych.

Słowa kluczowe

prawo autorskie, dozwolony użytek, integralność, text-and-data-mining, big data
<Wykaz maksymalnie 10 słów swobodnie wybranych>

Dziedzina pracy (kody wg programu Socrates-Erasmus)

10.3 prawo cywilne

{ *(opcjonalnie, klasyfikator zależny od dziedziny)*

Klasyfikacja tematyczna

083900

Tytuł pracy w języku angielskim

TDM vs research freedom from copyright perspective

Spis treści

Oświadczenie kierującego pracą.....	2
Oświadczenie autora pracy.....	2
Wykaz skrótów.....	5
Wstęp.....	6
Metody Text and Data Mining.....	9
Pozaautorskie zagadnienia prawne związane z TDM w badaniach naukowych.....	20
Kwestia dopuszczalności metod TDM w świetle prawa UE.....	22
Stosowanie metod TDM na podstawie przepisów o dozwolonym użytku.....	29
Dozwolony użytek a analiza baz danych o cechach utworu.....	46
Integralność utworów analizowanych metodami TDM.....	50
Ochrona prawnoautorska efektów TDM.....	61
Zasada wolności badań naukowych.....	67
Granice legalnego stosowania TDM. Podsumowanie.....	72
Bibliografia.....	76

Wykaz skrótów

CC - Creative Commons

Dyrektywa Infosoc - Dyrektywa 2001/29/WE Parlamentu Europejskiego i Rady z dnia 22 maja 2001 r. w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym

KC - kodeks cywilny

Konst. - Konstytucja RP z 1997 r.

PrAut – ustawa z dn. 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych

u.o.b.d. - ustawa o ochronie baz danych

u.s.o. - ustawa o systemie oświaty z 1991 r.

u.z.f.n. - ustawa o zasadach finansowania nauki z 30 IV 2010 r.

TDM – *text and data mining*

Wstęp

Migrena jest chorobą tajemniczą. Chorzy cierpią na dokuczliwe bóle głowy, a naukowcy nie znajdują przyczyn. W 1986 r. Don Swanson opublikował wyniki badań, które były zaprzeczeniem obowiązujących metod uzyskiwania nowej wiedzy medycznej. Nie badał pacjentów, ale wykorzystał komputery do tego, by maszynowo przeanalizować abstrakty artykułów znajdujące się w bazie MEDLINE. Zauważył, że migreny bardzo często są związane z niedoborem magnezu. Wcześniej nikt nie podejrzewał takiego związku. Swanson odnalazł relacje między faktami ukryte w ilości danych przerastających percepcję indywidualnych osób. Hipoteza Swansona została potwierdzona po kilku latach w badaniach klinicznych. Choć nie rozwiązało to całkowicie zagadki migren, to uczyniono duży krok, by ulżyć wielu cierpiącym. Istota zastosowanej metody polegała na objęciu maszynową analizą ogromnej ilości danych.

Dzisiaj metody maszynowej analizy tekstu i danych (*text-and-data mining*, TDM) upowszechniły się i znajdują zastosowanie w bardzo zróżnicowanych sytuacjach, w różnych obszarach wiedzy, także poza nauką. Nie co dzień zdarzają się równie spektakularne odkrycia, jak to, którego dokonał Swanson. Metody te są wykorzystywane także do bardziej prozaicznych czynności, jak np. analiza przyrastającej coraz szybciej literatury naukowej w celu jej selekcji, a nie zdobycia nowej wiedzy.

W miarę, jak nowe metody zdobywały popularność, prawnicy zaczęli zadawać pytania o legalność działań naukowców. Czy wystarczy, że badacz uzyskał legalny dostęp do utworów, by mógł legalnie analizować je metodami TDM? A może prawo stawia jakieś granice tej działalności w ramach dozwolonego użytku? Koncerny wydawnicze twierdziły, że analiza metodami maszynowymi nie mieści się w zakresie standardowych licencji i wymaga uzyskania od nich dodatkowej zgody.

Kwestie te zaczęto podnosić na forach coraz większej liczby państw. Zauważono paradoks: prawo autorskie nie chroni faktów, których opracowanie jest na ogół przedmiotem TDM, ale w celu uzyskania tych faktów metodami TDM należy wykonać działania, które są uznawane za naruszenie monopolu autorskiego (por. rozdział 3 o kontekście europejskim). Niektóre kraje, jak Japonia i Wielka Brytania, dostosowały swoje ustawy autorskie tak, by zagwarantować możliwość prowadzenia badań naukowych, przynajmniej niekomercyjnych, metodami TDM. Obecnie trwa dyskusja w Unii Europejskiej, która zmierza prawdopodobnie w kierunku wprowadzenia nowego wyjątku od monopolu autorskiego. Temat ten w Polsce nie znalazł dotychczas większego oddźwięku. Poza nielicznymi pracami nie został jeszcze zauważony.

Celem niniejszej pracy jest odpowiedź na pytanie (sformułowane w ślad za analogicznymi

pytaniem stawianym na forum europejskim), czy stosowanie metod TDM ma podstawy w obecnie obowiązujących przepisach polskiego prawa autorskiego? Jakie są granice legalnego stosowania TDM przede wszystkim w celach naukowych w naszym kraju? Okazuje się, że nie można udzielić jednoznacznej odpowiedzi na tak postawione pytanie. Sposób, w jaki często myśli się o kwestiach prawnych związanych z TDM, zbyt upraszcza i ujednolica zróżnicowane metody analizy.

Niniejsza praca dotyczy stanu prawnego po nowelizacjach PrAut z jesieni 2015 r., które istotnie zmieniły część przepisów o dozwolonym użytku. Choć kilkakrotnie nawiązuję do stanu sprzed tej nowelizacji, to nie przeprowadzam gruntownej analizy wpływu zmian na możliwość stosowania metod TDM do celów naukowych.

W rozdziale 1 wyjaśniam, czym są metody TDM. Jest to ważny, pierwszy krok w kierunku prawnej analizy dopuszczalności tych metod. Zaczynam od przeglądu zastosowań w różnych obszarach wiedzy. Metody TDM okazują się na tyle zróżnicowane, że ograniczanie się w analizach prawnych do stosowania zbiorczego określenia TDM wydaje się niewystarczające. Rozdział 2 zawiera skrótowy przegląd zagadnień prawnych związanych z metodami TDM, a które nie dotyczą prawa autorskiego. Ze względu na zakres tematyczny niniejszej pracy nie będę już wracał do tych wątków. Z kolei rozdział 3 koncentruje się na europejskiej dyskusji o prawnych aspektach TDM. Dyskusja zmierza w kierunku wprowadzenia nowego wyjątku od bezwzględnych praw autorskich.

Kolejne cztery rozdziały stanowią najważniejszą część pracy. W Rozdziale 4 przeanalizowałem przepisy o dozwolonym użytku pod kątem, czy mogą stanowić podstawę dla metod TDM, do jakich metod i pod jakimi warunkami? Najważniejszy wydaje się wyjątek dla instytucji naukowych w celu badań naukowych (art. 27). Zostały przeanalizowane także inne przepisy mogące stanowić podstawę do stosowania TDM w ramach dozwolonego użytku. Rozdział 5 dotyczy analizowania baz danych spełniających cechy utworu. W rozdziale 6 odpowiadam na pytanie o ryzyko naruszenia prawa do integralności utworów analizowanych. Ryzyko takie powstaje zwłaszcza, gdy badacze „czyszczą” dane przed analizą, „kaleczą” utwór poprzez ujednolicanie i usuwanie tzw. szumu. Rozdział 7 został poświęcony kwestii ewentualnej prawnoautorskiej ochronie produktów analizy TDM. Wreszcie, rozdział 8 zawiera analizę zasady wolności badań naukowych, która jest przeciwwagą w sytuacjach, gdy granice legalnego stosowania metod TDM wyznaczają klauzule generalne, zmuszające do wagi interesów. Szczególnej wagi nabierają te rozstrzygnięcia przy stosowaniu art. 35 PrAut. Ostatni rozdział zawiera wnioski i podsumowanie.

TDM w badaniach naukowych nieustannie nabiera coraz większego znaczenia. Niepewność prawna spowalnia rozwój tych metod. Stwarza bariery, zwłaszcza dla podmiotów nie dysponujących znaczącym zapleczem (młodzi badacze, start-up'y działające na styku nauki i biznesu), nawet gdy nie ma przeszkód technicznych.

Ostateczne rozstrzygnięcia, co do dopuszczalnych granic TDM wymagają przemyślanych decyzji. Nie powinny być one przypadkowe. Należy cofnąć się o krok i spojrzeć na przepisy z szerszej perspektywy. Jakie konkretnie dobra uważamy za warte ochrony? Jakie rozwiązania sprawią, że twórcy będą cieszyć się nadal wystarczającymi zachętami do tworzenia, a zarazem pozwolą społeczeństwu uzyskać jak najwięcej korzyści wynikających z działalności naukowców? Być może rozwiązanie jest tak proste, choć nieoczywiste, jak prostym, choć nieoczywistym rozwiązaniem było wskazanie w 1986 r. przez D. Swansona niedoboru magnezu jako jednej z przyczyn migreny.

Metody Text and Data Mining

*Wiedza może być publiczna, a zarazem wciąż nie odkryta, jeśli niezależnie stworzone jej fragmenty, logicznie ze sobą powiązane, nie zostały jeszcze ujawnione, połączone i zinterpretowane*¹ - tak rozpoczyna się jeden z najbardziej znanych artykułów o analizie dużych zbiorów tekstów w celu zdobycia nowej wiedzy. Jego autor, Don R. Swanson, opierając się wyłącznie na bazie abstraktów artykułów medycznych MEDLINE, postawił w 1986 r. hipotezę o związku występowania migren z niedoborem magnezu, potwierdzoną po trzech latach eksperymentalnie. Miał na swoim koncie również inne sukcesy w stosowaniu tej metody.

W czasie, gdy Swanson pisał to zdanie, pojęcia text-miningu i data-miningu (*Text-and-Data-Mining*, w skrócie TDM) nie były jeszcze popularne², a sama idea pozyskiwania zupełnie nowej wiedzy poprzez samą analizę statystyczną dużej liczby publikacji, była nowatorska i przyjmowana ze sceptycyzmem. Od połowy lat 1990-ych gwałtownie wzrosła popularność tych terminów. Moda na stosowanie pewnych słów i określeń często przemija. Nawet jeśli jednak spotka ten los także pojęcie TDM, to metody analizy określane dziś tym mianem mają już ugruntowaną pozycję. W niektórych dyscyplinach nauki metody zaliczane dziś do TDM są używane od kilku dziesięcioleci. Przedstawiciele innych dyscyplin dopiero uczą się nimi posługiwać.

Do dyscyplin, których przedstawiciele od dawna wykorzystują maszynową analizę tekstu i danych, należą przede wszystkim nauki biologiczne i medyczne. Przegląd metod wykorzystujących text mining w biologii molekularnej i biomedycynie od lat 1980-ych do 2004 r. pokazuje rozwój takich metod jak identyfikacja obiektów (*tagging biological entities*), wydobywanie informacji faktograficznych, czy nawet zdobywanie nowej wiedzy o białkach i ich interakcjach³. Po metody TDM sięgają naukowcy z kolejnych specjalności. Przykładowo: w celu doskonalszego rozpoznania chorób układu oddechowego analizuje się maszynowo nie tylko dane osobowe pacjentów, ich diagnostykę, wyniki badań laboratoryjnych, ale nawet szczegóły dotyczące historii kontaktów pacjentów ze służbą zdrowia⁴. *Te narzędzia, które już odgrywają znaczącą rolę w innych obszarach*

1

Knowledge can be public, yet undiscovered, if independently created fragments are logically related but never retrieved, brought together, and interpreted, D. Swanson, *Undiscovered public knowledge*, The Library Quarterly, 1986/56(2), s. 103.

² Por. https://books.google.com/ngrams/graph?content=text+mining%2Cdata+mining&year_start=1960&year_end=2008&corpus=15&smoothing=0&share=&direct_url=t1%3B%2Ctext%20mining%3B%2Cc0%3B.t1%3B%2Cdata%20mining%3B%2Cc0.

³ M. Krallinger, R.A. Erhardt, A. Valencia, *Text Mining approaches in molecular biology and biomedicine*, "Drug Discovery Today: Biosilico.", 2005, 10, s. 439-445. Wykres 1 przedstawia propozycję usystematyzowania stosowania metod TDM w biomedycynie w okresie od poprzedzającego lata 1990 do 2004 r.

⁴ D. Piedra, A. Ferrer, Joaquim Gea, *Text Mining and Medicine: Usefulness in Respiratory Diseases*, Arch

biomedycyny, zaczęły być wykorzystywane w ostatnim czasie także w medycynie układu oddechowego. Najbardziej oczywistym obszarem zastosowań medycznych text miningu (zarówno w badaniach, jak i w zastosowaniach klinicznych) są integracja oraz przeniesienie wyników uzyskanych w naukach podstawowych, a także lepsze zrozumienie procesów diagnostycznych, klasyfikacja i prognozowanie schorzeń. Text mining może być przydatny także w tworzeniu modeli predykcyjnych, opracowywaniu inteligentnych systemów alarmowych i wspieraniu lekarzy w podejmowaniu decyzji⁵.

Zastosowanie metod TDM wykracza daleko poza nauki biologiczne i medyczne. W naukach społecznych podejmowane są m.in. starania, by zbadać, jakie znaczenie mają w rzeczywistości relacje na portalach społecznościowych. Portale społecznościowe są wdzięcznym przedmiotem badań zarówno dlatego, że odgrywają coraz większą rolę w życiu ludzi, ale i z powodu łatwego dostępu do dużych ilości danych, umieszczanych w nich przez użytkowników. Najczęściej wykorzystywane są dane z Twittera i Facebooka. Zaproponowano także metodę przewidywania siły więzi społecznych w portalu LinkedIn⁶.

Jednym z trudniejszych zadań jest automatyczne rozpoznawanie ironii. Jest to zaś potrzebne do rozpoznawania nastrojów społecznych, zarówno jeśli chodzi o wybory konsumenckie, jak i o preferencje polityczne. Z porównania metod nadzorowanych (*supervised*) i częściowo nadzorowanych (*semi-supervised*) analiz nastrojów politycznych wśród greckich użytkowników serwisu Twitter wynika, że jedne i drugie metody wykazują podobną skuteczność⁷.

W bibliotekoznawstwie i nauce o informacji metody TDM wykorzystuje się od końca lat 1990-ych. Dość późno, biorąc pod uwagę, że dziedzina ta tradycyjnie zajmuje się analizą danych. W 2003 r. zostało ukute określenie *bibliomining*, sugerujące stosowanie metod TDM w bibliometrii⁸.

Wreszcie beneficjentami stosowania metod TDM staje się także środowisko prawnicze. Zostawiając na boku komercyjne bazy, które udoskonalają narzędzia służące przeszukiwaniu zgromadzonych zasobów, prowadzone są również badania mające dostarczyć nowej wiedzy. Przykładem jest analiza za pomocą metod TDM australijskiego orzecznictwa w sprawach

Bronconeumol. 2014;50(3):113.

⁵ These tools, which already play a significant role in other areas of biomedical knowledge, have recently begun to be used in respiratory medicine. The most obvious areas for the medical application of text mining (both in research and in the clinic) are the integration and transfer of advances made in the most basic sciences, and a better understanding of the diagnostic processes, severity classifications and determination of disease prognosis. Text mining may also be useful for generating predictive outcomes models, creating intelligent alert systems and supporting the clinician in the decision-making process, D. Piedra, A. Ferrer, Joaquim Gea, *Text Mining and Medicine: Usefulness in Respiratory Diseases*, Arch Bronconeumol. 2014;50(3), s. 119.

⁶ M.K. Sohrabi, S. Akbari, *A comprehensive study on the effects of using data mining techniques to predict tie strength*, "Computer in Human Behavior" 60 (2016), s. 534-541.

⁷ B. Charalampakis et al., *A comparison between semi-supervised and supervised text mining techniques on detecting irony in greek political tweets*, "Engineering Applications of Artificial Intelligence, 51(2016), s. 50-57.

⁸ L. Siguenza-Guzman et al., *Literature Review of Data Mining Applications in Academic Libraries*, "The Journal of Academic Librarianship" 41 (2015), s. 499-510.

cywilnych dotyczącego ochrony znaków towarowych⁹. Metoda grupowania pozwoliła ujawnić ewolucję orzecznictwa w latach 2004-2013 (por. wykres 9, s. 796). Z kolei hiszpańscy badacze zaproponowali metodę segmentacji aktów prawnych w oparciu o słownik pojęć¹⁰.

Popularność metod TDM wiąże się z zauważeniem rosnącego znaczenia, jakie w środowisku cyfrowym ma uzyskiwanie w sposób maszynowy nowej wiedzy z tekstów i danych. Często pojęcie TDM występuje w sąsiedztwie innego popularnego terminu, czyli „big data”. Big data jest używane na określenie wielkich zbiorów danych, do przechowywania których potrzebna jest specjalna infrastruktura i niezwykle pojemne serwery. Jednym ze sztandarowych przykładów *big data* są dane uzyskiwane przez zderzacz hadronów w CERN, których przybywa w tempie ok. 30 petabajtów każdego roku¹¹. O *big data* można mówić także w sytuacji, gdy zbioru danych nie trzeba mierzyć w petabajtach, ale gdy są one wystarczająco duże, by nie dało się ich analizować inaczej, niż za pomocą technik maszynowych. Humorystycznie określa się je czasem jako *big enough data*, czyli dane wystarczająco duże, by przekroczyć możliwości tradycyjnej analizy. To jest przypadek np. korpusów językowych. W humanistyce maszynowa analiza dużych zbiorów danych jest określana mianem *distant reading* (zwrot zaproponowany przez Franco Morettiego¹²), w przeciwieństwie do *close reading*, czyli tradycyjnych metod dogłębnej analizy pojedynczych tekstów¹³.

Metody, które stosuje się do analizy wielkich zbiorów danych, określane są jako Text and Data Mining, (choć, jak często bywa w przypadku modnych pojęć, nie ma pełnej zgody co do nazw i ich znaczeń). Jako tłumaczenia na język polski stosuje się najczęściej zwrot „maszynowa analiza danych i tekstu”, który nie oddaje uroku angielskiej metafory, przywodzącej na myśl prace górnicze w kopalnianych szybach. Angielska metafora sugestywnie wskazuje na podobieństwo poszukiwania wartościowych informacji w spowitej ciemnościami masie (ale – z drugiej strony – sugeruje wchodzenie coraz bardziej w głąb badanej materii, podczas gdy wiele metod TDM „pozostaje na powierzchni” tekstu, obrazów czy innych badanych obiektów¹⁴).

Big data nie da się analizować inaczej niż za pomocą metod TDM. Zależność ta nie działa jednak w drugą stronę. Niektóre metody TDM można z powodzeniem stosować do względnie

⁹ Ch. Trappey, A. Trapey, *Collective intelligence applied to legal e-discovery: A ten-year case study of Australia franchise and trademark litigation*, „Advanced Engineering Informatics” 29 (2015), s. 787-798.

¹⁰ I. Hasan, J. Parapar, R. Blanco, *Segmentation of legislative documents using a domain-specific lexicon*, „19th International Conference on Database and Expert Systems Application”.

¹¹ Por. <http://home.cern/about/computing>, a także „Dzielenie się danymi to tylko kolejna składowa komunikacji naukowej” – z Timem Smithem (CERN) rozmawia Maciej Chojnowski, <http://otwartanauka.pl/analysis/rozmowy/dzielenie-sie-danymi-to-tylko-kolejna-skladowa-komunikacji-naukowej>.

¹² Por. też tegoż autora esej poświęcony metodologii historii literatury przechodzącej od tekstów do modeli, F. Moretti, *Wykresy, mapy, drzewa. Abstrakcyjne modele na potrzeby historii literatury*, Kraków 2016.

¹³ Por. W. McCarty, *A Telescope for a mind*, w: „Debates in the Digital Humanities”, red. M. Gold, 2012, <http://dhdebates.gc.cuny.edu/debates/text/37>.

¹⁴ *Study on the legal framework of TDM...*, s. 10.

niewielkich zbiorów.

Choć określenie TDM kryje w sobie wyraźne rozdzielenie tekstu od danych, to należy pamiętać, że każdy tekst jest rodzajem danych. Cechą charakterystyczną tekstu jest zazwyczaj jego nieustrukturyzowany charakter, o czym będzie mowa jeszcze niżej. Tekst nie jest jednak jedynym rodzajem danych, który nie jest ustrukturyzowany. Zarówno nagrania audio, jak i obrazy również mogą mieć taki charakter. Problemy prawne są podobne, niezależnie od rodzaju przedmiotu badań: *Chociaż inny jest przedmiot, analizy tekstu, wideo, obrazów i zdjęć wiążą się z zasadniczo takimi samymi zagadnieniami prawnymi z obszaru prawa autorskiego i ochrony baz danych*¹⁵. Dane badawcze są definiowane bardzo szeroko. Przykładowa definicja to „zarejestrowane materiały o charakterze faktograficznym powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych”¹⁶. Taka definicja obejmuje nie tylko pomiary laboratoryjne, ale także tak różnorodne materiały, jak notatki terenowe z badań antropologicznych lub ankiety z badań socjologicznych.

Zwrot TDM może sugerować, że jego desygnatem jest jakaś jedna metoda analizy. W wielu analizach prawniczych wydaje się być w taki sposób używany. Nic bardziej mylnego. TDM jest to zbiorcze określenie wielu zróżnicowanych technik przetwarzania danych, mających na celu uchwycenie relacji między poszczególnymi elementami, przypisanie im odpowiednich właściwości, zhierarchizowanie elementów lub ich większych całości, ustalenie relacji lub odniesienie ich do jakichś zewnętrznych punktów odniesienia. Ostatecznym celem jest uchwycenie pewnych wzorów, zależności lub schematów. Wspólnym mianownikiem dla tych wszystkich technik jest ich funkcja: analiza danych w sposób maszynowy. Cechą charakterystyczną metod TDM jest skoncentrowanie na wykrywaniu schematów (*patterns*). Schematy wykrywa się za pomocą specjalnych algorytmów.

Poniżej pokrótce przedstawię kilka reprezentatywnych metod TDM, pomijając aspekty czysto techniczne, a koncentrując się na tych cechach, które mogą mieć wpływ na sytuację prawną i sferę uprawnień podmiotu przeprowadzającego analizy tymi metodami.

Text Mining – przegląd metod

Cechą tekstu jest brak ustrukturyzowania lub ustrukturyzowanie tylko częściowe. Język naturalny jest wieloznaczny. Pierwszym krokiem i podstawą do prowadzenia dalszych badań jest zatem ustrukturyzowanie tekstu. Polega to na precyzyjnym określeniu zależności między elementami analizowanych tekstów (np. pojedynczymi słowami) lub między elementami tekstów a zewnętrznym odniesieniem. Wyniki analizy materiałów jednego rodzaju wykorzystuje się do analizy materiału innego rodzaju, np. analiza tekstu pomaga analizować maszynowo grafikę, czy

¹⁵ *Although the objects are different, the analysis of texts, videos, images and pictures present basically the same legal issues in terms of copyright and database protection, Study on the legal framework of TDM...*, s. 8.

¹⁶ *Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings*, https://www.whitehouse.gov/omb/fedreg_a110-finalnotice.

nagrania audio lub wideo¹⁷.

Metody TDM dzieli się m.in. według kryterium zaangażowania człowieka w proces analizy. Metody, w których wykorzystywana jest ręczna praca człowieka, np. w celu korekty wyników lub na etapie przygotowania kodowania wzorcowej porcji materiału, który służy jako odniesienie we właściwej analizie (*training material*) określa się mianem „nadzorowanych” (*supervised*). Stopień ingerencji człowieka może być różny (stąd mówi się także o metodach częściowo nadzorowanych - *semi-supervised*). Z kolei metody zakładające brak bezpośredniej ingerencji człowieka w finalny efekt, to metody nienadzorowane (*unsupervised*), w których w zamian za mniejszy nakład pracy akceptuje się większy margines błędu. Zależnie od stosowanej techniki, akceptowalny margines błędu może wynosić od 70% do powyżej 90%.

Inne kryterium podziału metod *text miningu* to podział ze względu na podejście do elementów analizowanego materiału. Gdy każdy wyraz traktuje się jako oddzielną, niezależną jednostkę, mówimy o „worku słów” (*Bag of Words*, BOG). Alternatywnie analizuje się większe całości, z reguły pojedyncze zdania. Te całości stają się podstawowymi jednostkami podlegającymi analizie i noszą miano strun (*String of Words*), wyodrębnionych w postaci założeń danych badań, wyrażonych w ich algorytmie. W obu przypadkach w celu ustalenia lokalizacji danej jednostki (pojedynczego słowa lub struny) najczęściej trzeba posłużyć się większym fragmentem.

Dobrym przewodnikiem po technikach Text Miningu jest książka *Mining Text Data*¹⁸. Prezentując wybrane metody TDM, będę opierał się na tej publikacji. Przykłady zastosowań metod mogą obejmować obszary wykraczające poza badania naukowe (które są przedmiotem mojej pracy). Choć koncentrują się one na analizie tekstu, to na poziomie ogólności potrzebnym dla mojej pracy, techniki analizy innego rodzaju danych znajdują wiele analogii. Poszczególne metody TDM nie są rozłączne. Wręcz przeciwnie, często zastosowanie jednych metod analizy wymaga uprzedniego zastosowania innych.

Przed zastosowaniem większości technik maszynowej analizy potrzebne jest przygotowanie dokumentów i oczyszczenie ich z tzw. szumu, czyli słów i znaków, które zakłócają lub utrudniają analizę. Usuwa się słowa przeszkadzające (*stop-words*) i zastępuje różne formy gramatyczne tego samego słowa jedną formą. W przypadku technik wykorzystujących struny słów (*strings*) czasem usuwa się sufiksy i prefiksy. Można usunąć odmianę wyrazów przez czasy i przypadki, a także pozostawić tylko liczbę pojedynczą. Zakres takiego przygotowania zależy od konkretnego przypadku¹⁹.

Wydobywanie informacji (*information extraction*).

¹⁷ *Mining Text Data*, red. Charu C. Aggarwal, ChengXiang Zhai, Springer 2012 s. 239 i n.

¹⁸ *Mining Text Data*, Red. Charu C. Aggarwal, ChengXiang Zhai, Springer 2012.

¹⁹ Por. np. s. 106, 167.

*Podstawowy cel wydobywania informacji polega na odkryciu struktury informacji w nieustrukturyzowanym lub częściowo ustrukturyzowanym tekście*²⁰. Od zwykłego wyszukiwania odróżnia tę metodą to, że wymaga ona poradzenia sobie z synonimami i pojęciami niejednoznacznymi. Badacz rozpoczyna pracę od przygotowania zestawu potrzebnych informacji, których będzie szukał. Algorytm sprawdza więc, czy w danym tekście znajdują się odpowiedzi na z góry postawione pytania. Wśród wielu poszukiwanych informacji, można wymienić: wyszukanie odkryć dotyczących poszczególnych genów lub białek w obfitej literaturze z obszaru nauk medycznych lub biologicznych (zadanie nietrywialne z powodu synonimów i wieloznaczności); informacje o działalności przedsiębiorstw w określonym czasie w celu analiz finansowych; odpowiedzi na zapytania w wyszukiwarkach internetowych²¹.

Podstawowe techniki wydobywania informacji zakładają uprzednią znajomość struktury tekstów, w których szuka się informacji²², należą więc one do kategorii metod nadzorowanych (*supervised*).

Dwie podstawowe czynności wykonywane w ramach tej metody, to rozpoznawanie nazw jednostek (*Named Entity Recognition*, NER) oraz rozpoznawanie relacji między tymi jednostkami. NER polega na przypisaniu nazw do ustalonych uprzednio klas, takich jak „osoba”, „organizacja”, czy „lokalizacja”. NER jest stosowany jako niezbędny wstęp do wielu dalszych analiz. Rozpoznawanie relacji ma na celu odtworzenie semantycznych relacji między jednostkami (np. Jan Kowalski jest profesorem afiliowanym przy uniwersytecie X). Odtworzenie relacji semantycznych często wymaga sięgnięcia do zewnętrznych zasobów, gdzie znajdują się informacje potrzebne do przypisania cech analizowanym obiektom (w podanym wyżej przykładzie zewnętrzna informacja może dotyczyć faktu, że X jest uniwersytetem). Popularnym zewnętrznym źródłem informacji jest Wikipedia²³.

Kontrola skuteczności wydobywania informacji wymaga ręcznego anotowania (opracowanie warstwy informacyjnej) części dokumentów z tego samego korpusu w celu sprawdzenia stopnia zgodności wyników uzyskanych ręcznie z tymi uzyskanymi automatycznie.

Podsumowywanie (*text summarization*)²⁴.

Technika podsumowywania polega na odnajdywaniu kluczowych informacji lub najważniejszych zdań z korpusu tekstów. W efekcie otrzymuje się np. tekst długości jednego akapitu lub wykres przedstawiający znaczenie poszczególnych zdań według określonych kryteriów.

²⁰ *The general goal of information extraction is to discover structured information from unstructured or semi-structured text*, Jing Jiang, *Information extraction from Text*, w: „Mining Text Data”, ..., s. 11.

²¹ *Mining Text Data*, s. 11.

²² Tamże s. 14.

²³ Tamże, s. 24.

²⁴ Ani Nenkova, Kathleen McKeown, *A Survey of Text Summarization*, w: „Mining Text Data”, s. 43-76.

Osiąga się to poprzez ocenę znaczenia poszczególnych zdań i poruszanych w nich tematów. Tematy są identyfikowane np. przy użyciu bazy leksykalnej języka angielskiego WordNet, udostępnianej przez Uniwersytet Princeton²⁵.

Szczególnym przypadkiem tej techniki jest podsumowywanie artykułów naukowych. Charakterystyczne jest tu wykorzystywanie cytowań i zakresu wykorzystania artykułu w innych artykułach, które powołują się na niego. W badaniach społecznych mogą znaleźć zastosowanie metody podsumowywania dużych korpusów e-maili, takich jak korespondencja elektroniczna administracji rządowej²⁶ lub spuścizny osób prywatnych²⁷.

Algorytmy grupujące (*clustering algorithms*)²⁸

Grupowanie, klasyfikowanie, filtrowanie, indeksowanie – to są podstawowe problemy, z jakimi należy zmierzyć się przy analizie jakichkolwiek danych. *Grupowaniem nazywa się odnajdywanie w danych grup podobnych obiektów. Podobieństwo między obiektami mierzy się za pomocą funkcji podobieństwa*²⁹. Technika grupowania znajduje zastosowanie w organizacji gromadzonych danych i ich przeglądaniu, podsumowywaniu, klasyfikacji. Metody grupowania są różnorodne i są dostosowane do przedmiotu analizy. Niektóre bazują na założonych schematach lub częstotliwości występowania elementów, inne wykorzystują metody bazujące na prawdopodobieństwie.

Wstępna praca, niezbędna do prawidłowego przeprowadzenia analizy, polega na wyczyszczeniu danych z szumu, czyli słów czy znaków, które zakłócają prawidłowe wyniki. Z korpusu usuwa się słowa występujące zbyt często, a które nie niosą ze sobą istotnej treści (w języku angielskim są to takie wyrazy jak „a”, „the”, „of”, itd.). Są to tzw. *stop-words*. Zazwyczaj lista takich *stop-words* obejmuje 300-400 słów³⁰. Czasem stosuje się metody nadzorowane (*supervised*), kiedy człowiek ręcznie usuwa szum w materiale, który służy jako wzór dla komputera do czyszczenia zasadniczej części zbioru.

Ujednoznacznienie (*dimensionality reduction*)

Jednym z trudniejszych zadań, a zarazem należącym do najważniejszych w metodach analizy tekstu, jest ujednoznacznienie wyrażen użytych w tekście. Żywy człowiek znający język, w

²⁵ <https://wordnet.princeton.edu/>

²⁶ Głośnym echem odbiła się w USA sprawa „zgubienia” przez administrację G. W. Busha 22 mln e-maili z lat 2003-2005, a odzyskanych w 2009 r. Por. *Millions of Bush administration e-mails recovered*, <http://edition.cnn.com/2009/POLITICS/12/14/white.house.emails>.

²⁷ Por. S. Hangal i inni, *Historical Research Using Email Archives*, Extended Abstracts of the ACM CHI'15 Conference on Human Factors in Computing Systems, <http://mobisocial.stanford.edu/papers/chi2015.pdf>.

²⁸ Charu C. Aggarwal, ChengXiang Zhai, (...) s. 77-128.

²⁹ *The clustering problem is defined to be that of finding groups of similar objects in the data. The similarity between the objects is measured with the use of a similarity function*, Tamże, s. 77-78.

³⁰ Tamże, s. 81.

którym powstał tekst, z łatwością i niejako bezrefleksyjnie wie, jaki desygnat ma w danym kontekście np. słowo „zamek” (np. odróżnia zamek z piasku od zamka błyskawicznego i zamka w drzwiach). Sprawienie, by desygnaty wyrazów wieloznacznych stały się „zrozumiałe” dla komputera, wymaga zaawansowanej analizy semantycznej tekstu. Ujednoznacznienie polega także na rozpoznawaniu synonimów. Ten sam desygnat może być przecież określany na wiele sposobów. *Ujednoznacznianie może polegać na łączeniu pojęć, które mają tę samą semnatykę, by zidentyfikować i oddzielić pojęcia mające wiele znaczeń*³¹.

Metoda ujednoznacznienia bywa wspomagana innymi metodami, takimi jak grupowanie i identyfikowanie tematów. Te trzy metody są ze sobą ściśle związane. Modelowanie tematów jest trudne dlatego, że tylko w wyjątkowych przypadkach w samym tekście pojawia się słowo określające temat w najbardziej precyzyjny sposób. Przypomina to grę w tabu (Taboo®), w której gracze muszą odgadnąć słowo, przy czym zabronione jest opisywanie znaczenia za pomocą słów o zbyt bliskich znaczeniu³². Powiązania (linki) między dokumentami niosą duży bagaż informacji, wykorzystywany do ujednoznacznienia i określania tematów. Ze znacznym prawdopodobieństwem można założyć, że to samo pojęcie w dwóch powiązanych ze sobą dokumentach używane jest w tym samym znaczeniu.

Spodziewanym efektem tych operacji nie musi być materiał łatwy do analizy, zwłaszcza nie musi być on w ogóle zrozumiały dla człowieka. Ma on stanowić podstawę dalszych analiz wykorzystujących kolejne metody maszynowej analizy tekstu.

Analiza międzyjęzykowa (*translingual mining*)

Jedną z metod TDM specyficzną dla tekstów są automatyczne tłumaczenia z jednego języka na drugi. Tłumaczenie tekstu jest jednak tylko prototypiczną formą różnych metod transferu wiedzy między językami. Międzyjęzykowe wyszukiwanie informacji (CLIR – *cross-language information retrieval*) pozwala np. na wydobycie informacji z dokumentów anglojęzycznych poprzez zadanie pytania w języku chińskim.

*Wprawdzie dodatkowa wiedza translatoryjna może być niezbędna do przeprowadzenia porównania dokumentów i znaczenia zapytania w obu językach, ale należy odróżnić cel informacyjny CLIR od celu pełnego tłumaczenia maszynowego*³³. W obecnym stanie rozwoju techniki automatycznych tłumaczeń stosuje się trzy kategorie metod: modele bazujące na

³¹ *Dimension reduction can collapse together terms that have the same semantics, to identify and disambiguate terms with multiple meanings*, Steven Crain, Ke Zhou i in., „Dimensionality reduction and topic modeling: from latent semantic indexing to latent dirichlet allocation and beyond”, w: *Mining Text Data*, s. 130.

³² Tamże, s. 139.

³³ *Although additional translational knowledge may need to be brought to bear in order to compare the returned documents and the query in two languages, the information goal of CLIR is distinct from that of full text M[achine] T[ranslation]*, Jian-Yun Nie, Jianfeng Gao, Guihong Cao, „Translingual mining from text data”, w: *Mining Text Data*, s. 324.

pojedynczych słowach, modele bazujące na frazach oraz modele syntaktyczne³⁴. Te ostatnie pozwalają na formułowanie coraz bardziej poprawnych zdań w języku, na który dokonywane jest tłumaczenie. Poprawność językowa tłumaczeń w pozostałych modelach pozostawia wiele do życzenia. Kluczowe dla poprawności automatycznych tłumaczeń jest wykorzystywanie tekstów paralelnych, umożliwiających doskonalenie algorytmów.

Analiza opinii (*Opinion Mining and Sentiment Analysis*)

Metodą TDM szeroko stosowaną np. w badaniach marketingowych, ale znajdującą szerokie zastosowanie również w badaniach naukowych, jest analiza sentymentów i opinii. Analiza maszynowa powinna cechować się obiektywnością, w przeciwieństwie do analiz dokonywanych przez człowieka, które są w nieunikniony sposób skażone preferencjami badacza, zwłaszcza jego subiektywizmem. Opinie mogą być wyrażane na dowolny temat. Mogą dotyczyć jakiegoś przedmiotu jako całości lub tylko pojedynczych jego cech (funkcji). Mogą być pozytywne, negatywne lub neutralne z różnym stopniem natężenia. Przekształcenie nieustrukturyzowanego tekstu w ustrukturyzowane dane obejmuje w tym przypadku pięć kroków: identyfikację jednostek wyrażających opinie i ich pogrupowanie, ekstrakcję aspektów podlegających ocenie, przypisanie opinii do danego źródła i czasu, klasyfikację opinii i wreszcie uogólnienie wyników. Analiza ta opiera się na wykorzystaniu słowników zawierających słowa wyrażające opinie i sentymenty (w tym idiomy, które są źródłem wielu pomyłek).

Analiza danych biomedycznych (*Biomedical Text Mining*)

Metody TDM są intensywnie rozwijane w obszarze nauk biomedycznych.

*Zadania obejmują rozpoznawanie wyrażonych explicite faktów z biomedycznej literatury, odkrycie faktów wcześniej nieznanych lub wyrażonych tylko implicite, podsumowywanie dokumentów i odpowiadanie na pytania*³⁵. Z powodu wysokiej specjalizacji poszukiwanych informacji techniki TDM ogólnego zastosowania okazują się zbyt mało skuteczne. W centrum zainteresowania są tu w jeszcze większym stopniu fakty. Do analizy danych biomedycznych wykorzystuje się analizy gramatyczne języka naturalnego, grupowanie i klasyfikację. Nieocenioną pomocą są specjalistyczne słowniki i korpusy, zwłaszcza MEDLINE, pierwsza i wciąż podstawowa baza danych bibliograficznych i abstraktów artykułów biomedycznych oraz PubMed Central – największe repozytorium artykułów biomedycznych w otwartym dostępie. Teksty poddawane analizie często są anotowane. Anotacja jest dokonywana albo ręcznie przez ludzi, albo

³⁴ Tamże, s. 327 i nn.

³⁵ *These tasks include the recognition of explicit facts from biomedical literature, the discovery of previously unknown or implicit facts, document summarization, and question answering*, Matthew S. Simpson, Dina Demner-Fushman, „Biomedical Text Mining: A Survey of Recent Progress”, w: *Mining Text Data*, s. 465.

automatycznie z korektą wykonaną przez człowieka. Specjalistyczne słownictwo nie jest wolne od wieloznaczności. Jako synonimy określonych terminów naukowych należy rozpoznać np. „atak serca” (heart attack) i „myocardial infarction”, a akronim RA zależnie od kontekstu może oznaczać m.in. „right atrium”, „rheumatoid arthristis” lub „renal artery”.

Jeśli analiza ma prowadzić do uprawnionego ekstrapolowania wyników, to metodologiczna poprawność wymaga, by korpusy poddawane badaniu były wystarczająco duże i reprezentatywne i jednoznaczne.

Metody TDM wg raportu dla Komisji Europejskiej

Autorzy raportu o prawnych aspektach TDM, przygotowanego na zlecenie Komisji Europejskiej, zdecydowali się pominąć techniczne szczegóły tego, jak przebiega proces TDM³⁶. Nie odbiegają tym od wielu innych opracowań prawniczych tego tematu. Uzasadniają to szybkością zmian technologicznych. Wyróżniają pięć etapów, które ich zdaniem wyczerpująco – z prawnego punktu widzenia – opisują procesy TDM. Wydobycie (*extract*) treści z zewnętrznych zasobów, przekształcenie (*transform*) ich zgodnie z potrzebami operacyjnymi, umieszczenie (*load*) treści w zestawie danych lub repozytorium. W czwartym kroku dane podlegają analizie (*analysis*), a ostatecznym efektem jest nowa wiedza (*report* w szerokim rozumieniu). Pierwsze trzy kroki autorzy określają akronimem ETL. Czasem kolejność jest inna: umieszczenie treści w kolekcji wyprzedza przekształcenie, co znajduje odzwierciedlenie w jednym z dalszych wątków raportu.

Zamysł wydobycia samej esencji procesu TDM jest godne pochwały. Moim zdaniem autorzy raportu popadli jednak w nadmierną ogólnikowość. W szczególności etap przekształcenia wymaga bardziej zniuansowanego opisu. Choć w wielu miejscach autorzy raportu sugerują, że pod pojęciem TDM należy rozumieć różne metody, to zasadniczy tok ich rozumowania nie uwzględnia tego zróżnicowania. TDM jest traktowany jako jedna metoda badawcza.

Podsumowanie opisu procesu TDM

Podsumowując wstępne ustalenia, czym są metody TDM, należy podkreślić, że jest to przede wszystkim zestaw różnorodnych metod analizy. W wielu przypadkach zastosowanie jednej metody wymaga uprzedniego zastosowania innych. Pewne metody są jednak alternatywne. Postrzeganie metod TDM jako jednej metody badawczej, choć upraszcza złożone zagadnienie, wprowadza w błąd.

Dla analizy prawnej ważny jest podział metod TDM ze względu na **zaangażowanie człowieka** (metody nadzorowane i nienadzorowane), ze względu na **efekt pracy** (ustrukturyzowany zbiór danych, w którym zachowane zostało dzieło poddawane analizie; tłumaczenie lub

³⁶ Jean-Paul Triaille i inni, *Study on the legal framework of text and data mining (TDM)*, 2014, s. 28.

podsumowanie; baza danych zawierające same tylko fakty, bez przejmowania elementów chronionych prawem autorskim) oraz ze względu na **ingerencję w treść i formę** dzieła analizowanego (ewentualne czyszczenie analizowanego materiału z szumu, czyli elementów utrudniających analizę). Ten podział stanowi dla mnie punkt odniesienia w dalszych częściach pracy.

Pozaautorskie zagadnienia prawne związane z TDM w badaniach naukowych

Metody TDM wiążą się z szeregiem problemów prawnych. Są one związane z procesem analizy i wykonywanymi w jej ramach operacjami oraz z przedmiotem tych analiz. Ponieważ zarówno przedmiot badań, jak i metody TDM są bardzo zróżnicowane (o wiele bardziej niż zazwyczaj uwzględnia to literatura prawnicza, w której często w celu uproszczenia mówi się o TDM tak, jakby była to jednorodna metoda analizy), w niniejszej pracy można co najwyżej zarysować kilka problemów prawnych, które wydają się mieć największy ciężar gatunkowy. W konkretnych sytuacjach może okazać się, że zarówno zbiory poddawane analizie, jak i metody analizy sprawiają, że *in casu* pojawią się odmienne kwestie prawne od tych zawartych w niniejszej pracy, albo ze względu na dodatkowe okoliczności ich ocena wypadnie pod pewnymi względami inaczej.

Polska doktryna dotychczas zajmowała się kwestiami związanymi z TDM co najwyżej marginalnie. Na gruncie prawa autorskiego chyba jedyną pracą, która wprost dotyczy tych kwestii, jest artykuł Anny Górskiej i Rafała Górskiego o tworzeniu i udostępnianiu korpusów tekstowych i problemach z tym związanych na gruncie prawa autorskiego³⁷. Autorzy zauważyli, że wskutek nowelizacji ustawy o ochronie baz danych z 2007 r. część ich tez zdezaktualizowała się. Niestety, nie rozwinęli szerzej tego komentarza³⁸. Autorzy analizowali problemy prawne związane z korpusami językowymi, a w szczególności – narodowym korpusem języka polskiego. Przedmiot ich pracy częściowo pokrywa się z tematem mojej pracy. Opracowywanie korpusów językowych, a później ich analiza, to jeden z przykładów stosowania metod TDM. Jednocześnie ich zakres zainteresowania wykracza poza samo tylko prawo autorskie, choć to ono stanowi trzon ich pracy. Problemy prawne, jakie zauważają, są związane z korpusem językowym jako bazą danych oraz z korpusem językowym jako utworem w rozumieniu PrAut. Autorzy zastanawiają się, czy korpus spełnia wszystkie przesłanki utworu (twórczość i indywidualność), czy twórcy korpusu mogą korzystać z utworów włączanych do korpusu na zasadach dozwolonego użytku publicznego, dozwolonego użytku instytucji naukowej lub na zasadach cytatu. Analizują także, czy przy tworzeniu korpusu zachodzi ryzyko naruszenia integralności utworów włączanych w procesie

37

A. Górską, R. Górski, *Tworzenie i udostępnianie korpusów tekstowych – problemy na gruncie prawa autorskiego*, w: *Prace Instytutu Prawa Własności Intelektualnej UJ, Zagadnienia Prawa Autorskiego, Z. 97*, red. A. Matlak, Kraków 2007, s. 47-78.

³⁸ Tamże, s. 78.

anotowania (dodawania informacji gramatycznej do poszczególnych słów) oraz czy włączanie utworów do korpusu bez zgody uprawnionych może zostać zakwalifikowane jako nierzetelne wykorzystanie dzieła. Do wniosków, które z reguły nie są jednoznaczne, będę odnosił się w następnych częściach pracy. W tym miejscu chciałbym poprzestać na identyfikacji problemów prawnych.

Poza Polską dyskusje o prawnych aspektach TDM toczą się od dłuższego czasu. Wiele uwagi poświęca się prawnym aspektom TDM w obszarze ochrony danych osobowych i prywatności.

Dotychczas kwestie prawne związane z maszynową analizą danych były szeroko dyskutowane w obszarze ochrony danych osobowych, ponieważ te analizy są silnie powiązane z tematami profilowania i reklamy behawioralnej³⁹. Maszynowa analiza tekstów jako mniej związana z ochroną prywatności, nie cieszyła się dotąd takim wzięciem wśród prawników. Na działania będące częścią TDM należy patrzeć także z perspektywy ochrony danych osobowych, zarówno gdy TDM jest stosowany w celach marketingowych, jak i w kontekście badań naukowych, choćby medycznych lub społecznych. Analiza maszynowa danych medycznych lub pozyskanych w trakcie badań ankietowych napotyka na trudności prawne związane z wymogami uzyskiwania zgód pacjentów na przetwarzanie ich danych, ewentualnie wymagana jest odpowiednia anonimizacja lub przedstawianie danych w formie zagregowanej.

Prawdopodobnie w każdym przypadku maszynowej analizy danych badacze będą musieli odpowiedzieć sobie na pytania o zgodność podejmowanych czynności z przepisami konstytuującymi ochronę baz danych⁴⁰. Na poziomie unijnym odpowiednie regulacje znajdują się w dyrektywach: Infosoc i o bazach danych⁴¹.

W szczególnych przypadkach procesy TDM w badaniach naukowych mogą wymagać przeanalizowania szczególnych przepisów regulujących konkretne obszary, np. ustawy medyczne lub ustawy o nauce.

³⁹ *From a legal point of view, data mining has already been extensively discussed in the field of data protection, because it is strongly linked with the topics of profiling and behavioural advertising*, Maarten Truyens, Patrick Van Eecke, *Legal aspects of text mining*, LREC Conference 2014, Proceedings, <http://www.lrec-conf.org/proceedings/lrec2014/index.html>.

⁴⁰ Por. Maarten Truyens, Patrick Van Eecke, *Legal aspects of text mining*, LREC Conference 2014, Proceedings, <http://www.lrec-conf.org/proceedings/lrec2014/index.html>, s. 2184.

⁴¹ Por. Jean-Paul Triaille i inni, *Study on the legal framework of text and data mining (TDM)*, 2014, s. 29-40.

Kwestia dopuszczalności metod TDM w świetle prawa UE

Coraz powszechniejsze stosowanie maszynowej analizy tekstu i danych oraz płynące z tego korzyści skłaniają do postawienia pytania o dopuszczalność tych praktyk w świetle ustawodawstw poszczególnych państw. Świadomość problemu jest obecna także na poziomie europejskim. W grudniu 2015 r. Komisja Europejska opublikowała komunikat wskazujący proponowane kierunki „modernizacji” prawa autorskiego w UE⁴². Znalazło się w nim miejsce również na kwestie związane z TDM⁴³: *Brak jasnych przepisów unijnych regulujących TDM w celu badań naukowych stwarza sytuację niepewności wśród społeczności naukowej*⁴⁴. Komisja zapowiedziała przedstawienie propozycji wprowadzenia dozwolonego użytku dla instytucji naukowych uprawniającego do stosowania metod TDM w celach naukowych na wszelkich materiałach, do których te instytucje mają legalny dostęp bez uzyskiwania specjalnych zgód uprawnionych⁴⁵.

Pojawianie się wątku TDM w kontekście dyskusji o digital single market nie powinno zaskakiwać. Jest to efektem trwającego przynajmniej kilka lat zainteresowania tą tematyką w instytucjach unijnych, w tym Komisji. We wrześniu 2015 r. rozpoczął się dwuletni projekt FutureTDM, mający na celu zaproponowanie rozwiązań, które zlikwidują bariery prawne i polityczne ograniczające stosowanie metod TDM w Europie⁴⁶.

Maurizio Borghi i Stavroula Karapap zwięźle przedstawili paradoks, który sprawia, że stosowanie TDM budzi wątpliwości prawne: *Automatyczne analizowanie tekstu powoduje paradoks w prawie autorskim. Z jednej strony automatyczna analiza zakłada powtarzające się kopiowanie całych utworów; w tym sensie to jest przykład prima facie naruszenia. Z drugiej strony jednak, celem zwielokrotnienia jest wydobycie informacji z tekstu i o tekście, czyli działalność, która normalnie nie jest zaliczana do kategorii naruszeń prawa autorskiego*⁴⁷. W następnych rozdziałach

⁴² http://europa.eu/rapid/press-release_IP-15-6261_en.htm.

⁴³ Por. <https://blogs.ucl.ac.uk/copyright/2015/12/23/eu-reform-plans-text-and-data-mining-tdm/>.

⁴⁴ *The lack of a clear EU provision on TDM for scientific research purposes creates uncertainties in the research community*, COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Towards a modern, more European copyright framework, Bruksela, 09.12.2015, s. 7.

⁴⁵ *Allow public interest research organisations to carry out text and data mining of content they have lawful access to, with full legal certainty, for scientific research purposes*, Tamże, s. 8.

⁴⁶ <http://project.futuretdm.eu/project-overview/>.

⁴⁷ *Automated text processing presents a paradox for copyright law. On one side, automated processing presupposes the repeated copying of whole works; in this respect, it is an exemplary prima facie case for infringement. On the other side, however, the purpose of this reproduction is to extract information from texts and about texts, an activity that does not normally amount to an infringement in copyright law*, M. Borghi, S. Karapap, *Copyright and Mass*

zostanie wyjaśnione, na ile to stwierdzenie jest słuszne w odniesieniu do polskiego prawa autorskiego.

Dotychczasowy dorobek obejmuje m.in. dwa raporty opublikowane w 2014 r., zamówione przez Komisję Europejską. Pierwszy, autorstwa zespołu ekspertów pod kierownictwem Iana Hargreavesa, poświęcony został zarówno kwestiom ekonomicznym, jak i prawnym⁴⁸. Drugi raport koncentruje się wyłącznie na kwestiach prawnych⁴⁹. W dokumencie *Amsterdam Call for Action on Open Science*, określającym cele w obszarze komunikacji naukowej dla UE do osiągnięcia do 2020 r., wskazano m.in. na potrzebę uregulowania zasad prowadzenia badań metodami TDM⁵⁰.

Raport Hargreavesa składa się z trzech części, które razem zarysowują szerokie tło, a zarazem przegląd zagadnień związanych z TDM. *Coraz lepiej rozumiemy, że znajdujemy się u progu masowej automatyzacji naszych przemysłów usługowych (automatyzacja myślenia), porównywalnej z dokonaną we wcześniejszej erze automatyzacją linii produkcyjnych (automatyzacja mięśni)*⁵¹. Nie wchodząc w tym miejscu w polemikę co do faktycznego znaczenia upowszechnienia metod TDM w nauce, na ile zastąpią one dotychczasowe metody, a na ile będą stanowiły tylko ich uzupełnienie, chciałbym podkreślić, że nadzieje, jakie niektórzy wiążą z TDM są daleko zakrojone.

W pierwszej części raportu Hargreavesa znajduje się analiza grup interesariuszy zmian obejmujących metody TDM. Najostrzejszy spór rysuje się pomiędzy wielkimi, komercyjnymi wydawcami naukowymi z Elsevierem na czele, a środowiskami akademickimi. *Kluczowym zagadnieniem dla większości badaczy jest możliwość automatycznej analizy treści, do których mają zapewniony legalny dostęp*⁵². Tymczasem wydawcy sugerują raczej konieczność ponoszenia przez instytucje naukowe dodatkowych opłat za licencje upoważniające do analiz metodami TDM, wprowadzając dystynkcję pomiędzy „dostępem” a „analizą maszynową” (*mining*). Wielu naukowców jest przekonanych, że niski poziom stosowania metod TDM w Europie, w porównaniu do USA i państw azjatyckich, jest skutkiem dużej niepewności, co do obowiązującego prawa i tego,

Digitization. A Cross-Jurisdictional Perspective, Oxford University Press, 2013, cytata za: *Standardisation in the area of innovation...*, s. 30.

⁴⁸ *Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining. Report from the expert group*, zespół pod kierunkiem I. Hargreavesa, Unia Europejska 2014.

⁴⁹ J.-P. Traille, J. de Meeûs d'Argenteuil, A. Francquen, *Study on the legal framework of text and data mining (TDM)*, 2014.

⁵⁰ <https://wiki.surfnet.nl/display/OSCFA/2.+Facilitate+text+and+data+mining+of+content>.

⁵¹ *There is growing recognition that we are at the threshold of mass automation of our service industries (automation of thinking) comparable with the robotic automation of manufacturing production lines (automation of muscle) in an earlier era*, *Standardisation in the area of innovation...*, s. 10.

⁵² *For most researchers the key issue is being able to mine content for which they already have legal access*, Tamże, s. 16.

jakie działania są prawnie dopuszczalne, a jakie już nie⁵³.

Druga część raportu przedstawia ekonomiczny punkt widzenia, a trzecia poświęcona została kwestiom prawnym. Kluczowa kwestia prawna w raporcie dotyczy tego, czy optymalną równowagę między interesami różnych grup uda się osiągnąć drogą licencjonowania zasobów, czy poprzez wprowadzenie wyjątku od prawa autorskiego pozwalającego pod pewnymi warunkami stosować metody TDM w ramach dozwolonego użytku? Autorzy raportu zakładają, że stosowanie TDM wkracza w monopol autorski. Przeciwno uznaniu licencji za optymalny sposób uzyskiwania uprawnień do TDM podnosi się m.in. fakt upowszechniania treści naukowych w otwartym dostępie⁵⁴. Już w 2013 r. szacowano, że ok. 40% artykułów naukowych jest w ten sposób dostępnych. Jeśli nie zostały udostępnione na żadnej licencji (np. jednej z licencji Creative Commons), zakres uprawnień użytkowników określony jest ustawowo.

Analizę uwarunkowań prawnych w UE poprzedza szkic rozwiązań przyjętych w USA, Australii, Kanadzie, Izraelu i Japonii. Pierwsze państwa przyjęły doktrynę *fair-use*, oznaczającą otwarty katalog warunków dopuszczających korzystanie z utworów bez zezwolenia uprawnionego. Izrael dołączył do tej grupy w 2007 r. Zależnie od orzecznictwa, w tych państwach można stosować metody TDM w celach badawczych z większą lub mniejszą pewnością, że czyni się to legalnie. Japonia zdecydowała się wprowadzić w 2009 r. wyjątek od prawa autorskiego bezpośrednio dedykowany metodom TDM w celach badawczych. Maszynowa analiza tekstu i danych została opisana za pomocą kilku warunków (nie pojawia się termin TDM, ani podobny)⁵⁵.

W państwach członkowskich Unii Europejskiej, w przeciwieństwie do wielu innych państw, obowiązują przepisy przyznające ochronę *sui generis* baz danych, odrębną od ochrony autorskiej, ze względu na oryginalny dobór, układ lub zestawienie (art. 3 PrAut). Wątkowi „bazodanowemu” poświęca się z tego powodu dużo uwagi. Ochrona *sui generis* baz danych zharmonizowana w państwach członkowskich w Dyrektywie o Ochronie Baz Danych z 1996 r. jest również źródłem niepewności dla badaczy chcących stosować metody TDM.

Niepewność co do możliwości stosowania metod TDM w zgodzie z europejskim prawem autorskim (zwłaszcza z dyrektywą Infosoc) wynika w dużej mierze z faktu, że wyjątki pozwalające na dozwolony użytek miały charakter opcjonalny. Zakres wdrożenia ich został pozostawiony swobodnemu uznaniu państw członkowskich. Raport przedstawia cztery możliwe modele zmniejszania tej niepewności: licencje, przeformułowanie prawa zwielokrotnienia (*jeśli TDM wprowadza nietwórcze, nie naruszające i korzystne ze społecznego punktu widzenia typy*

⁵³ Tamże, s. 20.

⁵⁴ Podstawowe zagadnienia, także prawne, związane z otwartym dostępem do treści naukowych przedstawione zostały w: P. Suber, *Otwarty dostęp*, tłum. Warszawa 2014, <https://depot.ceon.pl/handle/123456789/6598>.

⁵⁵ Tamże, s. 47.

zwielokrotniania, to nie powinny być one objęte zakresem praw wyłącznych⁵⁶) oraz dodanie wyjątku wprowadzającego dozwolony użytek polegający na stosowaniu metod TDM do celów badawczych. Pojawia się pytanie, czy wyjątek ten powinien zostać połączony z obowiązkiem zapłaty wynagrodzenia uprawnionemu z tytułu praw autorskich⁵⁷? Takie wynagrodzenie, przekształcające wyjątek w przymusową lub ustawową licencję, pełniłoby funkcję zachęty do inwestowania w bazy danych, by były technicznie dostosowane do maszynowej analizy.

Drugi raport szczegółowo analizuje kwestie prawne TDM, nie odnosząc się do problemów pozaprawnych. Autorzy argumentują, dlaczego zamiast o TDM wolą mówić o „analizie danych” (*data analysis*). Przedstawiają słuszne argumenty przeciwko sformułowaniu TDM (tekst jest tylko jednym z rodzajów danych; z kolei pojęcie *mining* to nieprecyzyjna metafora). Ich propozycja jest jednak zbyt ogólna. Wymagałaby dodania określenia choćby takiego jak stosowane w Polsce „maszynowa analiza danych”. Kwestie terminologiczne nie są jednak kluczowe.

Autorzy raportu proponują podzielić dane ze względu na zasady dostępu do niego na cztery kategorie, zamiast na dwie (dostęp wolny – dostęp ograniczony): dane sieciowe (*web data*) dostępne dla wszystkich, dane z sieci społecznościowych udostępniane przez wielu dla wielu, dane udostępniane wielu podmiotom na podstawie umowy oraz dane poufne udostępniane indywidualnie.

W raporcie zostało przeanalizowane, czy obecne wyjątki od prawa autorskiego i od ochrony *sui generis* baz danych przewidziane w dyrektywach są wystarczające dla TDM. Ich zakres zainteresowania obejmował nie tylko badania naukowe, ale i inne cele analiz TDM. Poszczególne wyjątki i przesłanki, od których uzależnione jest powoływanie się na nie, analizowali dla następujących etapów procesu TDM: pozyskiwanie źródeł, przetwarzanie, dołączanie do zbioru (*load*), analiza oraz efekt finalny (*output*).

Dyrektywa Infosoc nakłada na państwa członkowskie obowiązek przyznania autorom wyłącznego prawa do zezwalania lub zabrania zwielokrotniania utworów w całości lub części w jakikolwiek sposób (art. 2 dyrektywy). Jednocześnie wprowadza obligatoryjny wyjątek, który umożliwia na dokonywanie zwielokrotnienia, jeśli ma ono charakter tymczasowy, przejściowy lub dodatkowy, stanowi integralną i podstawową część procesu technologicznego, a zarazem jego jedynym celem jest legalne korzystanie z utworu (art. 5 ust. 1). Dyrektywa dopuszcza wprowadzenie przez państwa członkowskie szeregu innych wyjątków od monopolu autorskiego, m.in. w celach badań naukowych (art. 5 ust. 3 a).

Zdaniem autorów raportu, wyjątek dopuszczający tymczasowe i przejściowe

⁵⁶ *If TDM constitutes non-expressive, non-infringing, and socially beneficial types of reproduction, then these should not fall within the ambit of the exclusive right*, Tamże, s. 53.

⁵⁷ Tamże, s. 56.

zwielokrotnianie dla zgodnego z prawem korzystania okazał się niewystarczający, by oprzeć na nim TDM⁵⁸. Zwielokrotnianie przejściowe oznacza bowiem tylko takie zwielokrotnienie, które jest konieczne np. do wyświetlenia treści w przeglądarce internetowej, a treści są przechowywane w pamięci *cache* lub RAM i są automatycznie kasowane po wykorzystaniu.

Z kolei wyjątek dla badań naukowych w celach niekomercyjnych prowadzonych przez instytucje naukowe stanowi wystarczającą podstawę dla badań metodami TDM. Należy jednak pamiętać, że ma on charakter opcjonalny i jego wdrożenia różnią się w zależności od państwa członkowskiego. Wiele problemów sprawia odróżnienie celów komercyjnych od niekomercyjnych⁵⁹.

Podstawową rekomendacją zawartą w raporcie jest postulat wprowadzenia wyjątku umożliwiającego prowadzenie badań naukowych metodą TDM. Rekomendacja ta ma na celu przede wszystkim usunięcie niepewności prawnej na obszarze całej UE, wynikającej z opcjonalnego charakteru wyjątku dla badań naukowych. Niepewność ta ogranicza współpracę między przynależnymi różnymi państwami członkowskimi. Pytania, na które należałoby odpowiedzieć, dotyczą tego, czy taki wyjątek powinien być ograniczony do celów badawczych lub niekomercyjnych, czy powinien być obowiązkowy dla państw członkowskich, czy należy umożliwić umowne zrzeczenie się go?

Dyskusja wokół prawnych aspektów TDM toczy się w Europie także poza instytucjami unijnymi. Dotychczas zostały przeprowadzone prawdopodobnie tylko jedno badanie empiryczne mające na celu zbadanie wpływu poziomu ochrony prawnoautorskiej na stosowanie metod TDM w celach naukowych⁶⁰. Autorzy tych badań patrzą na prawo autorskie przez pryzmat balansu, który należy osiągnąć pomiędzy niedostatecznym wykorzystaniem prac nadmiernie chronionych przez prawo autorskie, a brakiem wystarczających zachęt do tworzenia, gdy brak jest ochrony wystarczająco silnej⁶¹. Opierając się na danych zgromadzonych w bazie Web of Science obliczyli wskaźnik badań wykorzystujących metody TDM w stosunku do wszystkich badań z danego państwa (według afiliacji głównego autora) w ujęciu rocznym. Zauważyli negatywną korelację między silną ochroną prawnoautorską (zarówno jeśli chodzi o treść norm, jak i o „praworządność”, *rule of law*) a liczbą badań TDM. Dane dla Polski są odrobinę lepsze niż średnia dla piętnastu

⁵⁸ J.-P. Traille, J. de Meeûs d'Argenteuil, A. Francquen, *Study on the legal framework of text and data mining (TDM)*, 2014, s. 50.

⁵⁹ Tamże, s. 62.

⁶⁰ Handke, Christian and Guibault, Lucie and Vallbé, Joan-Josep, *Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research* (June 7, 2015). Available at SSRN: <http://ssrn.com/abstract=2608513>.

⁶¹ *Underutilization of existing works with strong copyright protection and underproduction of new works without it*, tamże, s. 3.

największych państw UE (por. tab. 1, s. 10). *Silna krajowa ochrona prawnoautorska może hamować wykorzystanie utworów z innych państw, gdy tymczasem wartościowe dane [wytworzone w tym kraju] będą dostępne na terytoriach ze słabszą ochroną prawnoautorską. Może okazać się, że państwa z mocną ochroną otrzymają to, co najgorsze z obu światów: rozległe, nieautoryzowane wykorzystanie zagranicą danych wytworzonych w kraju i wysokie koszty prowadzenia TDM w kraju*⁶².

W przypadku naukowców charakterystyczne jest to, że korzyści, na jakie liczą bezpośrednio, związane są z publikacjami w prestiżowych czasopismach i uzyskiwanie cytowań, co przekłada się na możliwość pozyskiwania grantów na kolejne badania. Nie odnoszą więc korzyści, jakie zazwyczaj łączy się z prawem autorskim⁶³.

Za wprowadzeniem szeroko zakrojonego wyjątku pozwalającego stosowanie metod TDM opowiadają się Diana Cocoru i Mirko Boehm⁶⁴. Z przeprowadzonych przez nich badań wynika, że praktycy TDM postrzegają swoją działalność albo w kontekście wyjątku pozwalającego wykonywanie tymczasowych kopii, albo w kontekście wyjątku uprawniającego do prowadzenia badań naukowych. W pierwszym przypadku postrzegają TDM jako czynności czysto techniczne. Związek z prawem autorskim wynika z przeniesienia zasad mających uzasadnienie w przypadku fizycznych egzemplarzy dzieła. W przypadku badań naukowych warunków prowadzenia badań w celach niekomercyjnych to otwarcie puszek Pandory, ponieważ granica między tym, co komercyjne, a tym co niekomercyjne jest nieostra, a szara strefa bardzo szeroka. Co więcej, wszelkie działania podejmowane przez naukowców w partnerstwie z przedsiębiorstwami zostają wyłączone.

Celem TDM jest na ogół wydobycie pewnych faktów. *Nagie fakty nie są chronione prawem autorskim i zgodnie z tą samą logiką wydobywanie faktów z utworów lub dzieł sztuki nie powinno powodować naruszenia prawa autorskiego*⁶⁵. Analiza TDM zdaniem autorów wzmacnia gospodarkę jako całość. Korzyści są rozłożone na wiele podmiotów (np. korzyści dla pacjentów), a w czasie negocjacji z uprawnionymi, np. z wydawcami, pomija się je⁶⁶. Szukając balansu nie powinno się o nich jednak zapominać.

Cocoru i Boehm starają się sprowadzić problem metod TDM do prostego zagadnienia i

⁶² *Then strong domestic copyright protection may inhibit transfer and use of input works into countries, whereas valuable data will be accessible in territories with less copyright protection. High protection countries may get the worst of both worlds: extensive unauthorized use of domestically produced data abroad and high costs of conducting DM domestically*, Tamże, s. 21.

⁶³ Tamże, s. 22.

⁶⁴ D. Cocoru, M. Boehm, *An analytical review of text and data mining practices and approaches in Europe Policy recommendations in view of the upcoming copyright legislative proposal*, Open Forum Europe, 1.05.2016, <http://www.openforumeurope.org/ofe-publishes-high-level-policy-paper-text-data-mining/>.

⁶⁵ *However, bare facts are not protected by copyright and by the same logic, the extraction of facts from a work of authorship or art should not constitute a copyright infringement*, Tamże, s. 8.

⁶⁶ Tamże, s. 15.

wskazać łatwe rozwiązanie. Maksymalistyczny postulat wprowadzenia wyjątku uprawniającego do stosowania metod TDM przez dowolny podmiot, w celach zarówno komercyjnych, jak i niekomercyjnych, niekoniecznie związanych z badaniami naukowymi być może bierze się z przecenienia rzeczywistych korzyści dla drobnych podmiotów oraz niedoceny argumentów sceptyków.

Tymczasem Jaron Lanier przestrzega przed obecnie obserwowanymi trendami, które sprawiają, że „zwycięzcy biorą wszystko”. Jeśli korporacje będą korzystać z twórczego wkładu dostarczanego przez osoby, które nie będą uzyskiwać z tego tytułu benefitów, gospodarka będzie coraz bardziej rozregulowana: *Jeśli oczekujemy od komputerów, by prowadziły samochody i kierowały fabrykami, zatrudnienie, które pozostanie, powinno wiązać się z zawodami kreatywnymi, z twórczością i prawami własności intelektualnej. Ale jeśli to podkopujemy, spowodujemy kryzys zatrudnienia na ogromną skalę (...). Krytyczna jest podstawowa zasada, że płacimy ludziom za ich informacje i wkład (...). Własność intelektualna musi stać się o wiele bardziej wyszukana i zniuansowana. Musi dawać korzyści wszystkim (...). Chciałbym zobaczyć więcej systemów, w których zwykli ludzie mogą uzyskać wynagrodzenie, kiedy dokładają od siebie wartość do cyfrowych sieci*⁶⁷.

23 grudnia 2015 r. sąd w Amsterdamie wydał orzeczenie w sporze o ochronę prawnoautorską *Dzienników Anny Frank*. Jednym z zagadnień była ewentualna konieczność uzyskania zgody od podmiotu uprawnionego z tytułu praw autorskich na przeprowadzenie analiz metodą *text-miningu* w celach naukowych. Holenderski sąd stwierdził, że taka zgoda nie jest potrzebna dopóty, dopóki badacze mają legalny dostęp do utworu. Sąd uzasadnił to zasadą swobody badań naukowych⁶⁸. Był to prawdopodobnie pierwszy wyrok w Europie w takiej sprawie.

⁶⁷ *If we expect computers to pilot cars and operate factories, the employment that is left should be the creative stuff, the expression, the I[n]tellectual P[roperty]. But if we undermine that, we are creating an employment crisis of mass proportions. (...) The general principle that we pay people for their information and contributions is critical (...). But IP needs to be made much more sophisticated and granular. It needs to be something that benefits everybody (...). I would like to see more systems where ordinary people can get paid when they contribute value to digital networks,* http://www.wipo.int/wipo_magazine/en/2016/02/article_0001.html

⁶⁸ Por. <http://www.communia-association.org/2016/01/08/what-the-diary-of-anne-frank-can-tell-us-about-text-and-data-mining/>.

Stosowanie metod TDM na podstawie przepisów o dozwolonym użytku

Czy utwory, do których badacze mają legalny dostęp, mogą być przedmiotem maszynowej analizy bez dodatkowej zgody uprawnionego i czy metody TDM stosowane w badaniach naukowych wyczerpują przesłanki dozwolonego użytku? Przedmiotem badań naukowych stosujących metody TDM są często duże zbiory danych lub korpusy tekstów. Jeśli prawa autorskie do nich są rozproszone wśród wielu uprawnionych, konieczność uzyskiwania od nich dodatkowego zezwolenia często oznaczałaby brak możliwości przeprowadzenia takich badań. O możliwości prowadzenia badań tymi metodami i o ich ograniczeniach prawnych decyduje status prawny tych danych i sposób uzyskania do nich dostępu.

W jakich sytuacjach stosowanie metod TDM nie wkracza w obszar objęty monopolem autorskim, a gdzie znajduje się granica poza którą w obecnym stanie prawnym uprawniony z tytułu praw autorskich może żądać zaprzestania takiej działalności?

Generalną zasadą jest, że *twórcy przysługuje wyłączne prawo do korzystania z utworu i rozporządzania nim na wszystkich polach eksploatacji* (art. 17 PrAut). *Prawo do korzystania z utworu w art. 17 powinno być rozumiane przede wszystkim jako uprawnienie do podejmowania czynności o charakterze faktycznym. Prawo to zawiera oczywiście także stronę negatywną, tj. możliwość zakazania innym podmiotom korzystania z dzieła bez zezwolenia autora*⁶⁹. Ograniczenia tego prawa wyłącznego w postaci dozwolonego użytku mają charakter wyjątku i nie należy interpretować ich w sposób rozszerzający, ani stosować analogii. W badaniach metodami TDM kluczowa jest możliwość zwielokrotniania utworów, opracowywania ich i prezentowania wyników. Wyprzedzając szczegółowe rozważania należy zwrócić uwagę na dozwolony użytek dla instytucji naukowych w celu prowadzenia badań naukowych (art. 27) i dozwolony użytek osobisty (art. 23). Instytucje naukowe zostały uprawnione do szerokiego korzystania z utworów w celach naukowych, także jeśli mają one charakter komercyjny. Nie mogą jednak na podstawie tego przepisu zwielokrotniać utworów (z wyjątkiem fragmentów większych utworów lub drobnych utworów w całości). Z kolei osoby fizyczne są uprawnione do swobodnego korzystania z utworów, w tym do zwielokrotniania ich, byleby to korzystanie mieściło się w ramach użytku osobistego.

Można wyróżnić zasadniczo trzy sposoby dostępu do danych. Pierwszy typ danych, to

⁶⁹ E. Traple, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. J. Barta, R. Markiewicz, wyd. 5, Warszawa 2011, s. .

materiały znajdujące się w domenie publicznej. Nie są one objęte ochroną prawnoautorską (przynajmniej jeśli chodzi o prawa majątkowe), w związku z czym nie ma ryzyko wkroczenia w monopol autorski. Dane (w tym teksty) mogą znaleźć się w domenie publicznej albo z powodu upływu czasu (najczęściej siedemdziesięciu lat od śmierci autora) i wygaśnięcia w związku z tym autorskich praw majątkowych, albo z powodu nie objęcia takimi prawami obiektów wymienionych w art. 4 PrAut, choćby posiadały cechy utworu (akty normatywne i ich urzędowe projekty, dokumenty urzędowe, opublikowane opisy patentowe i proste informacje prasowe).

Drugi typ materiałów, które mogą stać się przedmiotem analiz metodami TDM, stanowią materiały, do których dostęp określony jest na podstawie umowy, w tym umowy licencyjnej. Jeśli postanowienia umowne wyraźnie dozwolają stosowanie takich metod, sytuacja badacza jest prosta. Wątpliwości może budzić kwestia skuteczności umownej klauzuli zabraniającej analizowanie danych metodami TDM.

Szczególnym przypadkiem licencjonowania materiałów jest udostępnianie ich na licencji adresowanej do nieoznaczonej grupy odbiorców. Najpopularniejszymi takimi licencjami są licencje Creative Commons (CC)⁷⁰. Funkcjonuje sześć odmian licencji CC, które przyznają różne uprawnienia użytkownikom. Funkcjonują one w czterech wersjach (od 1.0 do 4.0) oraz odmianach językowych, w tym po polsku (najnowsza odmiana polska jest w wersji 3.0). Odmiany językowe stanowią nie tylko tłumaczenie tekstu angielskojęzycznego, ale również dostosowują tekst licencji do lokalnego prawodawstwa. Wszystkie licencje CC zawierają warunek uznania autorstwa. Pozostałe warunki, które występują w różnych konfiguracjach, to: „na tych samych warunkach” (utwory zależne należy udostępnić na takiej samej licencji), „bez utworów zależnych” (licencja nie pozwala na rozpowszechnianie utworów zależnych) i „użycie niekomercyjne” (nie pozwalający na wykorzystanie utworu w celach komercyjnych). Oprócz licencji CC funkcjonuje tzw. znak domeny publicznej, czyli Creative Commons Zero⁷¹. W Polsce skutek CC 0 jest ograniczony i przypomina bardzo szeroko zakrojoną licencję.

Trzecim typem materiałów są materiały udostępnione w internecie na zasadach dozwolonego użytku, tzn. każdy może swobodnie zapoznać się z tymi materiałami bez uzyskiwania na to specjalnej zgody, na ogół bez potrzeby rejestrowania się (rejestracja często jest związana z jakąś formą licencji). Nie uzyskuje jednak żadnych uprawnień wykraczających poza ustawowe przepisy o dozwolonym użytku. W tym rozdziale chciałbym skupić się na tej właśnie grupie materiałów.

Dozwolony użytek: charakter instytucji

⁷⁰ Por. <https://creativecommons.pl/poznaj-licencje-creative-commons/>.

⁷¹ Por. Krzysztof Siewicz, *Analiza prawna „Creative Commons 0”*, http://koed.org.pl/wp-content/uploads/2012/09/CC0_analiza.pdf.

Intensyfikacja dyskusji o charakterze instytucji dozwolonego użytku jest częścią szerszej dyskusji o charakterze prawa autorskiego w ogólności. Prawnicy coraz częściej dostrzegają, że konstrukcje prawne opracowane w czasach „analogowych” wpisują się tylko z najwyższym trudem w warunki, jakie powstały po gwałtownym rozwoju technik komputerowych i przeniesienia znaczącej części obiegu utworu do środowiska cyfrowego⁷².

J. Marcinkowska patrzy na prawo autorskie jako na regulacje, które nie tylko zapewniają ochronę autorom, ale i zabezpieczają interes społeczny. To były zawsze dwie strony monety: awers to wsparcie twórców i stwarzanie zachęt dla twórczości, a rewers to gwarancje korzystania z twórczości przez ogół społeczeństwa. *Zgłębianie dziejów prawa autorskiego pozwala zauważyć, że w niemal trzystuletnim okresie jego kształtowania zmierzano nie tylko do zapewnienia jak najszerzej ochrony majątkowych i osobistych interesów autorów oraz ich najbliższych, ale także dążono do uwzględnienia interesów społecznych w korzystaniu z osiągnięć kultury*⁷³.

W. Machała w artykule z 2001 r. postawił tezę, że dozwolony użytek jest prawem podmiotowym użytkownika⁷⁴. Tezę tę rozwinął po dwóch latach w swojej dysertacji w odniesieniu już tylko do dozwolonego użytku prywatnego⁷⁵. Wychodząc od stwierdzenia, że w cywilistyce sfera swobodnego działania podmiotu jest opisywana kategorią prawa podmiotowego i powołując się na definicję prawa podmiotowego zaproponowaną przez S. Grzybowskiego, zauważa, że w przypadku przepisów o dozwolonym użytku występuje *typowa dla prawa podmiotowego sytuacja, w której z uprawnieniem jednego podmiotu wiąże się obowiązek jego poszanowania (znoszenia) przez inne osoby*⁷⁶. Uznanie dozwolonego użytku za prawo podmiotowe miałooby istotne znaczenie praktyczne, np. użytkownik miałby prawo skutecznie zrzec się swojego uprawnienia w drodze umowy. Pojawiałyby się sprzeczność z dopuszczalnością zabezpieczeń technicznych utworów przed kopiowaniem (DRM), wprowadzonej do polskiego prawa zgodnie z tendencjami w prawie międzynarodowym.

Stanowisko W. Machały krytykuje L. Małek: opieranie się na koncepcji prawa podmiotowego było przejawem mody, która przeminęła⁷⁷. Pojęcie prawa podmiotowego przestało być wykorzystywane tak często, jak miało to miejsce wcześniej. Mimo krytyki metody argumentacji Machały, L. Małek prowadzi swoją argumentację w podobnym kierunku: szuka

⁷² O znaczeniu powstania cyfrowego obiegu dla możliwości uczynienia z niektórych tekstów dobra publicznego, tak jak dobrem publicznym jest wiedza: P. Suber, *Knowledge as a Public Good*, w: P. Suber, „Knowledge unbound. Selected writings on open access 2002-2011”, Cambridge, Massachusetts, London, 2016, <https://mitpress.mit.edu/books/knowledge-unbound> oraz *Understanding Knowledge as a Commons. From Theory to Practice*, red. Ch. Hess, E. Ostrom, Cambridge, Massachusetts, London, 2007.

⁷³ J. Marcinkowska, *Dozwolony użytek w prawie autorskim. Podstawowe zagadnienia*, Prace Instytutu Prawa Własności Intelektualnej UJ, z. 87, s. 15.

⁷⁴ W. Machała, *Dozwolony użytek utworów w polskim prawie autorskim w świetle cywilistycznej koncepcji prawa podmiotowego*, w: *Prace z Wynalazczości i Ochrony Własności Intelektualnej*, rok 2001, Z. 78, s. 53-80.

⁷⁵ W. Machała, *Dozwolony użytek prywatny w polskim prawie autorskim*, Warszawa 2003.

⁷⁶ W. Machała, *Dozwolony użytek utworów...*, s. 55-56.

⁷⁷ L. Małek, *Cytat w świetle prawa autorskiego*, Warszawa 2011, s. 58.

argumentów wzmacniających pozycję użytkownika względem autora. Jego konkluzje są jednak ostrożniejsze. Obaj autorzy są zgodni co do diagnozy, że powszechna opinia o zakazie wykładni rozszerzającej przepisów o dozwolonym użytku jest stanowiskiem zbyt sztywnym i nieadekwatnym do postępu technicznego⁷⁸. L. Małek swoje rozumowanie dotyczące charakteru dozwolonego użytku w prawie autorskim opiera na dwóch ujęciach tej instytucji: ujęciu zewnętrznym i wewnętrznym⁷⁹. Zgodnie z tym drugim, przepisy o dozwolonym użytku w taki sam sposób współtworzą uprawnienia twórcy, jak przepisy o przedmiocie prawa autorskiego. Takie ujęcie wzmacnia pozycję użytkownika. Z kolei ujęcie zewnętrzne wzmacnia autora, stwarzając domniemanie kompetencji na jego rzecz. Dozwolony użytek tylko ogranicza możliwość wykonywania jego praw, ale nie ogranicza ich przedmiotu. Panuje zgoda, że dozwolony użytek publiczny jest zewnętrzny względem przedmiotu prawa autorskiego. Oznacza to, że nie stanowi on elementu konstytuującego prawo autorskie. *Ujęcie „zewnętrzne” (...) jest dużo korzystniejsze dla podmiotu autorskouprawnionego; wzmacnia jego pozycję w sferze prawa materialnego, co wyraża się w przypisaniu mu domniemania kompetencji, czyli (...) przerwaniu na barki użytkownika ciężaru udowodnienia zachowania się w granicach dozwolonego użytku*⁸⁰.

Beata Giesen sprzeciwia się stosowaniu konstrukcji prawa podmiotowego w kontekście dozwolonego użytku. Zwraca jednak uwagę na ryzyka związane ze stosowaniem modelu własnościowego w prawie autorskim. Porównanie niematerialnego utworu i materialnej rzeczy wskazuje na szereg istotnych różnic, m.in. cel (monopol autora ma stanowić zachętę dla twórczości) i sztuczność bariery w dostępie do dzieła. Co więcej, *przetrasponowanie rzymskiej idei bezwzględnej władztwa nad rzeczą na grunt prawa autorskiego doprowadziło do stworzenia kategorii prawa absolutnego, ale tym razem w dosłownym znaczeniu, tzn. takiego, które nie ma żadnych granic. Dowodem na to jest porównanie formuły art. 140 k.c. oraz art. 17 pr.aut. Uwagę zwraca fakt, że pierwszy z wymienionych przepisów, definiując prawo własności, wskazuje nie tylko na uprawnienia właścicielskie, ale wytycza również granice prawa. Adekwatnej formuły brak natomiast w art. 17 pr.aut., który ogranicza się wyłącznie do wskazania na treść uprawnień twórcy. Potrzebie swobodnego dostępu do utworów mają z założenia czynić zadość przepisy o dozwolonym użytku, tym prywatnym oraz tym publicznym. W aktualnym kształcie prawnym dozwolony użytek ujęty został jednak jako usankcjonowany prawem akt wkroczenia w sferę cudzego monopolu. W konsekwencji skopiowanie kilku stron nabytej uprzednio książki na własny użytek oznacza naruszenie prawa autorskiego, co jednak nie jest bezprawne na mocy postanowień ustawy*⁸¹.

⁷⁸ L. Małek, *Cytat...*, s. 43.

⁷⁹ Tamże, s. 56 i nn.

⁸⁰ L. Małek, *Cytat...*, s. 56.

⁸¹ B. Giesen, *Własnościowy model prawa autorskiego – analiza koncepcji przyjętej w prawie polskim*, [w:] „Ruch Socjologiczny, Ekonomiczny i Prawniczy”, 2/2015, s. 72.

Nieuzasadnione, zdaniem autorki, rozciągnięcie monopolu autorskiego sprawia, że wszyscy są niezadowoleni: i twórcy, i użytkownicy.

Prawo autorskie stawia twórcę w uprzywilejowanej pozycji. Przejawia się to w szeregu instytucji prawnych, jak choćby uprawnieniu do dodatkowego wynagrodzenia z tytułu pewnych form korzystania z utworu (art. 70¹ PrAut), czy prawo do wynagrodzenia w przypadku dokonywanej zawodowo odsprzedaży oryginalnych egzemplarzy utworów plastycznych lub fotograficznych (art. 19 PrAut). Przede wszystkim jednak o uprzywilejowanej pozycji twórcy decydują podstawowe zasady, na jakich oparta została konstrukcja prawa autorskiego. Art. 16 konstytuuje prawa autorskie osobiste, które chronią więź twórcy z utworem. Są one nieograniczone w czasie i niezbywalne (co *notabene* chroni nie tylko interes autora, ale i interes publiczny, polegający np. na nielegalności takich praktyk jak *ghostwriting*; A. Michalak zwrócił uwagę, że interes publiczny w dyskusji publicznej jest zbyt często utożsamiany z interesem jednej tylko grupy, np. konsumentów lub twórców. Tymczasem interes publiczny należy rozumieć jako równowagę interesów, będącą kompromisem między kolidującymi wartościami. Taki stan równowagi może nie być wprost postulowany przez żadną ze stron sporu i należy do niego dopiero dojść drogą konsensusu⁸²).

Art. 17 PrAut stanowi swoistą bazę dla ochrony autorskich praw majątkowych. Szereg innych przepisów regulujących sposoby korzystania z utworów i wynagradzanie za to stanowi albo uszczegółowienie tego przepisu, albo wyjątki od niego. Przepis ten przyznaje twórcy *wyłączne prawo do korzystania z utworu i rozporządzania nim na wszystkich polach eksploatacji oraz do wynagrodzenia za korzystanie z utworu*, o ile ustawa nie stanowi inaczej. Oznacza to, że analizując w każdym konkretnym przypadku sytuację prawną różnych podmiotów, jako stan domyślny należy uznać monopol autora, który ma wyłączne prawo decydowania o sposobach korzystania z utworu.

Wyłącznemu prawu twórcy zostały jednak wyznaczone pewne granice. Ustawodawca uznał, że w pewnych sytuacjach, w których interes autora zderza się z interesem innych interesariuszy lub z interesem publicznym, interes autora musi ustąpić. Przykładem takich sytuacji jest np. utwór pracowniczy (art. 12 i 13), do którego prawa autorskie majątkowe przysługują pracodawcy, chyba że co innego wynika z umowy lub utwór naukowy powstały w ramach umowy o pracę (wyjątek od reguły z art. 12: uprawnienie instytucji naukowej zatrudniającej pracownika naukowego ograniczone zostało w art. 14 jedynie do prawa pierwszeństwa publikacji; w praktyce instytucje naukowe najczęściej rezygnują z tego uprawnienia w regulaminach własności intelektualnej,

⁸² A. Michalak, *Interes publiczny i jego oddziaływanie na powstanie, treść i wykonywanie praw własności intelektualnej*, Warszawa 2012, s. 79.

wymagając jedynie umieszczenia przez autora afiliacji). Najważniejsze przepisy ograniczające wyłączność praw twórcy do utworu zawarte zostały w Oddziale 3 rozdziału 3 PrAut. Są to przepisy wprowadzające dozwolony użytek. Dozwolony użytek to taki sposób korzystania z utworu, który nie wymaga zgody twórcy. Dozwolony użytek to ogólne pojęcie, pod którym kryje się szereg ograniczeń wyłącznych praw twórcy. Gdy prawo autorskie w reakcji na rozwój techniki przyznawało twórcom kolejne uprawnienia, zarazem konstruowało odpowiednie przepisy, które te uprawnienia ograniczały. *Wraz z rozszerzaniem się katalogu uprawnień autorskich rosła także liczba ich ustawowych ograniczeń*⁸³. Uzasadnione to jest tym, że każdy utwór, niezależnie od wartości współtworzy powszechne dziedzictwo kulturowe. To uzasadnia interes publiczny. *Przez udostępnienie utworu publiczności przestaje być on wyłączną własnością autora, nawet gdyby twórca go zniszczył. Zakomunikowane publiczności dzieło pozostaje już bowiem w umysłach jego odbiorców i może nadal oddziaływać na ich wiedzę, sferę przeżyć artystycznych czy też pracę twórczą i naukową*⁸⁴.

Doktryna dzieli dozwolony użytek przede wszystkim ze względu na dwa kryteria. Ze względu na podmiot, któremu przysługują uprawnienia ograniczające monopol autorski dzieli się dozwolony użytek na osobisty i publiczny, z kolei ze względu na wymóg odpłatności dzieli się na dozwolony użytek bezpłatny i płatny (w tym ostatnim przypadku podmiot korzystający nie musi uzyskiwać zgody twórcy, ale musi wypłacić wynagrodzenie za korzystanie). Wiązą się z tymi podziałami terminy „licencja ustawowa” i „licencja przymusowa”. Funkcjonują one w języku prawnym niektórych ustawodawstw zagranicznych, np. niemieckim. W Polsce posługuje się nimi część doktryny, ale nie ma zgody ani co do znaczenia, ani co do sensu posługiwania się nimi⁸⁵.

Zasady stosowania dozwolonego użytku wyznacza tzw. test trójstopniowy, wprowadzony po raz pierwszy w roku 1971 do konwencji berneńskiej. Nie od razu jego wprowadzenie zostało docenione. Przez pierwsze lata postrzegano ten test jako wyraz kompromisu, który mając zadowolić wszystkie strony konwencji, faktycznie niewiele wnosi. Znaczenie testu zaczęło gwałtownie rosnać, gdy wprowadzano go do kolejnych umów międzynarodowych (m.in. Porozumienie TRIPS, Traktat WCT). Przełomem okazała się opinia panelu WTO z 2000 r., dotycząca dozwolonego użytku w prawie amerykańskim. *Rozstrzygnięcie panelu WTO, jak i późniejsza dyskusja wokół tej decyzji, uświadomiły jej uczestnikom jak doniosłe znaczenie może odgrywać test trójstopniowy*⁸⁶.

⁸³ Sz. Rubisz, *Istota i proces kształtowania się instytucji dozwolonego użytku w prawie autorskim*, w: „Zagadnienia prawa autorskiego”, Prace Instytutu Prawa Własności Intelektualnej UJ, z. 93, 2006, s. 175.

⁸⁴ J. Marcinkowska, *Dozwolony użytek w prawie autorskim. Podstawowe zagadnienia*, Prace Instytutu Prawa Własności Intelektualnej UJ, z. 87, s. 16.

⁸⁵ Por. J. Marcinkowska, *Dozwolony użytek w prawie autorskim. Podstawowe zagadnienia*, Prace Instytutu Prawa Własności Intelektualnej UJ, z. 87, s. 78-83.

⁸⁶ R. Sikorski, *Ocena dozwolonego użytku w prawie autorskim w świetle kryteriów testu trójstopniowego*, „Zarys prawa własności intelektualnej”, t. I „Granice prawa autorskiego”, red. J. Kępiński, K. Klafkowska-Waśniowska, R. Sikorski, Warszawa 2010, s. 26.

Test trójstopniowy wyznacza dopuszczalne granice ograniczeń praw autorskich, jakie mogą wprowadzić państwa będące stronami konwencji. Z punktu widzenia prawa międzynarodowego, jest on ograniczeniem autonomii państw. Z kolei z punktu widzenia porządku prawnego krajowego jest środkiem pomagającym dokonać właściwej interpretacji przepisów o dozwolonym użytku⁸⁷.

Zgodność z testem wymaga spełnienia łącznie trzech warunków. Po pierwsze, ograniczenie wyłączności praw twórcy ma miejsce w szczególnych, wyraźnie określonych przypadkach. Po drugie, nie narusza normalnego korzystania z dzieła. Po trzecie, nie wiąże się z nieuzasadnionym naruszeniem interesów podmiotu uprawnionego. Pierwszy warunek przesądza, że przepisy o dozwolonym użytku mają charakter wyjątku od normy polegającej na przyznaniu twórcy wyłączności do korzystania ze swojego utworu na wszystkich polach eksploatacji. Nie powinno się więc stosować przy ich interpretacji wykładni rozszerzającej ani analogii. Pozostałe dwie przesłanki są przedmiotem dyskusji. Podnosi się czasem, że prowadzą do zachwiania balansu między twórcami a użytkownikami. Drugi i trzeci warunek testu znajdują odzwierciedlenie w art. 35 PrAut: *dozwolony użytek nie może naruszać normalnego korzystania z utworu, ani godzić w słuszne interesy twórcy*.

Komentując ten przepis, E. Traple zauważa, że może on bardzo zawęzić zakres zastosowania przepisów o dozwolonym użytku: *przeniesienie testu trzystopniowego do oceny poszczególnych działań, formalnie mieszczących się w granicach ustawowo dopuszczalnego wyjątku, prowadzi do bardzo restrykcyjnego traktowania przypadków dozwolonego użytku*⁸⁸. Naruszanie normalnej eksploatacji ma miejsce w przypadku takich form korzystania z utworu, które stanowią konkurencję dla korzystania z utworu przez uprawnionego. Normalne korzystanie łączy się z (przynajmniej spodziewanymi w sposób rozsądny) przychodami. Podobnie, słuszne interesy twórcy należy rozumieć przede wszystkim jako związane z uzyskiwaniem dochodu z eksploatacji utworu.

Również S. Stanisławska-Kloc podkreśla ryzyko nieuzasadnionego pomijania interesów podmiotów innych niż uprawnieni z tytułu praw autorskich: *Nie mam wątpliwości, że dwa elementy testu zawarte w art. 35 powinny służyć tylko jako narzędzie interpretacji poszczególnych postaci dozwolonego użytku. Uważam za zasadne nadawanie mu takiej treści, aby już na etapie ustalania znaczenia terminów „normalne korzystanie” oraz „słuszne interesy twórców” widzieć zarówno interesy uprawnionych, jak i użytkowników*⁸⁹.

Analizując znaczenie przepisów o dozwolonym użytku jako podstawie przeprowadzania

⁸⁷ Tamże, s. 47.

⁸⁸ E. Traple, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. J. Barta, R. Markiewicz, wyd. 5, Warszawa 2011, s. 290.

⁸⁹ S. Stanisławska-Kloc, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, 592.

badan naukowych metodami TDM nalezy przyjrzeć się następującym przepisom, które mogą mieć znaczenie dla TDM: art. 23 (dozwolony użytek osobisty), art. 23¹ (incydentalne zwielokrotnianie), art. 27 (korzystanie m.in. przez instytucje naukowe w celach prowadzenia badan naukowych), art. 29 (prawo cytatu) oraz art. 30¹ (bazy danych spełniające cechy utworu). Wreszcie nalezy zastanowić się nad art. 34 (wymóg oznaczenia twórcy) i 35 (zakaz naruszania normalnego korzystania z utworu i naruszania słusnych interesów twórcy), które mają charakter generalny i mają wpływ na określenie treści wszystkich norm o dozwolonym użytku.

Art. 27 – w celu prowadzenia badan naukowych

Najważniejszym z punktu widzenia prowadzenia badan metodami TDM w celach naukowych jest art. 27 PrAut. Upowaznia on w par. 1 instytucje oświatowe i naukowe do korzystania z rozpowszechnionych utworów woryginale i w tłumaczeniach. Mogą to robić w celach dydaktycznych i badawczych. W tych celach mogą zwielokrotnić jedynie drobne utwory w całości lub fragmenty większych utworów. Poniżej zostaną szczegółowo omówione poszczególne warunki i ich znaczenie dla TDM. Na wstepie jednak nalezy zauwazyc, że w tym przepisie polaczono działalność dydaktyczną i naukową, co nie sprzyja czytelności. Taka redakcja przepisu sprawia, że uprawnienie do korzystania z utworów w celach badawczych przysluguje takze jednostkom oświatowym, a w celach dydaktycznych – instytucjom *stricte* badawczym.

Zakres podmiotowy

Przepis dotyczy instytucji oświatowych, uczelni oraz jednostek naukowych w rozumieniu u.z.f.n. Przed 20 XI 2015 r. przepis mowil o „instytucjach naukowych i oświatowych”. Przepis w nowym brzmieniu wskazuje zatem, w jaki sposób nalezy rozumiec instytucje naukowe. Rozróznienie uczelni i jednostek naukowych jest częściowo powtórzeniem, poniewaz – o czym niżej – jednostkami naukowymi są m.in. podstawowe jednostki organizacyjne uczelni (np. wydzialy). Uczelnie są objęte hipotezą art. 27 jako całość, niezależnie od swoich jednostek organizacyjnych. Art. 5 par. 1 pkt 13 zawiera definicję instytucji oświatowej, odwołującej się do u.s.o.

U.s.f.n. definiuje jednostkę naukową w art. 2 pkt 9. Zgodnie z nią jednostkami naukowymi są prowadzące w sposób ciągły badania naukowe lub prace rozwojowe: podstawowe jednostki organizacyjne uczelni, jednostki naukowe Polskiej Akademii Nauk, instytuty badawcze działające na podstawie ustawy o instytutach badawczych z 30 kwietnia 2010 r., międzynarodowe instytuty badawcze działające w Polsce, Polska Akademia Umiejętności oraz inne jednostki organizacyjne posiadające siedzibę na terytorium Rzeczypospolitej Polskiej, będące organizacjami prowadzącymi badania i upowszechniającymi wiedzę w rozumieniu [art. 2 pkt 83](#) rozporządzenia Komisji (UE) nr

651/2014 z dnia 17 czerwca 2014 r.

Odwołanie do rozporządzenia nadaje wyliczeniu charakter otwarty, ponieważ określa ono organizacje prowadzące badania w następujący sposób: *podmiot (...) niezależnie od jego statusu prawnego (...) lub sposobu finansowania, którego podstawowym celem jest samodzielne prowadzenie badań podstawowych, badań przemysłowych lub eksperymentalnych prac rozwojowych lub rozpowszechnianie na szeroką skalę wyników takich działań*. Decydującym dla zakwalifikowania organizacji jako jednostki naukowej jest zatem, aby podstawowym celem organizacji było samodzielne prowadzenie badań. Nie muszą to być przy tym badania podstawowe. Pojawiają się jednak dwie zasadnicze wątpliwości.

Po pierwsze, co znaczy, że prace są prowadzone samodzielnie? Czy dyskwalifikuje, bardzo często przecież spotykana, forma konsorcyjna? Taka interpretacja byłaby nazbyt restrykcyjna i niezgodna z celem przepisu. Należy go rozumieć raczej jako wymóg rzeczywistego prowadzenia badań przez tę organizację. Nie spełnia tego warunku organizacja, która będąc członkiem konsorcjów naukowych wnosi do nich wkład inny niż rzeczywiste badania (np. tylko wkład finansowy, promocyjny lub organizacyjny). Jeśli jednak w ramach konsorcjum sama prowadzi jakieś prace badawcze, choćby nie były to prace decydujące dla powodzenia projektów, należy zakwalifikować ją jako prowadzącą badania samodzielnie.

Badania powinna prowadzić sama organizacja, a nie tylko jej pracownicy lub współpracownicy. Rzecz jasna, że poszczególne czynności w ramach badań będą wykonywać pracownicy. Istotny tu jest jednak podmiot prowadzący badania. Można mieć wątpliwości, co do spełnienia warunku w sytuacji, gdy badania nie są prowadzone przez organizację, ale przez samych naukowców realizujących tylko grant pozyskany z innego źródła w oparciu o zasoby organizacji. Gdyby taka praktyka dominowała, raczej nie można by było uznać tej organizacji za prowadzącą badania w świetle omawianego przepisu.

Druga wątpliwość jest związana z określeniem „podstawowy cel”. Czy organizacja może mieć więcej niż jeden podstawowy cel? Podobnie jak w przypadku wymogu samodzielności, interpretacja nie powinna być zawężająca. Intencją przepisu jest bowiem wskazanie organizacji, dla których prowadzenie badań jest rzeczywiście istotną częścią działalności. Niewątpliwie nie spełnia przesłanek przepisu koncern mający dział badawczy, który nie jest wyodrębniony w postaci odrębnego podmiotu prawnego. Działalność badawcza dla takiego koncernu będzie miała raczej charakter dodatkowy, a nie będzie jego podstawowym celem, choćby jednym z kilku. Stracił aktualność komentarz do ustawy sprzed reformy z 2015 r.: *istnieje wiele instytucji, które działalność naukową prowadzą w zakresie marginalnym, np. muzeum zatrudniające jednego pracownika w celu prowadzenia badań historycznych. Sądzę, że uprawnieniami wynikającymi z art. 27 powinna być objęta działalność naukowa wszelkich placówek, które choćby częściowo zajmują*

się badaniami naukowymi⁹⁰. Przed nowelizacją PrAut nie odwoływało się jednoznacznie do definicji instytucji naukowej z u.z.f.n.

Wracając do art. 27 PrAut, należy podkreślić, że powoływać się na niego mogą wyłącznie jednostki naukowe. Indywidualni naukowcy nie mogą być jego beneficjentami. Do pewnego stopnia jest on odpowiednikiem dozwolonego użytku osobistego do celów naukowych. Jak zauważają komentatorzy, *w praktyce czynności korzystania będą dokonywać przede wszystkim pracownicy tych instytucji, osoby wykonujące w ramach umów cywilnoprawnych w tych instytucjach działania dydaktyczne lub biorące udział w badaniach własnych tych instytucji*⁹¹. Przepis skierowany jest do osób, *które działają w ramach szeroko rozumianej struktury instytucji naukowej lub oświatowej (lub są beneficjentami tych instytucji)*⁹². W moim przekonaniu S. Stanisławska-Kloc zbyt swobodnie rozszerzyła zakres podmiotowy przepisu o osoby współpracujące z instytucją. Ustawa wyraźnie mówi o badaniach, których podmiotem jest instytucja. Instytucja może prowadzić badania za pomocą swoich pracowników lub współpracowników. Z art. 27 mogą korzystać jednak bezpośrednio tylko instytucje, a pracownicy – tylko pośrednio. Coraz powszechniejsza praktyka prowadzenia badań w oparciu o granty stwarza nieoczywiste sytuacje. O tym, kto jest podmiotem badań decydują w tych przypadkach umowy grantowe zawierane między agencją finansującą badania, kierownikiem projektu a instytucją udostępniającą infrastrukturę i zaplecze organizacyjne.

Kończąc analizę zakresu podmiotowego art. 27 należy zauważyć, że przepis ten upoważnia do prowadzenia badań naukowych także instytucje oświatowe, choćby nie prowadziły badań.

Zakres przedmiotowy art. 27

Z punktu widzenia stosowania metod TDM szczególnie ważny jest warunek „w celu prowadzenia badań naukowych”. PrAut nie określa, jak należy rozumieć badania naukowe. Pomocą służy tutaj ponownie u.z.f.n., która w art. 2 par. 3 podaje legalną definicję badań naukowych. Dzieli je na trzy kategorie: badania podstawowe są nakierowane na zdobycie nowej wiedzy bez nastawienia na bezpośrednie zastosowanie komercyjne; badania stosowane są również nakierowane na zdobycie nowej wiedzy, ale zorientowane na jej praktyczne aplikacje; wreszcie badania przemysłowe mają na celu zdobycie nowej wiedzy i umiejętności w celu opracowania nowych produktów lub usług, często polegające na budowie prototypów lub uruchamianiu linii pilotażowych. Metody TDM mogą być wykorzystywane we wszystkich tych rodzajach badań naukowych. Co oczywiste, ale z powodu swojego znaczenia warte podkreślenia: badania naukowe

⁹⁰ E. Traple, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. J. Barta, R. Markiewicz, wyd. 5, Warszawa 2011, s. 263.

⁹¹ S. Stanisławska-Kloc, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, 412.

⁹² Tamże.

nie muszą być skuteczne. Jeżeli badanie zakończy się fiaskiem i żadna nowa wiedza nie zostanie uzyskana, wykorzystanie utworów pozostaje uzasadnione.

Przepis pozwala „korzystać z rozpowszechnionych utworów”. Przed 2004 r. przepis pozwalał korzystać tylko z utworów opublikowanych. Było to znaczące rozszerzenie zakresu przedmiotowego tej postaci dozwolonego użytku. Zgodnie z definicją z art. 6 ust. 1 PrAut „utworem rozpowszechnionym jest utwór, który za zezwoleniem twórcy został w jakikolwiek sposób udostępniony publicznie”. W zdecydowanej większości przypadków badań metodami TDM te warunki powinny zostać spełnione.

Wykorzystać można utwory tak w oryginale, jak w tłumaczeniu. Jest to ważny przepis w przypadku metod TDM, których efektem jest tłumaczenie tekstu, albo które wykorzystują automatyczne tłumaczenie jako środek do uzyskania informacji z tekstu w innym języku. Przepis ten umożliwia dokonanie automatycznego tłumaczenia i wykorzystania go do celów badań naukowych. Należy jednak pamiętać, że celem takiego działania musi być zdobycie nowej wiedzy (co jest koniecznym warunkiem uznania działania za badanie naukowe). Redakcja artykułu w żaden sposób nie sugeruje, że tłumaczenie musi być utworem pochodnym, co w przypadku automatycznych translacji budzi uzasadnione wątpliwości (por. rozdział *Integralność utworów...*).

Wreszcie art. 27 pozwala zwielokrotniać drobne utwory w całości lub fragmenty większych utworów. Cel zwielokrotnienia musi mieć charakter naukowy, czyli musi zmierzać do uzyskania nowej wiedzy. To uprawnienie jest kluczowe dla metod TDM. Największe wątpliwości prawne związane są przecież właśnie z koniecznością wytwarzania, choćby tymczasowo, kopii utworów. Ważną granicą jest możliwość zwielokrotniania co najwyżej fragmentów większych utworów. Jest to określenie nieprecyzyjne i każdorazowo należy ocenić, czy dany utwór należy do kategorii „większych utworów” i jak obszerne mogą być to fragmenty. Można przyjąć, że dla niektórych metod TDM ten wymóg jest wystarczający, ale niektóre wymagają zwielokrotnienia, przynajmniej tymczasowego, całych utworów niezależnie od ich wielkości. Oznacza to, że w praktyce możliwość powoływania się na ten przepis jako podstawy prowadzenia badań metodami TDM jest ograniczona.

Oprócz postanowień, które ustawodawca włączył do art. 27, bardzo ważne jest to, czego w nim nie ma. Przede wszystkim nie ma ograniczenia ani liczby kopii, ani czasu wykorzystania. Zwielokrotniać można zatem w takiej liczbie, jaka jest uzasadniona celem prowadzenia badań naukowych. Brak ograniczeń czasowych sprawia, że efekty pracy badawczej, w tym efekty cząstkowe, osiągnane na różnych etapach zaawansowania prac, można zarchiwizować.

Ogromną doniosłość ma fakt, że ustawodawca zdecydował się nie wprowadzić do tego przepisu wymogu wykorzystania niekomercyjnego. O problemach z odróżnieniem wykorzystania komercyjnego i niekomercyjnego zostało wspomniane w rozdziale o TDM w kontekście prawa UE.

Tutaj należy wskazać, że badania naukowe mogą być prowadzone w celach komercyjnych. Co więcej, w ostatnich latach przynajmniej w warstwie retorycznej deklaruje się potrzebę komercjalizacji badań naukowych i wzmacniania współpracy między światem akademickim a biznesem. Badania naukowe z wykorzystaniem metod TDM na pewno często mają potencjał komercyjny.

Analizowany dotychczas ust. 1 przyznaje pewną swobodę obejmującą uzyskanie dostępu do utworów przez instytucje naukowe i etapu ich analizy maszynowej. Z kolei ust. 2 art. 27 określa granice udostępniania utworu w internecie. W pierwszym zdaniu zawarty został warunek: *W przypadku publicznego udostępniania utworów w taki sposób, aby każdy mógł mieć do nich dostęp w miejscu i czasie przez siebie wybranym*. Przepis wyraźnie zabrania umieszczenia takiego utworu w otwartym internecie. Wymaga ograniczenia dostępu do ograniczonego kręgu osób, a osoby te muszą być zidentyfikowane przez instytucję prowadzącą badania. Nie mogą zatem być to osoby anonimowe. Wraca tu opisane wcześniej zagadnienie, kto jest podmiotem uprawnienia, a w związku z tym – kto jest podmiotem zobowiązany. Ponieważ uprawniona jest instytucja naukowa, a nie indywidualny naukowiec (co zostało powtórzone w ostatnim zdaniu ust. 2), obowiązek identyfikacji ciąży na tej instytucji. Wydaje się, że można wyciągnąć z tego wniosek o konieczności zarządzania informacją o tych osobach na poziomie instytucji.

Art. 23 jako podstawa prawna stosowania TDM

Na podstawie przepisu o dozwolonym użytku osobistym osoby fizyczne (co wynika z określenia „osobisty” oraz z zakresu obejmującego krąg osób pozostających w związku osobistym, co może mieć miejsce wyłącznie w przypadku osób fizycznych) mogą swobodnie korzystać z pojedynczych egzemplarzy utworu. *Cel własnego użytku osobistego nie został bliżej określony; może to być zatem cel rozrywkowy, kolekcjonerski, naukowy, archiwalny*⁹³. Zakres podmiotowy obejmuje wszystkie osoby fizyczne. W szczególności na art. 23 nie może powoływać się jednostka naukowa (por. analizę art. 27). Jeżeli metody TDM wykorzystuje zespół osób, to o ile nie zachodzi między nimi więź osobista, nie mogą uzasadniać swojej działalności art. 23. Bardzo często metodami TDM posługują się jednak zespoły badawcze. Wiąż między członkami takiego zespołu ma charakter raczej zawodowy niż osobisty.

Zakres przedmiotowy dotyczy wykorzystywania utworu w takiej postaci, do jakiej korzystający ma legalny dostęp. Może wykonywać kopie. Gdyby w trakcie procesu analizy metodami TDM miało dojść do powstania utworu pochodnego, to zastosowanie znalazłby art 2 PrAut, z którego wynika, że można stworzyć utwór pochodny bez zgody pierwotnego

⁹³ E. Traple, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. J. Barta, R. Markiewicz, wyd. 5, Warszawa 2011, s. 239.

uprawnionego. Zgoda jest potrzebna, by móc z takiego utworu korzystać. Jeżeli jednak utworem pierwotnym jest baza danych spełniająca cechy utworu, zgoda jej twórcy jest wymagana także na samo wykonanie opracowania. O specyfice korzystania z baz danych spełniających cechy utworu więcej w rozdziale poświęconym temu wątkowi.

Osoba fizyczna raczej może korzystać z utworu w celu analizy niektórymi metodami TDM. Indywidualny badacz posługujący się takimi metodami np. w celu analizy literatury przedmiotu w zakresie swoich badań, jeśli wykonuje te działania samodzielnie, może czasem powoływać się na dozwolony użytek osobisty.

Art. 23¹ : przejściowe lub incydentalne zwielokrotnianie utworu

Przepis ten umożliwia wszystkim podmiotom (nie zostało przewidziane żadne zawężenie zakresu podmiotowego) zwielokrotnianie utworów (z wyjątkiem programów komputerowych), o ile spełnione zostają łącznie następujące przesłanki: 1) zwielokrotnienie jest tymczasowe, 2) ma charakter przejściowy lub incydentalny, 3) nie ma samodzielnego znaczenia gospodarczego, a przy tym 4) stanowi integralną i zarazem niezbędną część procesu technologicznego, 5) którego wyłącznym celem jest – w przypadku TDM zastosowanie może mieć tylko punkt drugi – zgodne z prawem korzystanie z utworu. Jak widać, ten wyjątek od wyłącznych praw twórcy został przez ustawodawcę na wiele sposobów ograniczony i obwarowany wieloma restrykcyjnymi warunkami. Został dodany do polskiej ustawy w 2004 r. jako wdrożenie odpowiednich postanowień dyrektywy 2001/29. Przepisy te mają na celu umożliwienie legalnego przeglądania treści dostępnych w internecie (*browsing*) i przechowywania ich w pamięci podręcznej komputera (*caching*). Zaznajamianie się z treścią utworów na ekranie komputera (lub innego urządzenia elektronicznego) *pomimo że stanowi ono „odpowiednik” czytania w odniesieniu do utworów udostępnianych w wersji tradycyjnej (np. czytania książek), to w środowisku cyfrowym wiąże się właśnie ze sporządzeniem krótkotrwałej kopii w pamięci komputera (czytnika). O ile czytanie tradycyjne nie jest objęte monopolem twórcy, o tyle sporządzanie krótkotrwałych kopii (niezbędnych do czytania na ekranie komputera) już jest objęte treścią prawa autorskiego. [...] Do tego, aby zapewnić użytkownikom korzystającym z utworów w środowisku cyfrowym podobny poziom „ochrony” jak korzystającym w świecie analogowym, niezbędne było wprowadzenie omawianej regulacji*⁹⁴.

Ta różnica w zapewnieniu swobody w zaznajamianiu się z utworami zależnie od nośnika – analogowego lub cyfrowego – jest centralnym problemem i źródłem wątpliwości związanych ze stosowaniem metod TDM. Z tym wiąże się popularne hasło zwolenników interpretacji dopuszczających legalność metod TDM także w obecnym stanie prawnym: „the right to read is a

⁹⁴ S. Stanisławska-Kloc, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, 361.

right to mine”⁹⁵ („prawo do czytania jest prawem do maszynowego analizowania”). Hasło to upraszcza jednak szerokie spektrum sytuacji, z jakimi mamy do czynienia w przypadku TDM. Na pierwszy rzut oka ten właśnie wyjątek, skoro został pomyślany jako odpowiedź na wyzwania cyfryzacji obiegu, powinien zapewnić odpowiedni poziom bezpieczeństwa prawnego dla korzystających z metod TDM do badań naukowych.

Podążając za wyczerpującym omówieniem tych warunków w raporcie „Study of the legal framework of TDM” należy stwierdzić, że ten wyjątek znajdzie zastosowanie w metodach TDM jedynie wyjątkowo. Warunek incydentalności oznacza, że dopuszczalne jest tylko przechowywanie utworu w pamięci RAM. Jeśli kasowanie utworów nie odbywa się automatycznie, czyli w jakikolwiek sposób zależy od aktywności człowieka, to warunek przejściowości lub incydentalności nie jest spełniony. Należy jednak zauważyć, że w pewnych, raczej wyjątkowych sytuacjach, warunki te mogą zostać spełnione. Może się tak stać np. w przypadku gdy proces analizy TDM odbywa się poprzez sekwencję analizy drobnych fragmentów przechowywanych wyłącznie w pamięci RAM i natychmiast kasowanych. Jako konkluzja analizy każdego warunku dla odrębnego etapu analizy TDM (przypomnę, że autorzy raportu wyróżniają ich pięć), pada stwierdzenie: *To oznacza, że ten wyjątek nie stanowi pomocy (chyba że zupełnie wyjątkowo) dla działań związanych z analizą danych*⁹⁶.

Art. 29 – prawo cytatu

W pewnych przypadkach stosowania metod TDM można zastanawiać się, czy znajduje zastosowanie prawo cytatu. Sytuacje takie obejmują te metody, których efekty wykorzystują fragmenty analizowanych utworów, w szczególności może to być technika podsumowywania, w której najważniejsze, oddające istotne treści zdania trafiają do streszczenia.

Prawo cytatu dotyczy korzystania z efektów cudzej pracy. Wstępnym warunkiem powołania się na prawo cytatu jest więc legalny dostęp do analizowanych treści. Sam proces analizy nie może opierać się na tym przepisie. Ewentualnie więc jako cytat można rozpatrywać dopiero udostępnianie publiczności efektów analizy. Metody TDM mieszczą się w kategorii uzasadnionego celu, jakim jest analiza krytyczna lub naukowa.

Nowelizacja PrAut. z 2015 r. znacząco zmieniła prawo cytatu, m.in. otwierając katalog ustawowych celów, które uzasadniają stosowanie tej odmiany dozwolonego użytku⁹⁷. Obecnie uzasadnione celem cytatu są takie działania jak wyjaśnianie, polemika, analiza krytyczna lub

⁹⁵ Por. np. <http://sparcopen.org/news/2015/the-right-to-read-is-the-right-to-mine/>

⁹⁶ *It means that this exception will not provide much relief (or really rarely) for data analysis activities, Study on the legal framework of TDM...*, s. 50.

⁹⁷ Ta zmiana częściowo zdezaktualizowała tezy książki L. Małka *Cytat w świetle prawa autorskiego*, Warszawa 2011.

naukowa, nauczanie lub prawa gatunku twórczości. Otwarcie katalogu zdecydowanie uelastyczyło tę konstrukcję. Nie ma to jednak bezpośrednio wpływu na ewentualne powoływanie się na to prawo w związku z metodami TDM. Podpadają one bowiem bez wątpienia pod kryterium analizy naukowej.

By móc powoływać się na prawo cytatu, należy spełnić kilka warunków. Po pierwsze, cytat musi zostać umieszczony w stanowiącym samoistną całość utworze, przy czym akceptuje się opracowania takie jak książki cytatów i wybory skrzydlatych słów⁹⁸ (ale L. Małek raczej sceptycznie odnosi się do takiego uzasadnienia dla książki cytatów⁹⁹). Efekt analizy TDM zawierający cytowane urywki musi spełniać cechy utworu, czyli mieć cechy indywidualne i twórcze. Musi być też dziełem człowieka. Co do możliwości zakwalifikowania efektów (niektórych) TDM jako utworów więcej w rozdziale dotyczącym integralności utworów. Nie można wykluczyć, że kompilacja urywków utworów analizowanych może mieć twórczy lub indywidualny charakter. Wydaje się, że byłaby to jednak sytuacja raczej wyjątkowa, ponieważ kolejność przejmowanych fragmentów jest układana zazwyczaj według ogólnych, zobiektywizowanych zasad, które uwzględniają algorytmy, a nie jest aktem twórczym. Drugi warunek stawiany przed cytatem, czyli konieczność uprzedniego rozpowszechnienia utworu cytowanego, nie stanowi przeszkody dla TDM. Trzeci warunek określa granice dopuszczalności przytaczania fragmentów utworów (lub drobnych utworów w całości) uzasadnionymi celami cytatu. W otwartym, przykładowym katalogu ustawodawca wymienił m.in. analizę naukową. Także ten warunek pozostaje więc spełniony.

Przyjmuje się, że jest jeszcze przesłanka nie wyrażona dosłownie w ustawie: *elementem legalności cytatu, pomimo że literalnie nieprzewidzianym w art. 29 lub 34, jest oznaczenie („wydzielenie”, wyodrębnienie) każdego cytowanego utworu lub każdego jego fragmentu (wskazanie, gdzie zaczynają się i gdzie kończą cytowany utwór lub jego fragmenty)*¹⁰⁰. Do tego warunku odnoszą się uwagi dotyczące art. 34 PrAut.

Tym, co w większości przypadków uniemożliwia powoływanie się na prawo cytatu we włączaniu fragmentów utworów w badaniach TDM i udostępnianiu wyników jest przede wszystkim brak spełnienia kryteriów samoistnego utworu przez produkt TDM. Naukowiec może natomiast po przeprowadzeniu badań metodami TDM przygotować utwór wykorzystujący ustalenia poczynione we wcześniejszym badaniu, w którym będzie cytował prace poddane uprzednio analizie w celu np. polemiki, wyjaśnienia czy analizy krytycznej. Taki utwór nie będzie jednak efektem badań metodami TDM.

Art. 34 – obowiązek wskazania twórcy i źródła

⁹⁸ S. Stanisławska-Kloc, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, 456.

⁹⁹ L. Małek, *Cytat...*, s. 332-333.

¹⁰⁰ S. Stanisławska-Kloc, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 471.

Ten przepis należy stosować razem z każdym przepisem o dozwolonym użytku.

Wprowadza dodatkowy warunek: korzystanie z utworu wymaga podania imienia i nazwiska twórcy (zgodnie z dominującymi w doktrynie poglądami chodzi o oznaczenie autora zgodnie z jego wolą, np. w formie pseudonimu) oraz źródła. Nie jest to warunek bezwzględny. Jeśli nie ma możliwości podania tych informacji, korzystający jest zwolniony z tego obowiązku. Takie zwolnienie ma charakter wyjątkowy i powinno być interpretowane w sposób ścisły.

Charakterystyczne dla metod TDM jest posługiwanie się dużymi zbiorami danych.

Spełnienie wymogu wskazania twórcy oznacza, że np. przy analizie maszynowej tysiąca artykułów naukowych należy wskazać wszystkich autorów tych artykułów. Ze względu na zwyczaj podawania jako autorów wszystkich uczestników badań, w niektórych dyscyplinach powszechnie jest współautorstwo kilkudziesięciu, a nawet kilkuset lub więcej osób. Jeśli analizowane są zbiory danych, które są efektem wcześniejszych prac podobnego rodzaju, następuje nawarstwienie tych informacji. W praktyce informacja o wykorzystaniu w analizie TDM czyjegoś tekstu lub zbioru danych jako jednego z bardzo wielu (np. z tysiąca lub więcej) raczej nie przynosi twórcy splendoru, z jakim zazwyczaj kojarzy się podawanie autorstwa.

Warto zwrócić tu uwagę na charakter współautorstwa artykułów naukowych. Zdaniem R. Markiewicza praktykę umieszczania jako autorów tych wszystkich, którzy wnieśli wkład w proces badawczy (a niekoniecznie wnieśli twórczy wkład w powstanie artykułu) należy rozumieć jako oznaczanie współuczestnictwa pracy naukowej, a zarazem rezygnację z oznaczenia autorstwa w sensie prawnautorskim. Praktyka ta budzi jednak wątpliwości, ponieważ PrAut nie ma na celu ochrony wyników badań, ale ochronę twórczości¹⁰¹.

Formalne spełnienie wymogu z art. 34 nie powinno nastęrczać dużych trudności technicznych, choć może być odbierane jako niepotrzebne obciążenie, nie przynoszące korzyści twórcom. Informacje takie mogą zostać umieszczone w bazie danych zawierającej wyłącznie informacje o autorach wykorzystanych utworów i źródłach.

Art. 35 – zakaz naruszania normalnego korzystania i godzenia w słuszne interesy twórcy

Podobnie jak art. 34, art. 35 należy uwzględniać w każdym przypadku stosowania dozwolonego użytku. Zawiera drugie i trzecie kryterium testu trójstopniowego. Zasady jego interpretacji w kontekście TDM zostały przedstawione w rozdziale o wolności badań naukowych. Jest to bardzo ważny przepis wyznaczający granice monopolu autorskiego i zakresu działań, które można podjąć bez uzyskiwania zgody od uprawnionego. Art. 35 zawiera dwie klauzule generalne, które sprawiają, że w każdy konkretny przypadek należy oceniać indywidualnie ważąc interes

¹⁰¹ R. Markiewicz, *Ochrona prac naukowych*, „Zeszyty Naukowe UJ”, z. 55/1990, s. 50-53.

autora, badacza stosującego metody TDM oraz społeczeństwa. Zasada ochrony praw twórcy zderza się z zasadą wolności badań naukowych.

Kończąc rozdział o dozwolonym użytku należy podkreślić potrzebę określenia formalnego podmiotu prowadzącego badania metodami TDM: czy jest to osoba fizyczna, która może powoływać się na dozwolony użytek osobisty (z ograniczeniami stąd wynikającymi, tj. możliwością korzystania co najwyżej przez krąg osób pozostających w związku osobistym, a nie zawodowym, w szczególności naukowym), czy też jest to instytucja naukowa, która może powoływać się na art. 27. W tym ostatnim przypadku możliwość zwielokrotniania utworów ograniczona jest do fragmentów większych utworów lub drobnych utworów w całości. Jest to poważne ograniczenie, ponieważ obecnie metodami TDM posługują się przede wszystkim zespoły badawcze, a wiele metod TDM wymaga na jakimś etapie zwielokrotniania utworów niezależnie od ich wielkości.

Dozwolony użytek a analiza baz danych o cechach utworu

Stosowanie metod TDM w wielu przypadkach wymaga korzystania z baz danych, także z baz danych spełniających cechy utworu. Poza zakresem niniejszej pracy pozostaje ochrona baz danych na podstawie u.o.b.d., która w art. 2 zawiera legalną definicję bazy danych: *baza danych oznacza zbiór danych lub jakichkolwiek innych materiałów i elementów zgromadzonych według określonej systematyki lub metody, indywidualnie dostępnych w jakikolwiek sposób, w tym środkami elektronicznymi, wymagający istotnego, co do jakości lub ilości, nakładu inwestycyjnego w celu sporządzenia, weryfikacji lub prezentacji jego zawartości*. Bazy danych mogą podlegać podwójnej ochronie. Na podstawie ustawy o ochronie baz danych chroniony jest nakład inwestycyjny poniesiony w celu utworzenia bazy. Oprócz tego, jeśli baza danych spełnia cechy utworu, czyli jest przejawem działalności twórczej o indywidualnym charakterze, uzyskuje ochronę również na gruncie PrAut., co jednoznacznie przesądza art. 1 ust. o ochronie baz danych. W niniejszej pracy interesuje mnie wyłącznie aspekt ochrony na gruncie PrAut.

Przepisy PrAut dotyczące baz danych spełniających cechy utworu są rozrzucone po całej ustawie. Jest ich zaledwie kilka i mają charakter wyjątków od przepisów PrAut dotyczących utworów w ogólności. Część wyjątków zawęża uprawnienia korzystającego z bazy danych, a część je rozszerza. Art. 2 PrAut określa relację pomiędzy utworem a jego opracowaniami. Ogólna zasada przyznaje korzystającemu z utworu swobodę sporządzania opracowań (np. tłumaczenie lub adaptację), a jedynie rozporządzenie i korzystanie z opracowania wymaga zgody autora utworu pierwotnego. Ust. 2 zd. drugie wprowadza wyjątek dla baz danych. Już samo sporządzenie opracowania bazy danych wymaga zgody. Jeśli metody TDM wymagają sporządzenia opracowania bazy danych, choćby takie opracowanie było potrzebne wyłącznie na krótki czas jako przejściowy etap wydobywania informacji, należy uznać, że przepis ten uniemożliwia legalne badanie takimi metodami.

Art. 3 przyznaje ochronę zbiorom, antologiom, wyborom i bazom danych jako całościom, niezależnie od istnienia lub nieistnienia ochrony przyznanej materiałom umieszczonym w takiej bazie. Chroniony zatem może być układ, dobór lub zestawienie, o ile mają twórczy charakter. Ta ochrona nie odbywa się jednak z uszczerbkiem dla praw do wykorzystanych w tej bazie materiałów.

Art. 17¹ upoważnia legalnego użytkownika bazy danych do stworzenia opracowania lub zwielokrotnienia tej bazy, jeśli jest to konieczne dla dostępu do zawartości i normalnego korzystania. Ten przepis pełni podobną funkcję, co art. 23¹, który pozwala tymczasowo

zwielokrotnić utwór w celu zgodnego z prawem korzystania z niego (por. rozdział o dozwolonym użytku). Upoważnienie dla użytkownika bazy danych jest znacznie bardziej liberalne. Pozwala nie tylko na zwielokrotnienie, ale również na stworzenie opracowania. Nie jest obwarowane, w przeciwieństwie do ogólnej zasady, szeregiem restrykcyjnych warunków.

Interesujące jest umiejscowienie przepisu. Znajduje się on bowiem w Rozdziale 3 „Treść prawa autorskiego”, a nie Rozdziale 4 „Dozwolony użytek chronionych utworów”, choć treściowo bardziej pasuje do rozdziału 4. Można wyciągnąć z tego wniosek, że stanowi on specyficzne ograniczenie monopolu autorskiego, inne niż dozwolony użytek. Czy w związku z tym mają do niego zastosowanie art. 34 i 35? Warunek normalnego korzystania, który pojawia się w art. 35, został powtórzony w art. 17¹. Nie ma natomiast warunku nie godzenia w słuszne interesy autora. Oczywiście, nie można dowodzić, że art. 17¹ pozwala w szerokim znaczeniu godzić w słuszne interesy autora (takie działanie byłoby nadużyciem prawa i spełniałoby hipotezę art. 5 KC). Warunek z art. 35 jest interpretowany przez doktrynę jako zakaz korzystania z utworu w sposób ograniczający czerpanie dochodu z utworu. Art. 17¹ jest być może w tym względzie mniej rygorystyczny.

Poważniejsza wątpliwość związana z umiejscowieniem art. 17¹ dotyczy przyjętej w odniesieniu do dozwolonego użytku zasady zakazującej interpretację rozszerzającą i stosowanie analogii ze względu na pozostawanie przepisów o dozwolonym użytku w relacji przepis szczególnie – przepis ogólny. Przepisem ogólnym jest zaś monopol autorski. Czy umiejscowienie art. 17¹ w rozdziale o treści prawa autorskiego sprawia, że nie stosuje się do niego ten zakaz?

Duże znaczenie dla korzystania z baz danych, także w celach analizy TDM, ma art. 23 wprowadzający instytucję dozwolonego użytku osobistego. Przepis ten *in fine* wyłącza z zakresu swojego działania „elektroniczne bazy danych spełniające cechy utworu, chyba że dotyczy to własnego użytku naukowego niezwiązanego z celem zarobkowym”. Oznacza to daleko idące ograniczenie korzystania z elektroczytnych baz danych do własnego użytku osobistego. Zwrot „chyba że” oznacza, oprócz wszystkiego innego, odwrócenie ciężaru dowodu. W tym przypadku jednak ciężar dowodu i tak spoczywa na korzystającym, ponieważ dozwolony użytek jest wyjątkiem od zasady monopolu autorskiego.

Aby móc korzystać z elektronicznych baz danych w oparciu o ten przepis należy spełnić łącznie trzy przesłanki. Użytek musi być „własny”, „naukowy” oraz „niezwiązany z celem zarobkowym”.

Znaczenie tych przesłanek okazuje się nieoczywiste. E. Traple stwierdza wprost, że wykładnia literalna jest niedopuszczalna: *Ścisła interpretacja określenia „własnego użytku naukowego niezwiązanego z celem zarobkowym” prowadziłaby do wniosku, że wykorzystywanie baz*

danych do celów badań naukowych prowadzonych przez konkretnego naukowca zatrudnionego w instytucie badawczym przekracza ramy wyznaczone w art. 23. Taka interpretacja wydaje się jednak nieuprawniona. Wykładnia celowościowa prowadzi do wniosku, że z zakresu dozwolonego użytku naukowego wyłączone tylko takie korzystanie z baz danych, które następuje w celach handlowych, a więc w pracy na zamówienie podmiotu gospodarczego¹⁰². Również S. Stanisławska-Kloc, choć odnosi się wyłącznie do przesłanki użytku naukowego, zdecydowanie odchodzi od wykładni literalnej: *Wydaje się, że użytek naukowy powinien być rozumiany zarówno jako związany z procesem nauki – edukacji (własnej – uczenia się, przygotowywania się do zajęć), jak i dotyczący działalności naukowej (własnej) związanej z prowadzeniem badań naukowych, o ile nie mają one celu zarobkowego*¹⁰³.

Analizując ten przepis należy pamiętać o art. 17¹, który pozwala na „normalne korzystanie”. Wąskie uregulowanie zakresu dozwolonego użytku osobistego w stosunku do elektronicznych baz danych, nawet gdyby oceniać je negatywnie, nie powinno jawić się jako rozwiązanie absurda.

Warunek „własnych” badań nie powinien być rozciągany na badania prowadzone przez instytucje, w szczególności zespoły badawcze. Należy pamiętać, że przepis mówi o użytku osobistym. Instytucje naukowe mogą stosować art. 27, w którym nie ma ograniczenia dla korzystania z baz danych. Nie chodzi zatem o miejsce, w którym wykorzystuje się utwór, ale o to, kto jest podmiotem tego działania. Jeśli indywidualna osoba, to może korzystać z ograniczeniami wynikającymi z art. 23, a jeśli instytucja naukowa – zakres dozwolonego użytku wyznacza art. 27. Art. 23 nie wprowadza żadnego innego ograniczenia podmiotowego, poza tym, że musi to być osoba fizyczna. Na ten przepis może powołać się każdy, kto „korzysta w sposób naukowy” z utworu. Nie musi to być ani naukowiec (osoba zatrudniona w instytucji naukowej), ani osoba o statusie studenta lub doktoranta. Z powodu braku zawężających wymogów podmiotowych ważna jest przesłanka „użytku naukowego”. Interpretacja S. Stanisławskiej-Kloc rozszerza to sformułowanie na cele edukacyjne, jako „związane z procesem nauki”. Być może takie rozszerzenie dozwolonego użytku byłoby pożądane. W obecnym kształcie nie znajduje jednak uzasadnienia w przepisach. Argument w tym przypadku opiera się na sztuczce słownej i wykorzystuje dwuznaczność wyrażenia „proces nauki”. Nauczanie to jednak wyraźnie odrębna aktywność od działalności naukowej. Nauczanie polega na przekazywaniu wiedzy już ustalonej. Działalność naukowa to działania mające na celu uzyskanie nowej wiedzy. Nawet gdyby rozumieć „użytek naukowy” szerzej niż tylko jako badania naukowe (zdefiniowane w u.z.f.n., co zostało już

¹⁰²

E. Traple, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. J. Barta, R. Markiewicz, wyd. 5, Warszawa 2011, s. 241.

¹⁰³ S. Stanisławska-Kloc, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 347.

wcześniej omówione), to musi być w jakiś sposób z badaniami naukowymi powiązany. Musi być więc, choćby pośrednio, nastawiony na pozyskiwanie nowej wiedzy. Ustawodawca konsekwentnie w PrAut rozdziela działalność naukową od edukacyjnej. Robi tak także w różnych przepisach konstytuujących dozwolony użytek. Przyjęcie, że tylko w tym jednym przepisie postąpił niekonsekwentnie byłoby wyrazem zbyt swobodnej wykładni.

Ostatni warunek ogranicza korzystanie z elektronicznych baz danych bez zezwolenia twórcy tylko do celów nie związanych z celami zarobkowymi. Jest to warunek będący odpowiednikiem anglojęzycznego wykorzystania niekomercyjnego („non commercial purpose”) lub niemieckiego użytku handlowego („gewerblicher Zweck”). Linia podziału pomiędzy użytkowaniem komercyjnym a niekomercyjnym jest jedną z najtrudniejszych do przeprowadzenia. Skutki przeprowadzenia granicy w taki lub inny sposób ma daleko idące konsekwencje. Dlatego ważne są niuanse językowe. Słownik Języka Polskiego PWN podaje trzy definicje słowa „zarobek”: „zapłata za wykonaną pracę”, „możliwość zarobienia pieniędzy” oraz „zysk osiągnięty z transakcji handlowej”¹⁰⁴. Pierwsza z tych definicji odnosi się do stosunków pracowniczych, druga definicja w zasadzie też dotyczy sfery stosunków pracowniczych (por. definicję „zarobienie”), wreszcie trzecia definicja dotyczy głównie relacji między przedsiębiorcami.

Dozwolony użytek osobisty jest adresowany do osób fizycznych. Z tego względu wiązanie go z celami zarobkowymi wydaje się logiczne i spójne. Oznacza jednak, że naukowiec może korzystać z baz danych w celach naukowych o ile nie są związane z prowadzeniem badań, za które dostaje wynagrodzenie (niezależnie od formy zatrudnienia: umowa o pracę lub umowy cywilnoprawne). Art. 23 upoważnia go do korzystania z elektronicznych baz danych spełniających cechy utworu w sposób nie związany z obowiązkami naukowymi.

¹⁰⁴ <http://sjp.pwn.pl/sjp/zarobek;2543837.html>

Integralność utworów analizowanych metodami TDM

Proces maszynowej analizy tekstu składa się z wielu działań, w czasie których dochodzi najpierw do kopiowania, a następnie do przekształcania materiałów do postaci umożliwiającej analizę. Przekształcanie polega np. na ujednoznacznianiu: usuwaniu znaków interpunkcyjnych, zmienianiu wszystkich liter na małe, zamianie różnych form gramatycznych wyrazu na jedną. Fragmenty wywołujące szum (np. często występujące słowa nie niosące ze sobą istotnego znaczenia) są usuwane. W kolejnym kroku, zależnie od stosowanych metod, może powstać baza zawierająca zidentyfikowane fakty, a wykorzystane wersje (zniekształconych) utworów zostają usunięte (mogą też jednak zostać zarchiwizowane w celu umożliwienia późniejszej weryfikacji wyników lub ponownego wykorzystania, co jest coraz częściej wymogiem agencji finansujących badania naukowe). Celem metod TDM może być też jednak uzyskanie wersji utworów wzbogaconych o pewne informacje (np. o informacje gramatyczne czy leksykalne). Takie wzbogacone wersje zostają utrwalone i mogą stanowić podstawę dalszych badań. Mogą być udostępniane różnemu kręgowi odbiorców, w tym nieokreślonymu kręgowi odbiorców. Przykładowo przedmiotem dodawania informacji mogą być utwory literackie znajdujące się w domenie publicznej.

Trzeba w związku z tym odpowiedzieć na pytanie, czy w związku z tym nie ma ryzyka naruszenia autorskich dóbr osobistych, a w szczególności ingerencji w integralność utworu? Czy omawiane metody analizy nie stoją w sprzeczności z art. 16 PrAut? Ochrona dóbr osobistych autorskich jest niezależna od praw majątkowych, co oznacza, że ewentualne uzyskanie zgody na analizę TDM od uprawnionych z tytułu praw autorskich majątkowych nie może zastąpić zgody podmiotów uprawnionych z tytułu autorskich praw osobistych, gdyby taka zgoda okazała się niezbędna. Podobnie przepisy o dozwolonym użytku ograniczają prawa (alternatywnie: wykonywanie praw) majątkowe, nie wpływając na ochronę dóbr osobistych. Autorskie dobra osobiste są nieograniczone w czasie i niezbywalne.

Problem ochrony dóbr osobistych został zauważony przy okazji korpusów językowych. *Analizując możliwość umieszczenia utworu (lub fragmentu utworu) w korpusie, należy również mieć na względzie to, że twórcy korpusu, włączając dany utwór do zbioru, działają poniekąd w braku poszanowania owej szczególnej więzi łączącej dzieło z jego autorem. Dla twórców korpusu sam pojedynczy utwór nie przedstawia bowiem żadnej wartości, dopiero zbiór tysięcy utworów*

świadczy o wartości korpusu jako takiego¹⁰⁵. Autorzy zastanawiają się nad ryzykiem nierzetelnego wykorzystania oraz ingerencji w integralność. Potencjalne ryzyko nierzetelnego wykorzystania upatrują w subiektywnym dyskomforcie twórcy, wynikającym z poczucia uwypuklenia nieistotnych aspektów jego dzieła, ewentualnie z oceną sprawności posługiwania się przez niego językiem¹⁰⁶. Jeżeli zaś chodzi o ryzyko ingerencji w integralność w procesie anotacji, to stwierdzają, że standardowe w czasie przygotowywania korpusu usuwanie równań matematycznych, tabel, ilustracji, odwołań bibliograficznych i dłuższych fragmentów obcojęzycznych niewątpliwie jest ingerencją w integralność dzieła. *Ryzyko takiego uznania jest tym większe, że autorzy zamieszczanych w korpusie utworów nie opracowują ich specjalnie w celu zamieszczenia w korpusie (...). Stąd trudno przyjąć, że zmiany w utworach wprowadzane przez twórców korpusu są zmianami, którym autor (...) nie mógłby się sprzeciwić*¹⁰⁷. Ostatecznie A. Górską i R. Górski uznają, że wyraźne zaznaczanie miejsc, w których znajdowały się usunięte fragmenty, za wystarczającą przesłankę do uznania, że do naruszenia nie dochodzi, przez analogię do praktyki stosowania cytatu. Powszechna praktyka dopuszcza zaznaczanie wielokropkiem w nawiasie opuszczone fragmenty.

Prawo do integralności utworu jest jednym z autorskich dóbr osobistych. Należy zrekonstruować znaczenie literalne przepisu. Nie można jednak na tym poprzestać, ponieważ ustawa nie podaje precyzyjnych wskazówek, jak rozumieć dobra osobiste (a wręcz przeciwnie – ze względu na krytykowaną redakcję przepisu wprowadza pewne pojęciowe zachwianie, co do przedmiotu ochrony: czy jest nim dobro osobiste, czy prawo do ochrony tego dobra? *Prawidłowa wykładnia art. 16 pr. Aut. z 1994 r. jest znacznie utrudniona z uwagi na rażąco niedbale sformułowanie tego przepisu*¹⁰⁸). Charakter dóbr osobistych sprawia, że podlegają one bezustannej ewolucji. Niezbędna będzie więc wykładnia funkcjonalna i celowościowa, która wyjaśni zakładane cele przepisu.

Autorskie a powszechne dobra osobiste

Jedną z nierozstrzygniętych w doktrynie kontrowersji jest relacja autorskich dóbr osobistych do dóbr osobistych powszechnego prawa cywilnego. *W literaturze przedmiotu zwraca się (...) uwagę na problematyczną relację pomiędzy dobrami osobistymi prawa cywilnego (z art. 23 k.c.) oraz autorskimi dobrami osobistymi (art. 16). Poglądy polaryzują się wokół dwóch koncepcji. Według pierwszej z nich dobra osobiste autorskie są szczególnymi dobrami osobistymi prawa cywilnego (stanowią zatem niewymienioną wprost w art. 23 kategorię dóbr osobistych), druga*

¹⁰⁵ A. Górską, R. Górski, *Tworzenie i udostępnianie korpusów tekstowych...*, s. 60.

¹⁰⁶ Tamże, s. 61-62.

¹⁰⁷ Tamże, s. 62.

¹⁰⁸ A-M. Niżankowska, *Prawo do integralności utworu*, Warszawa 2007, s. 60.

zakłada, że są odmienną kategorią dóbr osobistych. Obecnie większość zwolenników wyznaje pogląd pierwszy¹⁰⁹. Kontrowersja ta nie ma charakteru wyłącznie teoretycznego, ale ma przełożenie na praktykę. Od odpowiedzi na pytanie o tę relację zależy możliwość dochodzenia przez twórcę ochrony równolegle na podstawie przepisów kodeksowych i prawnoautorskich. *Uznanie ich za odrębną kategorię możliwości takiej nie daje*¹¹⁰. Przekonanie o tym, że autorskie dobra osobiste są szczególnym przypadkiem powszechnych dóbr osobistych wyrażał SN. Takie stanowisko nie jest jednak przekonujące. Przeciwną opinię wyrażał m.in. J. Błeszyński, zwracając uwagę, że przedmioty obu rodzajów dóbr są różne: *Moim zdaniem na podstawie art. 23-24 k.c. chroniona jest swoboda rozwijania pewnej działalności. Prawo autorskie natomiast chroni określony rezultat działalności twórczej, jakim jest dzieło*¹¹¹. Należy jednak zwrócić uwagę na fakt, że art. 78 PrAut przyznaje twórcom, w przeciwieństwie do ustawy z 1952 r., roszczenia służące ochronie jego dóbr osobistych podobne do tych z art. 24 KC. Zabieg ten zmniejszył praktyczne znaczenie dyskusji o relację między oboma rodzajami dóbr osobistych¹¹².

K. Czub analizował zagadnienia konstrukcyjne praw osobistych twórców dóbr niematerialnych¹¹³. Główne zadanie, jakie przed sobą postawił, dotyczyło możliwości ujęcia tych dóbr w sposób uniwersalny. Roboczo nazywa te dobra „intelektualnymi dobrami osobistymi”¹¹⁴. Odpowiedź na pytanie, na ile wartościowe dla nauk prawnych jest wprowadzenie takiej nowej kategorii może pozostać na uboczu moich rozważań. Ważnym założeniem K. Czuba jest teza o tym, że autorskie prawa osobiste nie są odrębne w stosunku do „ogólnych” dóbr osobistych. Ta teza jest przedmiotem trwającej od wielu lat kontrowersji. Moim zdaniem dobra te są odrębne, choćby z powodu czasu trwania. Autorskie dobra osobiste w polskim prawie autorskim są nieograniczone w czasie, w przeciwieństwie do dóbr osobistych z KC, które trwają tylko do śmierci człowieka.

Dyskutowany w doktrynie od dawna jest charakter dóbr osobistych powszechnego prawa cywilnego. Podstawowy spór rozgrywał się pomiędzy zwolennikami koncepcji subiektywnej, a zwolennikami obiektywnej, czy, lepiej powiedzieć, obiektywizującej. Koncepcja subiektywna, którą na gruncie polskiego prawoznawstwa rozwinął S. Grzybowski, koncentruje się na stanie życia psychicznego człowieka, wiąże dobro osobiste z „kondycją emocjonalną uprawnionego”¹¹⁵. Już jednak S. Grzybowski, jak również inni zwolennicy koncepcji subiektywnej (m.in. A. Wolter, S. Szer), zauważali potrzebę odniesienia tych subiektywnych przeżyć uprawnionego do pewnej

¹⁰⁹ S. Stanisławska-Kloc, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 220-221.

¹¹⁰ A-M. Niżankowska, *Prawo do integralności utworu*, Warszawa 2007, s. 57.

¹¹¹ J. Błeszyński, *Prawo autorskie*, wyd. 3, Warszawa 1988, s. 117-118.

¹¹² Odpowiadając na wątpliwość, czy twórca może dochodzić roszczeń powołując się na art. 448 KC należy zwrócić uwagę, że art. 23 KC dotyczy „dóbr osobistych człowieka”, a art. 448 mówi o naruszeniu „dobra osobistego”. „Dobro osobiste” z art. 448 obejmuje zatem „dobro osobiste człowieka” i „autorskie dobro osobiste”.

¹¹³ K. Czub, *Prawa osobiste twórców dóbr niematerialnych. Zagadnienia konstrukcyjne*, Warszawa 2011.

¹¹⁴ Tamże, s. 20.

¹¹⁵ K. Czub, *Prawa osobiste twórców dóbr niematerialnych. Zagadnienia konstrukcyjne*, Warszawa 2011, s. 26.

zewnętrznej miary, takiej jak stopień rozwoju psychicznego. Na drugim biegunie sporu znajdowali się zwolennicy koncepcji obiektywnej. A. Szpunar to stanowisko podsumował następująco: dobra osobiste są niemajątkowymi wartościami powiązаныmi z osobowością istoty ludzkiej, powszechnie uznawanymi w danym społeczeństwie¹¹⁶. Zgodnie z tą koncepcją, która jest obecnie uznawana za słuszniejszą przez większość przedstawicieli doktryny, o naruszeniu dóbr osobistych świadczy nie indywidualne przeżycie psychiczne poszkodowanego, ale porównanie ze wzorcem względem niego zewnętrznym. Nie można jednak zupełnie abstrahować od tych przeżyć, dlatego koncepcje te lepiej nazywać obiektywizującymi, niż obiektywnymi. Sądy rozpatrujące spory o naruszenie dóbr osobistych przychylają się z reguły do koncepcji obiektywizujących¹¹⁷. Indywidualne przeżycia są trudno weryfikowalne. Co więcej, uzależnienie naruszenia dóbr osobistych od poczucia krzywdy wyłącza spod ochrony dobra, których istnienia uprawnieni nie są świadomi.

Doktryna, przychylając się do koncepcji obiektywizującej, na ogół zgadza się też, że dobra osobiste są przykładem praw podmiotowych¹¹⁸. Pomijając w tym miejscu dyskusję o konstrukcji dóbr podmiotowych, należy zwrócić uwagę, że określają one sferę możliwości podmiotu uprawnionego. W granicach wyznaczonych przez prawo podmiotowe, uprawniony może działać, a wszyscy muszą to działanie znosić. Konstrukcja praw podmiotowych wzmacnia pozycję prawną uprawnionego w stosunku do zestawu nakazów i zakazów.

Jak wyżej wspomniałem, K. Czub wychodzi z założenia, przyłączając się w tym poglądzie do znacznej części doktryny, że autorskie dobra osobiste są szczególną postacią dóbr osobistych powszechnego prawa cywilnego¹¹⁹. Stwierdza, że dobra osobiste twórcy dóbr niematerialnych *dają się (...) sprowadzić do węzła uczuciowego łączącego danego twórcę ze stworzonym przez niego konkretnym dobrem o charakterze intelektualnym. Więż ta jest emanacją osobowości twórcy, jego godności*¹²⁰. Określanie węzła łączącego twórcę z jego utworem ma długą tradycję (*już S. Ritterman pisał, charakteryzując autorskie prawa osobiste, że chronią węzeł uczuciowy, jaki łączy twórcę z jego dziełem, przed wszelkimi bezzasadnymi atakami, które mogłyby uczucia twórcy do jego dzieła zadrasnąć (S. Ritterman, Komentarz..., s. 77)*¹²¹). S. Stanisławska-Kloc niuansuje znaczenie tej więzi. Stwierdza bowiem dalej, że ta więź, upraszczając, stanowi warunek *sine qua non* każdego autorskiego dobra osobistego. Pojawia się tu pytanie, czy wnioski byłyby takie same, gdyby nie upraszczać analizy? A także, czy oprócz tego, że jest to warunek konieczny, to czy jest to także warunek wystarczający do powstania ochrony?

Nie zgadzam się, że dobra osobiste niematerialne, także autorskie, można sprowadzić do

¹¹⁶ Za: K. Czub, *Prawa osobiste twórców dóbr niematerialnych. Zagadnienia konstrukcyjne*, Warszawa 2011, s. 26.

¹¹⁷ Por. K. Czub, *Prawa osobiste...*, s. 28-31.

¹¹⁸ Tamże, s. 40.

¹¹⁹ Tamże, s. 53.

¹²⁰ Tamże, s. 53.

¹²¹ S. Stanisławska-Kloc, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, 225.

węzła uczuciowego. Jest to niewątpliwie bardzo ważny aspekt, być może nawet najważniejszy. Węzeł uczuciowy łączący twórcę z jego dziełem nie może trwać jednak dłużej niż życie twórcy. Wraz ze śmiercią człowieka zanikają wszelkie więzy uczuciowe. Trwanie autorskich dóbr osobistych pomimo śmierci twórcy można uzasadniać tylko i wyłącznie tym, że więź twórcy z utworem, o której mówi art. 16 PrAut nie ma charakteru uczuciowego, ale jest pewnym obiektywnym stanem rzeczy. Interes w ochronie tego obiektywnego stanu rzeczy ma nie tylko twórca, ale także społeczeństwo (notabene w wielu sytuacjach twórcy mają interes w odstąpieniu od tej ochrony, np. w przypadku tzw. *ghostwritingu*; zakaz *ghostwritingu* wynika z ochrony interesu publicznego).

Podkreślenie obiektywnego charakteru więzi łączącej twórcę z utworem pozwala znaleźć „trzecią drogę” w dyskusji o charakterze praw osobistych twórcy po jego śmierci. *Czy może istnieć prawo podmiotowe bez podmiotu prawa*^{122?} - pyta S. Stanisławska-Kloc. Odstąpienie od postrzegania autorskich dóbr osobistych wyłącznie w kategoriach praw podmiotowych umożliwiłoby odejście od koncepcji fikcji prawnej, jakoby po śmierci autora nadal istniały jego dobra osobiste rozumiane jako jego prawa podmiotowe¹²³.

Nie ma w tym miejscu, niestety, możliwości na rozwinięcie tego wątku i wyczerpujące przeanalizowanie skomplikowanej tematyki konstrukcji autorskich dóbr osobistych. W tym miejscu należy przyjąć następującą konstatację: autorskie dobra osobiste, odrębne od powszechnych dóbr osobistych, chronią obiektywnie istniejącą (a więc nie ograniczającą się do więzi uczuciowej) więź twórcy z utworem.

Dobro osobiste jako pojęcie dynamiczne

Dobra osobiste są pojęciami dynamicznymi. Przemiany społeczno-gospodarcze odciskają się na nich zarówno w ten sposób, że powstają nowe dobra osobiste (katalogi z art. 23 KC oraz art. 16 PrAut są otwarte), ale również w ten sposób, że znaczenie dotychczasowych dóbr ulega ewolucji. Sądy wielokrotnie potwierdzały dynamiczny charakter powszechnych dóbr osobistych (np. wyrok SN z dn. 10.06.1977 r., II CR 187/77). *Pojęcie to [dóbr osobistych - MS] należy odnosić do określonego poziomu rozwoju technologicznego i cywilizacyjnego, przyjętych w społeczeństwie zasad moralnych i prawnych, istniejącego rodzaju stosunków społecznych, gospodarczych czy nawet politycznych. Pojęcie naruszenia określonego dobra osobistego jest również pojęciem dynamicznym w czasie i dotyczącym konkretnych środowisk*¹²⁴.

Mimo odrębności, w moim przekonaniu, autorskich dóbr osobistych, tezy te znajdują zastosowanie i do nich. Zakres i sposób ochrony więzi twórcy z jego dziełem musi odpowiadać

¹²² S. Stanisławska-Kloc, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 219.

¹²³ Por. np. J. Barta, R. Markiewicz, *Prawa autorskie i prawa pokrewne*, wyd. 6, Warszawa 2014, s. 49-50.

¹²⁴ Uchwała SN w składzie siedmiu sędziów z dn. 16.07.1993 r., I PZP 28/93

polom eksploatacji tego dzieła oraz możliwym naruszeniom.

Prawo do autorstwa

Jak zostało już wspomniane w rozdziale dotyczącym stosowania metod TDM na podstawie przepisów o dozwolonym użytku, wymóg oznaczenia autorstwa może być uciążliwy, ale nie powinien zazwyczaj nastęrczać trudności sprawiających, że uzasadniona stawałoby się powołanie na zdanie drugie art. 34 PrAut, pozwalające nie podawać imienia i nazwiska autora, gdy przekracza to istniejące możliwości. Do analizowanych zbiorów danych można dołączyć plik opisujący autorstwo każdego wykorzystanego materiału. Taki obowiązek nie ma jednak faktycznie znaczenia, jakie ma oznaczenie autorstwa w przypadku tradycyjnych form eksploatacji. Podanie autora jako jednego z wielu tysięcy twórców dzieł poddawanych analizie nie wpływa ani pozytywnie, ani negatywnie na postrzeganie twórcy i jego dzieła przez odbiorcę analiz TDM.

Integralność utworu

PrAut przyznaje ochronę w zakresie integralności utworu. Art. 16 pkt 3 tej ustawy wskazuje, że jednym z praw (dóbr) osobistych jest nienaruszalność treści i formy utworu oraz jego rzetelnego wykorzystania. Doktryna zazwyczaj rozpatruje to jako dwa odrębne prawa (dobra): prawo do integralności i prawo do rzetelnego wykorzystania¹²⁵. Wydaje się, że ten przepis można rozumieć także inaczej: jako jedno dobro, które zakłada spełnienie dwóch warunków. Wątpliwości budzi konstrukcja listy przykładowych praw: w każdym punkcie wymienione zostało jedno dobro osobiste twórcy (autorstwo, oznaczenie utworu nazwiskiem lub pseudonimem, decydowanie o pierwszym udostępnieniu publiczności, nadzór nad sposobem korzystania z utworu). Wyjątkiem – zgodnie z przyjętą w doktrynie interpretacją – ma być nienaruszalność treści i formy oraz rzetelne wykorzystanie. Czy ustawodawca nie miał intencji opisanego jednego dobra, które łącznie obejmowałoby nienaruszalność i rzetelne wykorzystanie?

Nierzetelne wykorzystanie wiąże się ze sposobem szeroko rozumianej niewłaściwej prezentacji (udostępnienia) utworu publiczności, która narusza więź twórcy z utworem¹²⁶. Wydaje się zatem, że istotnym elementem jest sposób, w jaki publiczność postrzega dzieło ze względu na formę jego udostępnienia. Prawo to należy rozumieć jako możliwość zakazywania wykorzystania dzieła w sposób wprowadzający w błąd co do charakteru i podstawowych przyjętych w nim kryteriów wartości¹²⁷. Należy oceniać m.in. kontekst eksploatacji lub tendencyjny wybór fragmentów. Wydaje się, że metody TDM stosowane w celach naukowych nie stwarzają ryzyka

¹²⁵ Por. np. J. Barta, R. Markiewicz, *Prawa autorskie i prawa pokrewne*, wyd. 6, Warszawa 2014, s. 54-56.

¹²⁶ S. Stanisławska-Kloc, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 243-244.

¹²⁷ J. Barta, R. Markiewicz, *Prawo autorskie...*, s. 55.

nierzetelnego wykorzystania.

Inaczej przedstawia się sytuacja w przypadku prawa do nienaruszalności treści i formy utworu, czyli prawa do jego integralności. Z tego prawa wynika zakaz wprowadzania w utworze jakichkolwiek zmian bez zgody twórcy. *Inaczej niż w przypadku ochrony majątkowych praw autorskich ochrona prawa do integralności dzieła nie jest, naszym zdaniem, ograniczona do elementów twórczych*¹²⁸. Prawo do rzetelnego wykorzystania, w przeciwieństwie do prawa do integralności, nie jest ściśle związane z udostępnianiem utworu publiczności lub umieszczaniem go w odpowiednim kontekście. Przepis został sformułowany jednoznacznie i nie zawiera żadnych warunków: co do zasady forma i treść utworu nie może zostać zmieniona bez zezwolenia twórcy. A. Niżankowska zauważa, że w ustawodawstwach wielu państw udostępnienie szerszej publiczności jest warunkiem koniecznym naruszenia integralności, a w wielu innych taki warunek jest postulowany przez doktrynę. Sytuacja pod tym względem jest prosta w państwach, w których warunkiem bezprawnego naruszenia integralności jest przynajmniej potencjalna szkodliwość dla honoru lub reputacji twórcy¹²⁹. Zasada ta wydaje się autorce „prosta i logiczna”, z czym należy się zgodzić: *Wydawałoby się, że skoro podstawowym ideologicznym uzasadnieniem udzielenia ochrony integralności dzieła jest umożliwienie jego odbiorcom zapoznania się z nim w postaci nadanej mu przez autora, to tym samym w sytuacji gdy ingerencja w postać dzieła nie jest komunikowana „na zewnątrz” (pozostaje w sferze prywatnej), nie dochodzi do naruszenia jego integralności*¹³⁰.

Polska regulacja nie wiąże bezprawnego naruszenia integralności ze szkodą dla reputacji twórcy. *Ustawodawca niejako „z góry” zakłada, że każda ingerencja w treść lub formę utworu, lub jego nierzetelne wykorzystanie prowadzi do naruszenia prawa do integralności dzieła (...). Dotyczy to nawet ingerencji w dzieło znajdujące się w sferze prywatnej*¹³¹. A. Niżankowska ostatecznie ten wniosek bagatelizuje i uznaje, że w praktyce nie ma istotnego znaczenia, ponieważ autor tylko wyjątkowo może dowiedzieć się o fakcie ingerencji w integralność dzieła, jeśli nie opuści ono sfery prywatnej.

W kontekście analiz metodami TDM należy jednak rozróżnić dwie sytuacje. Po pierwsze, gdy ingerencja w integralność jest tymczasowym etapem, a udostępniane są efekty analizy, które zawierają same tylko fakty niepodlegające ochronie prawnoautorskiej. Po drugie, gdy efekty analizy zachowują elementy twórcze analizowanych utworów. Nie można zaakceptować, by jakkolwiek etap analizy TDM prowadzony był *contra legem*, zwłaszcza, jeżeli podmiotem dokonującym analizy jest instytucja naukowa, nawet gdyby ryzyko „złapania” przez uprawnionego było niewielkie. Zresztą w przypadku badań naukowych ryzyko wcale nie jest niewielkie.

¹²⁸ Tamże, s. 54.

¹²⁹ A. Niżankowska, *Prawo do integralności...*, s. 182.

¹³⁰ Tamże, s. 182.

¹³¹ Tamże, s. 183.

Metodologia wymaga przecież wyczerpującego opisanie dokonywanych czynności.

Rzetelne wykorzystanie a integralność utworu

Wyjście poza koleiny wytyczone przez dotychczasową doktrynę pozwoli zrekonstruować z obecnie obowiązujących przepisów normę wymagającą udostępnienia zniekształconego utworu szerszej publiczności. Jak wcześniej wspomniałem, art. 16 pkt 3 wymienia nienaruszalność treści i formy utworu razem z rzetelnym wykorzystaniem tego utworu. Skoro pozostałe dobra osobiste wymienione w pkt. 1-2 oraz 4-5 są wymienione pojedynczo, to jest logiczne, że w pkt. 3 również wymienione zostało tylko jedno dobro osobiste. Nie należy zakładać, że przyjęta redakcja przepisu jest przypadkowa. Sens interpretacji rekonstruującej dobro osobiste łączące naruszenie integralności i rzetelne wykorzystanie leży w rozróżnieniu prawnej i bezprawnej ingerencji w integralność. Naruszenie integralności, aby stało się bezprawne, musi zostać przedstawione publiczności. To nie jest pierwsza interpretacja, jaka przychodzi na myśl przy lekturze tego przepisu. Redakcja całego art. 16 jest jednak mocno krytykowana za wewnętrzną niespójność.

SN w wyroku z dn. 6 grudnia 2013 r. o sygn. I CSK 109/13 wypowiedział się m.in. w kwestii zakresu prawa do rzetelnego wykorzystania utworu. Opowiedział się za szeroką wykładnią, która powinna *dotyczyć wszystkich aspektów korzystania z utworu, w tym jego eksploatacji dokonywanej bez związku z innymi utworami, oceny miejsca i sposobu prezentacji, czynienia użytku z utworu innego niż tworzenie dzieł zależnych*. Wobec braku legalnej definicji rzetelnego wykorzystania utworu, uznał za *wskazane sięgnięcie do kryteriów określonych w art. 49 ust. 1 p.a.p.p. [PrAut – MS] Przepis ten ma charakter klauzuli generalnej, zastępującej dyspozycje stron umowy i zezwalającej na odtworzenie zakresu uprawnień korzystającego z utworu*. Rzetelne korzystanie z utworu to zatem takie korzystanie, które jest zgodne z charakterem i przeznaczeniem utworu, a także przyjętymi zwyczajami.

Przesłanki z art. 49 ust. 1 są nieprecyzyjne i umożliwiają różną interpretację. *Kryteria charakteru utworu, jego przeznaczenia oraz zwyczajów trudno potraktować jako całkowicie rozłączne. Charakter utworu często bowiem współdecyduje o jego przeznaczeniu, przeznaczenie zaś wynika z określonej praktyki postępowania (zwyczaju). Ocena tych kryteriów powinna być więc oceną całościową. Innymi słowy, jakkolwiek z brzmienia art. 49 ust. 1 wynika wymóg koniunkcji (tj. korzystanie ma być zgodne z charakterem utworu oraz jego przeznaczeniem oraz panującymi zwyczajami), to w praktyce dokonywać należy oceny globalnej. Każde z użytych przez ustawę pojęć jest wysoce podatne na interpretację i niemożliwe do precyzyjnego zdefiniowania, co dodatkowo utrudniałoby stosowanie każdego z kryteriów całkowicie odrębnie od pozostałych¹³²*. Postulat stosowania „oceny globalnej” jest bardzo niejasny. Chyba nie należy rozumieć go jako

¹³² T. Targosz, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 709-710.

dopuszczenia, wbrew literze przepisu, rezygnacji w pewnych sytuacjach z wymogu koniunkcji trzech warunków. Stosowanie kryteriów „globalnie”, czyli nie odrębnie od pozostałych, wydaje się prowadzić przeciw do koniunkcji.

Metody TDM nie spełniają tej przesłanki, jeśli przeznaczenie utworu rozumie się jako intencję, z jaką utwór on stworzony. Utwory dotychczas raczej nie powstają z intencją stania się przedmiotem analiz metodami TDM, choć wraz z upowszechnianiem się tych metod świadomość, że może dojść do takiego wykorzystania będzie również się upowszechniać. Nie spodziewam się, by kiedykolwiek TDM był podstawowym przeznaczeniem utworu. Charakter utworu można rozumieć jako dostosowanie utworu do swojego przeznaczenia, czyli w szczególności obejmowałby raczej aspekty techniczne lub formalne (np. w przypadku książki: czy została wydana drukiem, czy w wersji elektronicznej, a jeśli w wersji elektronicznej, to w jakim formacie). Charakter utworu oznacza jednak także takie aspekty jak przynależność do określonego gatunku, np. parodii lub pracy naukowej. Gdyby wymóg zgodności z przyjętymi zwyczajami rozumieć wąsko, to uniemożliwiłoby on zastosowanie tego przepisu do wszystkich nowych sposobów korzystania z utworów, w tym z TDM z wyjątkiem obszarów, w których już przyjęło się stosować te metody.

Należy zwrócić uwagę na wątpliwość dotyczącą zakresu art. 49. Wątek ten rozwijam w następnym podrozdziale.

Przesłanki dopuszczające ingerencję w integralność

Prawo do integralności nie jest prawem bezwzględny. Art. 49 ust. 2 stanowi, że następca prawny nie może, bez zgody twórcy, czynić zmian w utworze, chyba że są one spowodowane oczywistą koniecznością, a twórca nie miałby słusznej podstawy im się sprzeciwić. Oznacza to, że po spełnieniu dwóch warunków (oczywistej konieczności i braku podstaw do sprzeciwu) następca prawny może dokonywać zmian. *Założeniem przepisu [art. 49.2] jest przekonanie, że w pewnych przypadkach podmiot uprawniony z tytułu autorskich praw majątkowych musi mieć możliwość dokonania zmian w dziele, nawet bez zgody autora lub wręcz wbrew jego woli. Wykładnia art. 49 ust. 2 powinna dążyć do znalezienia właściwego punktu równowagi między prawem do integralności a potrzebami obrotu oraz interesami osób korzystających z utworów*¹³³. Literalna wykładnia przepisu uprawniałaby do takich ingerencji wyłącznie nabywców autorskich praw majątkowych lub spadkobierców. W doktrynie przyjęła się interpretacja rozszerzająca ten przepis także o licencjobiorców, a nawet podmioty korzystające z utworów na zasadach dozwolonego użytku i właścicieli egzemplarzy¹³⁴. Jestem przekonany o celowości rozszerzenia zakresu podmiotowego tego przepisu. Taka interpretacja wydaje się jednak nazbyt swobodna w obecnym

¹³³ Tamże, s. 712.

¹³⁴ Tamże, a także A. Niżankowska, *Prawo do integralności...*, s. 169 i n.

stanie prawnym. Rozszerzenie zakresu podmiotowego powinno być formułowane jako postulat *de lege ferenda*. Na przeszkodzie stoi nie tylko dosłowne znaczenie zwrotu „następca prawny” (które jest w ogólnym prawie cywilnym raczej jednoznaczne). Przynajmniej na wykładnię nazbyt rozszerzającą sens literalny nie pozwala umiejscowienie tego przepisu w Rozdziale 5 „Przejście autorskich praw majątkowych”, a także ust. 1, który dotyczy sytuacji umownych. Przeciwno wykładni rozszerzającej przemawia także zasada zabraniająca stosowania takiej wykładni do przepisów ograniczających prawa bezwzględne autora.

Treść Rozdziału 5 uzasadnia rozszerzenie zakresu podmiotowego art. 49 ust. 2 na licencjobiorców. Podobną opinię wyrazili J. Barta i R. Markiewicz¹³⁵. Choć zgadzam się z A. Niżankowską, że wąskie rozumienie „następcy prawnego” może prowadzić do sytuacji nieracjonalnych z punktu całokształtu systemu prawa autorskiego¹³⁶, to nie widzę jednak możliwości w obecnym stanie prawnym rozszerzenia m.in. na podmioty korzystające z utworów na zasadach dozwolonego użytku.

Nawet gdy nie ma możliwości powołania się na art. 49 ust. 2 w celu wprowadzania koniecznych zmian do utworu, jest inny sposób uzasadnienia takich koniecznych działań. Rację ma bowiem Niżankowska, gdy stwierdza: *Nie ulega zaś wątpliwości, że cel wykorzystania utworów w ramach dozwolonego użytku niejednokrotnie wymaga wprowadzenia pewnych zmian, jak chociażby skrótów*¹³⁷. Skoro uprawnieni mogą korzystać z utworów w ramach dozwolonego użytku, to mogą podejmować wszelkie działania, które są oczywiście konieczne, a zarazem takie, którym autor nie ma podstaw sprzeciwić się. Próba zabraniań takich działań przez autora kwalifikowałaby się jako nadużycie prawa i zgodnie z art. 5 KC nie mogła być uznana za wykonywanie prawa, ani nie korzystałaby z ochrony.

Jeśli możliwość wprowadzania koniecznych zmian gwarantuje art. 5 KC, to czy nie oznacza to, że art. 49 ust 2 stanowi *superfluum*? Literalny sens przepisu nie prowadzi do takiego wniosku. Literalny sens potwierdza i wzmacnia rozdzielnosc autorskich praw osobistych i majątkowych: choćby ktoś nabył pełnię praw majątkowych do utworu, lub prawa majątkowe do utworu już wygasły, to nie wpływa to na ochronę integralności. Ewentualne *superfluum* dotyczyłoby powtórzenia treści nie art. 5 KC, ale różnicy wyrażonej w art. 16 i 17 PrAut.

Czy podmiot prowadzący badania naukowe metodami TDM spełnia przesłanki pozwalające ingerować w treść lub formę utworów, obojętnie czy na podstawie art. 49 ust. 2, czy art. 5 KC? Przypomnijmy, że takie ingerencje mają miejsce w procesie czyszczenia utworu z szumu, deduplikacji, ujednoznaczniania oraz ujednoczania np. form gramatycznych. Należy rozpatrzyć obie przesłanki. Niewątpliwie takie działania są oczywiście konieczne, by analizy metodami TDM

¹³⁵ Barta, Markiewicz, *Prawo autorskie...*, s. 54.

¹³⁶ A. Niżankowska, *Prawo do integralności...*, s. 169.

¹³⁷ Tamże, s. 171.

mogły zostać przeprowadzone zgodnie z rygorami metodologicznej poprawności. A czy twórca ma słuszne podstawy sprzeciwienia się takim praktykom?

Obie przesłanki mają charakter obiektywny¹³⁸. Za A. Kopffem podnosi się, że określenie „słuszna podstawa” obejmuje zarówno sferę interesów osobistych, jak i majątkowych twórcy. Nie da się określić „słusznego interesu twórcy” *in abstracto*. Domaganie się przez twórcę zaprzestania ingerowania w treść lub formę utworów w celu analizy ich metodami TDM oznaczałoby w praktyce zakaz prowadzenia takich analiz w ogóle. Zderzają się tu sprzeczne interesów autora i badacza. Takie interesy należy wyważyć.

Na potrzebę ważenia interesów w przypadku stosowania uprawnienia do koniecznych modyfikacji zwraca uwagę A. Niżankowska. Zachęca do sięgnięcia po zagraniczne doświadczenia związane ze stosowaniem analogicznych przepisów. *Wypracowane i sprawdzone przez lata w doktrynie i orzecznictwie niemieckim kryteria są niezwykle wyczerpujące, co pozwala na możliwie najbardziej słuszne i sprawiedliwe porównanie interesów obu stron we wprowadzaniu zmian do utworów i ocenę, czy wprowadzenie zmian było spowodowane oczywistą koniecznością (interes drugiej strony) i czy twórca miałby słuszną podstawę im się sprzeciwić (interes autora)*¹³⁹.

Wątek ważenia interesów zostanie rozwinięty w rozdziale o wolności badań naukowych.

¹³⁸ Tamże, s. 173, por. też: J. Błeszyński, *Prawo autorskie...*, s. 122.

¹³⁹ A. Niżankowska, *Prawo do integralności...*, s. 174.

Ochrona prawnoautorska efektów TDM

Metody TDM są wyrazem coraz bardziej subtelnej wykorzystywania techniki w obszarach życia społecznego, które dotychczas wiązano z twórczością i intelektualną działalnością człowieka. Efekty pracy komputerów będą coraz bardziej ludzako przypominać efekty twórczej pracy człowieka. W wielu sytuacjach wysiłek człowieka wzmacniany jest narzędziami komputerowymi. *Należy liczyć się z tym, że coraz powszechniejsze wykorzystanie programów komputerowych będzie sprzyjać większej liczbie przypadków łączenia wysiłku ludzkiego ze sprawnością maszyn, co może prowadzić do istotnych trudności w identyfikacji rzeczywistego twórcy utworu*¹⁴⁰.

Dotychczas niedocenianym zagadnieniem są skutki prawne stosowania metod nadzorowanych (*supervised*) oraz częściowo nadzorowanych (*semi-supervised*) TDM. Przypomnę, że są to metody, w których na pewnym etapie prac istotną rolę odgrywa ręczna praca człowieka. Może ona polegać na ręcznym otagowaniu lub anotowaniu, albo na weryfikacji ostatecznego efektu pracy skryptów. Często polega na opracowaniu niewielkiej części analizowanego materiału, tak by algorytmy na podstawie tego wzorca skutecznie opracowały cały badany materiał.

Istotną, a niedocenianą, jak się wydaje, kwestią, jest charakter obiektów powstałych w efekcie analiz metodami TDM. Czy można mówić, by przynajmniej w niektórych sytuacjach stosowania TDM powstawał nowy utwór? Od odpowiedzi na to pytanie zależy, w jaki sposób będzie ten obiekt chroniony, czy i jak będzie mógł być udostępniany, na jakich zasadach będą mogły z niego korzystać inne osoby? Wreszcie, będzie miało to wpływ na możliwość zastosowania prawa cytatu.

Utwór

Art. 1 PrAut określa przedmiot ochrony prawnoautorskiej. W rozumieniu PrAut utwór jest dobrem niematerialnym, różnym od przedmiotu materialnego (*corpus mechanicum*), na którym jest często utrwalany. Musi spełniać łącznie następujące warunki: powinien stanowić rezultat pracy człowieka, być przejawem działalności twórczej, cechować się indywidualnym charakterem oraz zostać w jakikolwiek sposób ustalony¹⁴¹. Warunek uzależniający przyznanie ochrony prawnoautorskiej od faktu, że obiekt jest rezultatem pracy człowieka nie został wprost wyrażony.

140

D. Flisak, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 25.

¹⁴¹ J. Barta, R. Markiewicz, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. J. Barta, R. Markiewicz, wyd. 5, Warszawa 2011, s. 19.

Wynika jednak z charakteru twórczości oraz z art. 8, który określa podmiot praw autorskich jako osobę mającą imię i nazwisko. Ochronie prawnoautorskiej podlegają twórcze elementy dzieła oraz utwór jako całość. Nie podlegają natomiast ochronie prawnoautorskiej te elementy, które nie spełniają przesłanek utworu.

*Odwołanie się do rezultatów pracy człowieka stawia poza zakresem autorskoprawnej ochrony efekty działania zwierząt czy przyrody. Podobnie produkty całkowicie zautomatyzowanych maszyn*¹⁴².

Twierdzenie, że utwór stanowi przejaw „działalności twórczej”, oznacza zatem, że utwór powinien stanowić rezultat działalności o charakterze kreacyjnym. Przesłanka ta, niekiedy określana jako przesłanka „oryginalność” utworu, zrealizowana jest wówczas, gdy istnieje subiektywnie nowy wytwór intelektu; jest ona zatem ujmowana wyłącznie w płaszczyźnie nowości subiektywnej i zorientowana retrospektywnie¹⁴³ - syntetycznie podsumował Sąd Apelacyjny w Poznaniu. Wyjaśnianie znaczenia „działalności twórczej” jako „działalności o charakterze kreacyjnym” jest definiowaniem *ignotum per ignotum*, a nawet posługiwaniem się synonimem. W każdym razie, o twórczości można mówić w przypadku subiektywnie nowego wytworu intelektu. Nie może być to efekt czynności rutynowych lub zdeterminowanych. Twórca musi dysponować pewną swobodą¹⁴⁴.

Przyjmuje się, że twórczość oznacza „subiektywną nowość”. Przyjęcie za punkt wyjścia tezy, że Prawo autorskie nie wymaga nowości w znaczeniu przedmiotowym, lecz w ujęciu subiektywnym (oryginalność pokrywa się z nowością podmiotową), nie oznacza, iż deklarowane subiektywne przekonanie twórcy, że dane dzieło stanowi nowy wytwór intelektu, nie poddaje się jakiegokolwiek weryfikacji¹⁴⁵. SN w cytowanym orzeczeniu dostrzegł ryzyko zbliżania się w praktyce do domniemania faktycznego, że każdy wytwór pracy umysłowej człowieka jest dziełem, któremu przysługuje ochrona prawnoautorska. W konsekwencji grozi to deprecjacją pojęcia twórczości w ogóle.

W. Machała przeciwstawia dominującej koncepcji minimalistycznej koncepcję rygorystyczną twórczości. Podnosi, że stosowanie jako kryterium „nowości subiektywnej” prowadzi do absurdu¹⁴⁶. Uzasadnia swoją propozycję w argumentie, że czym innym jest ochrona utworu niezależnie od jego wartości (co jest wprost wyrażone w ustawie), a czym innym stosowanie kryteriów ocennych na etapie uznawania obiektu za utwór. To ostatnie nie jest wartościowaniem utworu. Skoro

¹⁴² J. Barta, R. Markiewicz, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. J. Barta, R. Markiewicz, wyd. 5, Warszawa 2011, s. 21.

¹⁴³ Wyrok SA w Poznaniu z dn. 7 listopada 2007 r., sygn. I Aca 800/07.

¹⁴⁴ J. Barta, R. Markiewicz, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. J. Barta, R. Markiewicz, wyd. 5, Warszawa 2011, s. 21.

¹⁴⁵ Wyrok SN z dn. 22 czerwca 2010 r., sygn. IV CSK 359/09.

¹⁴⁶ W. Machała, *Utwór. Przedmiot prawa autorskiego*, Warszawa 2013, s. 136.

ustawodawca nie podaje żadnych wskazówek, jak interpretować przesłankę twórczości, należy stosować reguły semantyczne języka polskiego, a także do teorii psychologicznych i estetycznych. Te zaś wiążą twórczość z nowymi wartościami i doniosłymi skutkami¹⁴⁷. Powołuje się na W. Tatarkiewicza, który łączył twórczość z doniosłością kulturową (na marginesie należy wyrazić ubolewanie, że współczesna doktryna prawnicza tylko wyjątkowo powołuje się na dorobek teorii estetyki nowszy niż prace W. Tatarkiewicza czy R. Ingardena).

Z kolei indywidualny charakter dzieła wyraża się w tym, że odbiega ono od innych przejawów podobnego działania w sposób świadczący o jego swoistości i niepowtarzalności. Ta cecha - niepowtarzalność - bywa rozumiana w sposób rygorystyczny jako rezultat statystycznie jednorazowy¹⁴⁸. Rezultat pracy nie może być całkowicie zdeterminowany przez otoczenie. Muszą występować w nim elementy, których kształt zależał od osobistego ujęcia autora. O indywidualności świadczy statystyczna niepowtarzalność, zgodnie z teorią zaproponowaną przez M. Kummera. Teoria ta pozostawia jednak wątpliwości.

TDM

Nie ma wątpliwości, że metody nienadzorowane maszynowej analizy tekstu i danych nie spełniają kryterium działalności człowieka. Związek pomiędzy napisaniem skryptu a zastosowaniem go do konkretnego materiału ma charakter nazbyt pośredni. Skrypt podlega ochronie jako program komputerowy. Samo zastosowanie go do analizy konkretnego materiału nie jest działalnością ani twórczą, ani indywidualną.

Sytuacja przestaje być oczywista w momencie stosowania metod nadzorowanych. Nie można przedstawić jednego, uniwersalnego stanowiska, ponieważ te metody są zbyt zróżnicowane. Każdą metodę TDM należałoby przeanalizować indywidualnie pod tym kątem. W tym miejscu można wskazać tylko pewne ogólne spostrzeżenia dotyczące spełniania poszczególnych przesłanek objęcia ochroną prawnoautorską.

Działalność człowieka

W przypadku metod nienadzorowanych (*unsupervised*) nie można mówić o tym, że efekt jest przejawem działalności człowieka. Jest to efekt procesu w pełni zautomatyzowanego. Tym samym produkty wytworzone tymi metodami nie mogą uzyskać ochrony prawnoautorskiej, choćby wyglądały podobnie do utworów, np. automatyczne tłumaczenia. *Powszechne wykorzystanie komputerów w procesie twórczej kreacji (...) może niekiedy utrudniać rozstrzygnięcie, czy ostateczny efekt takiego tworzenia należy przypisać ludzkiemu intelektowi, czy też technicznej sprawności maszyny. Z jednej strony jest oczywiste, że posłużenie się oprogramowaniem (...) nie*

¹⁴⁷ Tamże, s. 114.

¹⁴⁸ Wyrok SN z dn. 30 czerwca 2005 r., sygn. IV CK 763/04.

może stać na przeszkodzie zakwalifikowaniu jako utwór (...). Z drugiej jednak strony faktyczne zastąpienie człowieka przez maszynę, np. przez wykorzystanie translatorskiego oprogramowania do przetłumaczenia tekstu, nie będzie prowadziło do powstania utworu, gdyż końcowy efekt w postaci tłumaczenia nie będzie pochodną wysiłku ludzkiego intelektu, lecz zautomatyzowanej logiki (...)¹⁴⁹.

Im większe znaczenie mają metody nadzorowane, w tym większym stopniu efekt może być przejawem działalności człowieka. Nie można jednak uznać, że w każdym przypadku stosowania metod nadzorowanych przesłanka ta zostaje spełniona. Przede wszystkim należy rozważyć, czy przesłankę tę spełnia opracowanie, zwłaszcza szczegółowe, niewielkiej części analizowanego zbioru danych, który służy za wzór dla algorytmów przeprowadzających analogiczne operacje dla całości korpusu (*training material*). Na pewno opracowanie wzorca jest przejawem działalności człowieka. W moim przekonaniu można bronić tezy, że efekty analizy całego korpusu są w takim przypadku również przejawem działalności człowieka. Skrypt jest tu narzędziem, które należy dostosować do konkretnego konkretnego materiału. Bez ingerencji człowieka albo w ogóle nie poradziłby sobie z zadaniem, albo skuteczność byłaby na niezadowalająco niskim poziomie.

Należy rozważyć także sytuację, gdy ingerencja badacza polega na kontroli efektu działania programu komputerowego. Algorytmy kategoryzują materiał, który dopiero podlega weryfikacji. Czy następcze oddziaływanie na ostateczny kształt produktu jest wystarczające, by spełnić przesłankę działalności człowieka (przy czym w tym miejscu nie zastanawiam się jeszcze nad przesłankami twórczości i indywidualności)?

Można zastosować dwa alternatywne kryteria oceny. Pierwsze dotyczy ilości pracy włożonej przez badacza. To kryterium wiązałoby się z wysiłkiem włożonym w sprawdzenie efektów. Drugie kryterium dotyczy stopnia ingerencji. Niezależnie od wysiłku, jeśli w trakcie weryfikacji badacz wprowadza znaczną liczbę poprawek, a produkt dzięki temu znacząco się zmienia, całość można uznać za przejaw działalności człowieka. Zastosowanie jednego lub drugiego kryterium prowadzi do wykluczających się wniosków, gdy w wyniku kontroli okupionej znacznym wysiłkiem potwierdzona zostaje skuteczność programów, a zmiany wprowadzane bezpośrednio przez człowieka pozostają znikome. Według pierwszego kryterium spełniona zostałaby przesłanka działalności człowieka, a według drugiego kryterium – nie.

Twórczość (subiektywna nowość)

Jak wyżej zaznaczyłem, twórczość jest rozumiana na gruncie PrAut jako subiektywna nowość. Twórca powinien być przekonany, że dzieło stanowi wytwór jego wyobraźni¹⁵⁰. Praca nad dziełem musi różnić się od czynności rutynowych, co określa się jako niepewność rezultatu (takie określenie budzi jednak wątpliwości i nie jest przyjmowane jako oczywiste). Jednocześnie

¹⁴⁹ D. Flisak, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 24.

¹⁵⁰ D. Flisak, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 23-24.

subiektywna nowość podlega zobiektywizowanej ocenie. Kryterium jest zatem odróżnianie się dzieła od dotychczasowych wytworów także w oczach postronnych obserwatorów.

Niektóre produkty analiz metodami TDM w sposób oczywisty nie spełniają przesłanki twórczości, np. baza faktów wydobytych z artykułów biomedycznych i ułożonych według obiektywnego kryterium. W przypadku jednak takich obiektów jak np. automatyczne tłumaczenia (zwłaszcza przy stale rosnącej ich jakości) lub podsumowania, odpowiedź na pytanie, czy spełniają one cechę subiektywnej nowości, okazuje się już nie tak oczywista, ponieważ zewnętrzny obserwator może być przekonany, że ma do czynienia z obiektem, który jest nowy.

Jeśli ograniczać pojęcie twórczości do subiektywnej nowości, to należałoby uznać, że wskutek stosowania metod TDM mogą powstać obiekty spełniające tę przesłankę. Na przeszkodzie temu wnioskowi może stać warunek nieprzewidywalności efektu, czy brak działania rutynowego. Na nieprzewidywalność wskazują autorzy komentarza: *Wskazuje się, że wyróżnikiem pracy twórczej jest brak możliwości przewidzenia końcowego rezultatu podjętej aktywności*¹⁵¹. Powołują się przy tym na wyrok NSA, w którego uzasadnieniu padło następujące stwierdzenie: *praca twórcza cechuje się tym, że jej rezultat jest niepewny*¹⁵². Zdanie to, wyrwane z kontekstu, nabiera trochę innego znaczenia, niż miało w uzasadnieniu. W sprawie, którą rostrzygał NSA, zagadnienie dotyczyło kwestii podatkowych. Niepewność dotyczyła tego, czy powstało jakiegokolwiek dzieło w ramach stosunku pracy, a w związku z tym, czy 60% wynagrodzenia można było zakwalifikować jako honorarium autorskie, co pozwoliłoby powodowi odliczyć od podatku koszty uzyskania. Niepewność dotyczyła więc osiągnięcia jakiegokolwiek efektu w ogóle. Niepewność, o której piszą autorzy Komentarza, wydaje się dotyczyć raczej tego, jaki będzie utwór, niż tego, czy w ogóle powstanie. Oba rodzaje niepewności są ze sobą związane, ale nie są tożsame.

Niepewność co do efektu finalnego może być cechą procesu twórczego, ale nie musi. Niektórzy twórcy przystępując do ustalania dzieła w postaci nadającej się do zaprezentowania mają już precyzyjny obraz tego, jaki ma być efekt finalny. Osiągnięcie efektu finalnego zgodnego z założoną wizją jest przejawem dobrego rzemiosła i sprawnego posługiwania się narzędziami. Polega na wykonywaniu działań rutynowych i wyćwiczonych. *Nieprzewidywalność końcowego rezultatu nie może stanowić bezwzględnego warunku tworzenia, trudno byłoby bowiem stwierdzić, co tak naprawdę oznacza „nieprzewidywalność” tworzenia*¹⁵³.

Podobnie należy zachować pewną ostrożność przy formułowaniu zakazu działania rutynowego. Jeśli działanie rutynowe to „wykonywany często i niemal automatycznie”¹⁵⁴, to należy uznać że działania rutynowe są częścią wielu procesów twórczych, z tym że nakierowane są one na

¹⁵¹ D. Flisak, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 24.

¹⁵² Wyrok NSA z dn. 12 marca 2010 r., sygn. II FSK 1791/08.

¹⁵³ D. Flisak, w: *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Warszawa 2015, s. 24.

¹⁵⁴ Słownik Języka Polskiego PWN, <http://sjp.pwn.pl/sjp/rutynowy;2574573.html>.

twórczy cel i temu celowi podporządkowane. Rutyna leży u podstaw wszelkiej pracy wymagającej wyspecjalizowanych umiejętności, w tym pracy twórczej.

Wydaje się zatem, paradoksalnie, że w niektóre produkty TDM spełniają przesłankę subiektywnej nowości.

Indywidualność (oryginalność)

W przypadku metod nienadzorowanych raczej nie można mówić o spełnianiu cechy indywidualności. W tych przypadkach oryginalność może cechować program komputerowy (który może być przygotowany specjalnie z myślą o danym badaniu), ale nie efekt jego pracy, który jest powtarzalny.

W przypadku stosowania metod nadzorowanych, należy ocenić rodzaj ingerencji badacza. Jeżeli badacz np. anotuje, czyli dodaje informacje gramatyczne do tekstu, to poprawnie wykonane zadanie będzie wyglądało tak samo, niezależnie od wykonującej tę pracę osoby. Nie można jednak wykluczyć, że pewne zadania wykonywane przez człowieka mogą mieć wyjątkowo charakter indywidualny, a wykonane przez inną osobę będzie wyglądało inaczej.

Zasada wolności badań naukowych

Zakres uprawnień wynikających z prawa autorskiego często zależy od ważenia interesów lub skonkretyzowania znaczenia pojęć nieostrych lub nie zdefiniowanych ustawowo. Ma to miejsce m.in. w przypadku stosowania różnych instytucji dozwolonego użytku, np. własny użytek osobisty, zwielokrotnianie o charakterze incydentalnym lub przejściowym, uzasadnione cele cytatu. Art. 34 PrAut ogranicza wymóg podania twórcy i źródła poprzez uwzględnienie istniejących możliwości.

Szczególnie newralgiczny jest art. 35 PrAut. Decyduje on o zakresie każdego dozwolonego użytku. Zgodnie z nim dozwolony użytek nie może naruszać normalnego korzystania z utworu lub godzić w słusne interesy twórcy. Ustawa nie precyzuje jednak, czym jest normalne korzystanie z utworu, ani nie podaje wprost kryteriów oceny słusnych interesów twórcy. Można – co zostało przedstawione we wcześniejszych częściach pracy – interpretować te warunki jako ochronę eksploatacji utworu przynoszącą korzyści finansowe oraz jako ochronę autorskich dóbr osobistych.

TDM, niekoniecznie w celach naukowych, to nowe sposoby korzystania z utworów. Nie ma jeszcze jasnej odpowiedzi na pytanie, czy na gruncie polskiego prawa te metody mieszczą się w kategorii normalnego korzystania, czy nie? Czy słusznym interesem twórcy jest możliwość zablokowania korzystania z jego utworu w celu analizowania go maszynowego? A jeśli tak, to w jakim zakresie? Odpowiedź nie musi być przecież albo-albo. Nie ma prostych analogii, które pomogłyby rozstrzygnąć te dylematy.

Pomocne w odpowiedzi będzie przyjrzenie się zasadzie wolności badań naukowych, która została wyrażona w Konstytucji RP z 1997 r., jej wyrazem są też przepisy ograniczające wyłączność praw autorskich w celach naukowych i wreszcie powinna być ona wykorzystywana jako jedno z kryteriów konkretyzacji nieostrych lub niejasnych sformułowań ustawowych.

Konstytucja RP a prawo autorskie i jego ograniczenia

Konstytucja RP z 1997 r. nie zawiera żadnego przepisu, który wprost mówi o ochronie praw autorskich, choć pojawiła się taka propozycja w 1994 r. na posiedzeniu Podkomisji Praw i Obowiązków Obywateli¹⁵⁵. W takiej sytuacji do praw własności intelektualnej należy odnosić konstytucyjne uregulowania własności (art. 21 i art. 64 Konst.). Wątpliwości budziło rozumienie terminu „własność” w obu przepisach i ich stosunek do cywilistycznego pojęcia własności. Należy

155

uznać, że pojęcia konstytucyjne są autonomiczne i nie mogą być definiowane w aktach niższego rzędu¹⁵⁶. Art. 21 znajduje się w Rozdziale I i jest częścią zasad ustrojowych RP. Art. 64 znajduje się w Rozdziale 2, w związku z czym jest jednym z praw człowieka.

Ograniczenia wolności i praw dopuszczalne są tylko po spełnieniu pięciostopniowego testu, określonego w art. 31 ust. 3: prawa i wolności ogranicza ustawa, zasada proporcjonalności, uwzględnienie zasad państwa demokratycznego, ograniczenia muszą służyć określonym wartościom (m.in. wolnościom i prawom innych osób) oraz nie mogą naruszać istoty tych wolności lub praw. Ograniczenia wynikające z instytucji dozwolonego użytku spełniają te kryteria i są zgodne z art. 31 Konstytucji¹⁵⁷. Czwartym warunkiem są prawa innych osób, zagwarantowane w art. 73 Konstytucji. *Dokonując zestawienia „autorskich uprawnień własnościowych” wynikających z art. 64 oraz uprawnień składających się na treść art. 73, dostrzec można, iż dochodzi między nimi do kolizji. Realizacja jednego z nich prowadzi do uszczuplenia zakresu drugiego¹⁵⁸. Ponieważ żadnemu z praw nie zostało przyznane pierwszeństwo, ustawodawca powinien dążyć do zagwarantowania obu praw w możliwie największym stopniu. J. Marcinkowska w kontekście art. 73 pisze o dwóch z pięciu praw zagwarantowanych w tym przepisie: o wolności twórczości artystycznej oraz o wolności korzystania z dóbr kultury. Pomija wolność badań naukowych, która również wpływa na przepisy o dozwolonym użytku.*

Wolność badań naukowych

Wolność badań naukowych nie wzbudzała dotychczas większego zainteresowania wśród polskich prawników. Kiedy stawała się przedmiotem analizy, zwracano uwagę przede wszystkim na jej aspekt związany z doborem tematu badań i swobodą upowszechniania ich wyników. B. Banaszak *implicite* utożsamia wolność badań naukowych z wolnością prowadzenia badań naukowych. Ta druga jest co najwyżej częścią szerszej wolności badań naukowych. *Wolność badań naukowych oznacza skierowany do władz publicznych nakaz wstrzymywania się od wszelkiej, pozbawionej wyraźnej podstawy prawnej, ingerencji w dziedzinę badań naukowych¹⁵⁹. W dalszej części omówienia tego przepisu B. Banaszak koncentruje się na raczej drugorzędnej kwestii możliwości zatrudnienia naukowców na wielu etatach.*

Wolność badań naukowych uważa się nawet za odmianę wolności wyrażania poglądów. *Niewątpliwie wolność badań naukowych i nauczania są ściśle związane z wolnością wyrażania poglądów (art. 54 Konstytucji), ponieważ stanowią jej formę. Obie wolności wymagają wolności duchowej, manifestującej się wolnością wyrażania poglądów, uzewnętrzniania w nieskrępowany*

¹⁵⁶ Tamże, s. 150-152.

¹⁵⁷ Tamże, s. 175.

¹⁵⁸ Tamże, s. 191.

¹⁵⁹ B. Banaszak, *Konstytucja Rzeczypospolitej Polskiej. Komentarz*, wyd. 2, Warszawa 2012, s. 434-435.

*sposób swoich myśli. Obie należą do praw o charakterze komunikacyjnym, wolność nauki bowiem ze swej istoty oparta jest na dyskursie*¹⁶⁰. Inny charakter ma już jednak dostęp do informacji, który jest warunkiem prowadzenia badań naukowych¹⁶¹. Dotyczy on przecież nie artykułowania swoich poglądów, ale dostępu do materiałów, na podstawie których te poglądy mają dopiero zostać skryształizowane. W dalszej części artykułu autorzy podają następującą definicję: *wolność badań naukowych to wolność od nieuzasadnionej, a więc arbitralnej ingerencji w decyzje dotyczące rozpoczęcia i prowadzenia badań w danej dziedzinie (...), wyboru przedmiotu badań, wyboru miejsca prowadzenia badań naukowych, określenia metod badań naukowych, ogłaszania ich wyników*¹⁶².

Pomimo znaczenia badań naukowych dla życia społecznego i rozwoju gospodarczego, ustrojodawca nie zdecydował się poświęcić tej wolności odrębnemu artykule w Konstytucji, pomimo wyeksponowania takich wolności jak np. wolność prasy (art. 14 Konstytucji). *Nie sposób jednak nie zauważyć, iż do wolności tej ustrojodawca zdaje się przywiązywać znacznie mniejszą wagę niż do wolności prasy i innych środków społecznego przekazu, zawartych w art. 14, które znalazły się w rozdziale I Konstytucji, stając się tym samym jedną z zasad ustrojowych*¹⁶³.

Wolność badań naukowych z art. 73 jest prawem człowieka (przysługuje każdemu). Z jednej strony oznacza to szeroki zakres podmiotowy: przysługuje każdemu znajdującemu się na terytorium RP, niezależnie od obywatelstwa. Z drugiej strony, uważa się, że jest ograniczona raczej do osób fizycznych, choć stanowisko to w moim przekonaniu nie jest oczywiste. *Pojęcie „każdy” odnosi się niewątpliwie do osób fizycznych, powstaje jednak pytanie, czy wyłącznie. Zasadniczo do twórczego, indywidualnego działania zdolny jest wyłącznie człowiek (osoba fizyczna)*¹⁶⁴. Błędem jest utożsamianie prowadzenia badań naukowych z twórczością naukową w znaczeniu autorskoprawnym. PrAut chroni zaledwie niewielką część pracy badawczej. Utwory w znaczeniu PrAut mogą powstawać na różnych etapach procesu badawczego, począwszy od wstępnych notatek, poprzez dane badawcze¹⁶⁵ zbierane lub wytwarzane w trakcie prowadzenia badań, aż po ostateczne publikacje. Nie są jednak wyłącznym elementem. Przykładowo niektóre dane badawcze takie jak pomiary laboratoryjne nie mają charakteru twórczego. Zawężanie zakresu podmiotowego

¹⁶⁰ S. Jarosz-Żukowska, Ł. Żukowski, *Wolność badań naukowych i nauczania*, w: „Realizacja i ochrona konstytucyjnych wolności i praw jednostki w polskim porządku prawnym”, Wrocław 2014, oai:www.repozytorium.uni.wroc.pl:53684, s. 713.

¹⁶¹ Tamże.

¹⁶² Tamże, s. 723.

¹⁶³ J. Sobczak, *Wolność badań naukowych. Standardy europejskie i rzeczywistość polska*, „Roczniki Nauk Społecznych”, 1/2008, s. 19.

¹⁶⁴ S. Jarosz-Żukowska, Ł. Żukowski, *Wolność badań naukowych i nauczania...*, s. 718.

¹⁶⁵ Dane badawcze są definiowane bardzo szeroko. Przykładowa definicja to „zarejestrowane materiały o charakterze faktograficznym powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych” (*Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings*, https://www.whitehouse.gov/omb/fedreg_a110-finalnotice). Taka definicja obejmuje nie tylko pomiary laboratoryjne, ale także tak różnorodne materiały, jak notatki terenowe z badań antropologicznych lub ankiety z badań socjologicznych.

wolności badań naukowych tylko na podstawie utożsamienia badań naukowych z twórczością naukową jest nieporozumieniem.

Z błędnego przekonania o zakresie podmiotowym wolności badań naukowych jako ograniczonym tylko do indywidualnych badaczy, wynika raczej niewłaściwe budowanie relacji między tą wolnością a autonomią uczelni wyższych. Odpowiednikiem jej w odniesieniu do instytucji naukowych jest autonomia uczelni wyższych (art. 70 ust. 5 Konst.). *Oznacza to zatem, że na płaszczyźnie idei odpowiednikiem autonomii szkół wyższych jest wolność nauki będąca zarazem prawem jednostki (badacza), a nie szkoły wyższej*¹⁶⁶. Gdyby faktycznie ustrojodawca przyznał indywidualnym badaczom wolność badań naukowych, a uczelniom wyższym autonomię mającą być odpowiednikiem wolności badań naukowych, to instytuty badawcze, Polska Akademia Nauk i inne instytucje naukowe, które nie są uczelniami wyższymi zostałyby pozbawione konstytucyjnych gwarancji w zakresie prowadzenia działalności naukowej.

Wydaje się, że zawężająca interpretacja wolności badań naukowych nie jest jedyną dopuszczalną. Należy uzupełnić ją o interpretację rozszerzającą zakres podmiotowy tej wolności o instytucje naukowe. Za szeroką interpretacją przemawia fakt, że – zwłaszcza w dzisiejszych czasach, gdy nauka uprawiana jest zespołowo, a badania często wymagają kosztownego zaplecza i infrastruktury – wolność indywidualnych badaczy żeby nie pozostać pustym postulatem wymaga wsparcia udzielanego przez instytucje wyposażone w odpowiednie uprawnienia. Wolność badań naukowych instytucji naukowych jest gwarancją realizacji takiej wolności indywidualnych badaczy.

Wolność badań naukowych w prawie autorskim

Ograniczenia bezwzględnych praw autorskich w celach naukowych należy postrzegać w kontekście konstytucyjnej wolności badań naukowych. Użytek naukowy został wprost przewidziany w art. 23 PrAut (jedno z uzasadnień dozwolonego użytku osobistego) oraz w art. 27-29 PrAut (prowadzenie badań naukowych, umieszczanie utworów w wypisach i antologiach, prowadzenie bibliotek, cytowanie w celu analizy naukowej). Z pewnością przepisy te należy uznać za ukonkretnienie wolności konstytucyjnej w zakresie, w jakim ich beneficjentami są indywidualni naukowcy. Jeśli zgodzić się na szeroką interpretację wolności badań naukowych, czyli objąć nią także instytucje naukowe, to odpowiednio należy uznać wymienione przepisy PrAut za ukonkretnienie tej wolności także w odniesieniu do instytucji.

Trybunał Konstytucyjny uzasadnił potrzebę ograniczania uprawnień innych osób w celu prowadzenia badań naukowych, wiążąc wolność z odpowiedzialnością za wysoką jakość. *Jakkolwiek ustrojodawca szeroko zakreślił krąg podmiotów korzystających z tej wolności, stanowiąc, że korzysta z niej „każdy”, to nie ulega wątpliwości, że wolność twórczości artystycznej,*

¹⁶⁶ S. Jarosz-Żukowska, Ł. Żukowski, *Wolność badań naukowych i nauczania...*, s. 717.

wolność badań naukowych, wolność ogłaszania wyników badań naukowych oraz wolność nauczania są szczególnie istotne dla m.in. pracowników naukowych/nauczycieli akademickich. Korzystanie z tych wolności wiąże się ze szczególną odpowiedzialnością nauczycieli akademickich oraz szkół wyższych, w których oni pracują. Jest to odpowiedzialność za świadczenie dydaktyki oraz uzyskiwanie wyników badań naukowych na możliwie najwyższym poziomie¹⁶⁷. Skoro uzasadnieniem wolności badań naukowych jest m.in. umożliwienie prowadzenia badań naukowych na najwyższym możliwym poziomie, to wynika z tego, że wszystkie przepisy nieostre, albo wykorzystujące klauzule generalne, należy interpretować w sposób maksymalnie wspierający prowadzenie badań naukowych, przynajmniej dopóty, dopóki ta wolność nie zderzy się z prawem lub wolnością innych osób (albo nie zajdzie inna przesłanka z art. 31 ust. 3 Konstytucji).

Warto przypomnieć w tym miejscu wyrok sądu w Amsterdamie w sprawie *Dzienników Anny Frank* (por. rozdział *Kwestie dopuszczalności metod TDM w świetle prawa UE*). Sąd uznał, że naukowcy, którzy dysponują legalnym dostępem do tekstu mogą analizować go metodami TDM bez dodatkowej zgody uprawnionych. Przeciwnie stanowisko, zdaniem sądu, prowadziłoby do naruszenia wolności badań naukowych.

Przepisy zapewniające swobodę badań naukowych w PrAut są ukonkretnieniem konstytucyjnej wolności badań naukowych. Fakt ten sprawia, że ograniczenie bezwzględnych praw twórcy znajduje mocne uzasadnienie. Oczywiście, PrAut zostało zbudowane wokół podstawowej zasady ochrony bezwzględnych praw twórcy.

Odpowiedź na pytanie, czy i w jakim stopniu prowadzenie badań naukowych z wykorzystaniem metod TDM, mieści się w zakresie obecnie obowiązujących przepisów o dozwolonym użytku, w szczególności, czy są zgodne z warunkami z art. 35 PrAut, czyli z normalnym korzystaniem z utworu oraz ze słusznymi interesami twórcy, powinna uwzględniać fakt, że za interpretacją korzystną dla badaczy stoi wolność konstytucyjna.

¹⁶⁷ Wyrok TK z dn. 28 kwietnia 2009 r., sygn. K 27/07.

Granice legalnego stosowania TDM.

Podsumowanie

Dotychczasowe rozważania prowadzą do następujących konkluzji. Wzrost znaczenia metod TDM w badaniach naukowych jest przejawem nie tylko mody, ale rzeczywistej potrzeby i nowych możliwości. W wielu państwach, także na poziomie europejskim, jest świadomość niepewności wynikającej z obecnego stanu prawnego, czy metody analizy, które z technicznego punktu widzenia stają się standardem, są dopuszczalne z prawnego punktu widzenia i w jakich granicach? W państwach z systemem *fair use* orzecznictwo wypracowuje interpretacje zakreślające granice legalnego stosowania metod TDM. W państwach opierających dozwolony użytek na zamkniętej liście wyjątków mówi się o potrzebie wprowadzenia wyjątku, który wprost zezwoli na stosowanie metod TDM przynajmniej do niekomercyjnych celów naukowych. Należy jednak zauważyć, że metody TDM są bardzo zróżnicowane. Wydaje się, że zazwyczaj stosowane uproszczenie polegające na analizie uwarunkowań prawnych do stosowania metod TDM w ogólności jest zbyt daleko idącym uproszczeniem. W rzeczywistości metody TDM różnią się pod względem zaangażowania człowieka, sposobu przetwarzania analizowanych danych (ich kopiowanie, zniekształcanie, tworzenie utworów pochodnych, itd.), rodzaju spodziewanego efektu (baza danych zawierająca same fakty lub obiekt przypominający utwór pochodny, np. tłumaczenie lub anotacje), planowane sposoby upowszechnienia wyników badań.

Proces badań naukowych metodami TDM można, upraszczając, podzielić na trzy etapy: 1. uzyskanie dostępu do utworu, 2. przetworzenie go i analiza, 3. archiwizacja i udostępnienie efektów.

1. Uzyskiwanie dostępu

Badacze mogą analizować dane, do których dostęp uzyskują na różny sposób. Dane mogą znajdować się w domenie publicznej. Takie dane mogą być analizowane metodami TDM. Badacze mogą korzystać również z formy umownej zawieranej z uprawnionym z tytułu praw autorskich do utworu, w szczególności z formy licencji. Licencje mogą być indywidualnie zawierane lub mogą to być licencje skierowane do nieokreślonego kręgu odbiorców (najpopularniejsze licencje z tej kategorii to licencje Creative Commons). Jeśli umowa zawiera postanowienia dotyczące analizowania TDM, to sytuacja jest jasna (można zastanawiać się nad skutecznością ewentualnego umownego zakazu stosowania TDM). Jeśli umowa w żaden sposób nie odnosi się do warunków

potrzebnych do analiz TDM, to należy stosować odpowiednio przepisy o dozwolonym użytku.

Wreszcie, dostęp do danych jest możliwy na zasadach dozwolonego użytku. Prawa autorskie są co do zasady bezwzględne. Przepisy szczególne ograniczają jednak monopol autorski chroniąc interes społeczny polegający na możliwości korzystania z utworów bez uzyskiwania zgody uprawnionego. To właśnie ten sposób uzyskiwania dostępu i granice wyznaczone przez przepisy o dozwolonym użytku wzbudzają najwięcej kontrowersji dotyczących legalności stosowania TDM. Od przepisów ogólnych o dozwolonym użytku odróżnia się przepisy regulujące wykorzystanie baz danych spełniające cechy utworu.

Możliwe są zasadniczo dwie sytuacje: dostęp do danych uzyskuje osoba fizyczna, która działa na podstawie art. 23 PrAut (dozwolony użytek osobisty), albo jednostka naukowa w rozumieniu ustawy o zasadach finansowania nauki (art. 27).

2. Przetwarzanie i analiza

Ten etap obejmuje zasadniczą część pracy metodami TDM i składa się z szeregu zadań. Większość metod TDM wymaga, by we wstępnej fazie opracować materiał poddawany analizie. Opracowanie to polega na czyszczeniu go z „szumu” (np. słów wieloznacznych lub nie niosących ze sobą treści). W związku z tym pojawia się wątpliwość, czy nie stanowi to naruszenia integralności utworu i czy nie jest nierzetelnym wykorzystaniem? A jeśli jest to naruszenie integralności, to czy jest ono bezprawne? W przypadku dostępu opartego na umowie wyraźnie dozwalającej stosowanie metod TDM, odpowiedź jest prosta: takie działania są dozwolone. W pozostałych przypadkach (zarówno dozwolonego użytku, jak i domeny publicznej) odpowiedź nie jest oczywista. Pojawia się potrzebaważenia interesów autora, badacza i interesu publicznego.

Zależnie od podmiotu różny jest zakres przedmiotowy dozwolonego użytku. Raz jeszcze należy wspomnieć o odrębnościach ochrony baz danych spełniających cechy utworu. Podstawowe znaczenie dla badań naukowych metodami TDM ma art. 27 PrAut. Daje on jednostkom naukowym szerokie uprawnienia, w tym na wykorzystanie komercyjne (o ile zachowany jest cel naukowy, czyli działanie nakierowane na uzyskanie nowej wiedzy) i wykorzystanie tłumaczeń (w tym opracowanych automatycznie). Bardzo dużym utrudnieniem jest jednak dozwolenie na zwielokrotnianie tylko fragmentów większych utworów lub drobnych utworów w całości. Tymczasem często potrzebne jest zwielokrotnianie całości utworów niezależnie od ich rozmiarów.

Art. 23¹ (przejściowe lub incydentalne zwielokrotnianie utworów) znajdzie zastosowanie przy TDM jedynie w wyjątkowych sytuacjach. Przepis ten, choć pomyślany jako odpowiedź na wyzwania cyfrowego obiegu kultury, obwarowany jest zbyt restrykcyjnymi wymaganiami.

W niektórych sytuacjach stosowania metod nadzorowanych (*supervised*) może zdarzyć się, że w wyniku działań TDM powstanie nowy utwór. Jest to warunek konieczny do powoływania się na prawo cytatu, co może w pewnych sytuacjach okazać się przydatne.

Kluczowe dla określenia granic legalnego stosowania metod TDM jest spełnienie tzw. testu trójstopniowego, który został wdrożony w art. 35 PrAut. Zabrania on takich sposobów korzystania z przepisów o dozwolonym użytku, które naruszają normalne korzystanie z utworu lub godzą w słuszne interesy twórcy. Oba warunki mają charakter klauzul generalnych i konkretyzują się *in casu*. Dochodzi tu do zderzenia interesów. Z jednej strony znajdują się interesy autora, a z drugiej strony interes publiczny i interes badacza.

3. Archiwizacja i udostępnienie

Efekty badań naukowych powinny być archiwizowane i udostępniane po to, by umożliwić kolejnym badaczom wykorzystanie dotychczas wykonanej pracy, a także w celu weryfikacji. Możliwości udostępnienia zależą od rodzaju finalnego produktu. Nie ma oczywiście ograniczeń wynikających z prawa autorskiego, gdy produkt finalny nie wykorzystuje elementów twórczych utworów analizowanych. Jeśli elementy twórcze zostały przejęte, podstawą umożliwiającą udostępnianie, jest art. 27 ust. 2. Wprowadza on jednak warunek restrykcyjnie ograniczający zakres osób, którym można produkt finalny udostępnić.

Odrębną kwestią związaną z produktem finalnym jest ewentualne objęcie go ochroną prawnoautorską. Co do zasady należy przyjąć, że produkty finalne wytworzone w procesie TDM nie spełniają cech utworu i nie podlegają takiej ochronie ze względu na brak przejawu działalności człowieka oraz brak indywidualności. Nie można wykluczyć, że w wyjątkowych sytuacjach, raczej przy korzystaniu z metod nadzorowanych, zostaną spełnione przesłanki objęcia ochroną prawnoautorską. Efekty badań metodami TDM mogą stanowić jednak podstawę do opracowania utworu wykorzystującego te ustalenia, np. artykułu naukowego lub wizualizacji danych. Należy odróżnić w ramach procesu badawczego etap, w którym stosuje się metody TDM od innych etapów, w tym ewentualnego opracowania nowych utworów z wykorzystaniem produktów etapu TDM. Takie utwory uzyskiwałyby, oczywiście, ochronę prawnoautorską.

Zasada wolności badań naukowych

Granice dopuszczalności badań naukowych metodami TDM zależą, jak wcześniej wspomniałem, od ważenia interesów autora, badacza i społeczeństwa. Prawo autorskie jest prawem bezwzględny. Nie wolno wyjątków od tej zasady interpretować rozszerzająco, ani stosować analogii. Na drugiej szali kładzie się jednak konstytucyjną zasadę wolności badań naukowych, której wyrazem są niektóre przepisy o dozwolonym użytku. Wbrew dotychczas wyrażanej interpretacji, wolność badań naukowych polega nie tylko na zakazie ingerencji państwa w proces badawczy. Polega także na wymogu interpretowania prawa na korzyść możliwości prowadzenia badań naukowych dopóty, dopóki nie następuje konflikt z uprawnieniem innej osoby.

Interpretacja przepisów PrAut uniemożliwiająca prowadzenie badań naukowych metodami TDM stanowiłaby pogwałcenie konstytucyjnej zasady wolności badań naukowych, jeśli tylko dopuszczalna byłaby interpretacja umożliwiająca stosowanie tych metod, ponieważ byłoby to naruszenie możliwego do znalezienia balansu.

Sprowadzając powyższe wnioski na poziom jeszcze bardziej ogólny: zwyczaj analizowania statusu prawnego wszystkich łącznie metod TDM jest zbyt dużym uproszczeniem. Pewne przepisy prawnoautorskie upoważniają do stosowania pewnych, ale nie wszystkich, metod TDM w celach naukowych. Nie można odpowiedzieć poprawnie, że wszystkie metody TDM są dozwolone lub są niedozwolone. Należałoby dokonać systematyki w obrębie metod TDM (zręby takiej systematyki zaproponowałem w niniejszej pracy) i dopiero w oparciu o taką systematykę zbadać granice uprawnień. Granice te wyznaczone są przez ważenie interesów twórcy, badacza i społeczeństwa. Zasadzie ochrony twórcy przeciwstawia się zasada wolności badań naukowych.

Z analizy zaprezentowanej w poprzednich rozdziałach wynika potrzeba uregulowania zasad stosowania metod TDM w PrAut., tak by jasno i precyzyjnie wyznaczyć dopuszczalne granice tych działań. Newralgicznymi kwestiami jest dopuszczalność zwielokrotniania utworów na potrzeby badań oraz ingerencji w ich integralność. Z reguły metodami TDM posługują się zespoły badawcze, a nie indywidualni naukowcy. Obecnie instytucje naukowe nie dysponują podstawą do zwielokrotniania utworów niezależnie od ich wielkości. W przypadku wielu metod TDM jest to bardzo dotkliwe ograniczenie.

Prawdopodobnie, jeśli polski ustawodawca nie zadba o odpowiednią regulację we własnym zakresie, we względnie niedługim czasie zostanie do tego zobligowany z powodu wspólnej polityki w tym zakresie Unii Europejskiej.

Niezależnie od działań, które może podjąć ustawodawca, należy skierować rekomendacje do podmiotów udostępniających swoje utwory: w celu minimalizacji niepewności prawnej powinni udostępniać swoje dane (w tym teksty) na licencjach wyraźnie uprawniających do prowadzenia badań metodami TDM, przynajmniej do celów niekomercyjnych. Takie podmioty powinny rozważyć udzielenie licencji zezwalającej na TDM także w celach komercyjnych, ponieważ granica oddzielająca cele komercyjne od niekomercyjnych jest bardzo trudna do wyznaczenia. Pozostaje to jednak autonomiczną decyzją takiego podmiotu. Nawet gdyby pewne metody TDM były dopuszczalne bez takich licencji, to dzięki wyraźnemu zezwoleniu sytuacja prawna prowadzących badania będzie klarowna.

Bibliografia

Orzecznictwo:

- Uchwała SN w składzie siedmiu sędziów z dn. 16.07.1993 r., I PZP 28/93
- Wyrok SN z dn. 30 czerwca 2005 r., sygn. IV CK 763/04.
- Wyrok SA w Poznaniu z dn. 7 listopada 2007 r., sygn. I Aca 800/07.
- Wyrok TK z dn. 28 kwietnia 2009 r., sygn. K 27/07.
- Wyrok SN z dn. 22 czerwca 2010 r., sygn. IV CSK 359/09.
- Wyrok NSA z dn. 12 marca 2010 r., sygn. II FSK 1791/08.

Literatura:

- Banaszak B., Konstytucja Rzeczypospolitej Polskiej. Komentarz, wyd. 2, Warszawa 2012.
- Barta J., Markiewicz R., Prawa autorskie i prawa pokrewne, wyd. 6, Warszawa 2014.
- Barta J., Markiewicz R., w: Prawo autorskie i prawa pokrewne. Komentarz, red. J. Barta, R. Markiewicz, wyd. 5, Warszawa 2011.
- Błęzyński J., Prawo autorskie, wyd. 3, Warszawa 1988.
- Borghi M., Karapap S., Copyright and Mass Digitization. A Cross-Jurisdictional Perspective, Oxford University Press, 2013.
- Charalampakis B. et al., A comparison between semi-supervised and supervised text mining techniques on detecting irony in greek political tweets, "Engineering Applications of Artificial Intelligence, 51(2016).
- Cocoru D., Boehm M., An analytical review of text and data mining practices and approaches in Europe Policy recommendations in view of the upcoming copyright legislative proposal, Open Forum Europe, 1.05.2016, <http://www.openforumeurope.org/ofe-publishes-high-level-policy-paper-text-data-mining/>.
- Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions. Towards a modern, more European copyright framework, Bruksela, 09.12.2015.
- Crain S., Zhou Ke i in., „Dimensionality reduction and topic modeling: from latent semantic indexing to latent dirichlet allocation and beyond”, w: Mining Text Data.
- Czub K., Prawa osobiste twórców dóbr niematerialnych. Zagadnienia konstrukcyjne, Warszawa 2011.
- Dzielenie się danymi to tylko kolejna składowa komunikacji naukowej – z Timem Smithem (CERN) rozmawia Maciej Chojnowski, <http://otwartanauka.pl/analysis/rozmowy/dzielenie-sie-danymi-to-tylko-kolejna-skladowa-komunikacji-naukowej>.
- Flisak D., w: Prawo autorskie i prawa pokrewne. Komentarz, red. D. Flisak, Warszawa 2015.
- Giesen B., Własnościowy model prawa autorskiego – analiza koncepcji przyjętej w prawie polskim, [w:] „Ruch Socjologiczny, Ekonomiczny i Prawniczy”, 2/2015.
- Górska A., Górski R., Tworzenie i udostępnianie korpusów tekstowych – problemy na gruncie prawa autorskiego, w: Prace Instytutu Prawa Własności Intelektualnej UJ, Zagadnienia Prawa Autorskiego, Z. 97, red. A. Matlak, Kraków 2007.

- Handke Ch., Guibault L., Vallbé J.-J., Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research (June 7, 2015). Available at SSRN: <http://ssrn.com/abstract=2608513>.
- Hangal S. i inni, Historical Research Using Email Archives, Extended Abstracts of the ACM CHI'15 Conference on Human Factors in Computing Systems, <http://mobisocial.stanford.edu/papers/chi2015.pdf>.
- Hasan I., Parapar J., Blanco R., Segmentation of legislative documents using a domain-specific lexicon, "19th International Conference on Database and Expert Systems Application".
- Jarosz-Żukowska S., Żukowski Ł., Wolność badań naukowych i nauczania, w: „Realizacja i ochrona konstytucyjnych wolności i praw jednostki w polskim porządku prawnym”, Wrocław 2014, oai:www.repozytorium.uni.wroc.pl:53684.
- Jing Jiang, Information extraction from Text, w: „Mining Text Data”.
- Krallinger M., Erhardt R.A., Valencia A., Text Mining approaches in molecular biology and biomedicine, "Drug Discovery Today: Biosilico.", 2005, 10.
- Machała W., Dozwolony użytek prywatny w polskim prawie autorskim, Warszawa 2003.
- Machała W., Dozwolony użytek utworów w polskim prawie autorskim w świetle cywilistycznej koncepcji prawa podmiotowego, w: Prace z Wynalazczości i Ochrony Własności Intelektualnej, rok 2001, Z. 78.
- Machała W., Utwór. Przedmiot prawa autorskiego, Warszawa 2013.
- Małek L., Cytat w świetle prawa autorskiego, Warszawa 2011.
- Marcinkowska J., Dozwolony użytek w prawie autorskim. Podstawowe zagadnienia, Prace Instytutu Prawa Własności Intelektualnej UJ, z. 87, Zakamycze 2004.
- Markiewicz R., Ochrona prac naukowych, „Zeszyty Naukowe UJ”, z. 55/1990.
- McCarty W., A Telescope for a mind, w: „Debates in the Digital Humanities”, red. M. Gold, 2012, <http://dhdebates.gc.cuny.edu/debates/text/37>.
- Michalak A., Interes publiczny i jego oddziaływanie na powstanie, treść i wykonywanie praw własności intelektualnej, Warszawa 2012.
- Millions of Bush administration e-mails recovered, <http://edition.cnn.com/2009/POLITICS/12/14/white.house.emails>.
- Mining Text Data, red. Charu C. Aggarwal, ChengXiang Zhai, Springer 2012.
- Moretti F., Wykresy, mapy, drzewa. Abstrakcyjne modele na potrzeby historii literatury, Kraków 2016.
- Nenkova A., McKeown K., A Survey of Text Summarization, w: „Mining Text Data”.
- Nie J., Gao J., Cao G., „Translingual mining from text data”, w: Mining Text Data.
- Niżankowska A-M., Prawo do integralności utworu, Warszawa 2007.
- Piedra D., Ferrer A., Gea J., Text Mining and Medicine: Usefulness in Respiratory Diseases, Arch Bronconeumol. 2014;50(3):113.
- Rubisz Sz., Istota i proces kształtowania się instytucji dozwolonego użytku w prawie autorskim, w: Zagadnienia prawa autorskiego, Prace Instytutu Prawa Własności Intelektualnej UJ, z. 93, 2006.
- Siguenza-Guzman L. et al., Literature Review of Data Mining Applications in Academic Libraries, "The Journal of Academic Librarianship" 41 (2015).
- Siewicz K., Analiza prawna „Creative Commons 0”, http://koed.org.pl/wp-content/uploads/2012/09/CC0_analiza.pdf.
- Sikorski R., Ocena dozwolonego użytku w prawie autorskim w świetle kryteriów testu trójstopniowego,

- „Zarys prawa własności intelektualnej”, t. I „Granice prawa autorskiego”, red. J. Kępiński, K. Klafkowska-Waśniowska, R. Sikorski, Warszawa 2010.
- Simpson M. S., Demner-Fushman D., „Biomedical Text Mining: A Survey of Recent Progress”, w: Mining Text Data.
 - Słownik Języka Polskiego PWN, <http://sjp.pwn.pl/sjp/rutynowy;2574573.html>.
 - Sobczak J., Wolność badań naukowych. Standardy europejskie i rzeczywistość polska, „Roczniki Nauk Społecznych”, 1/2008.
 - Sohrabi M.K., Akbari S., A comprehensive study on the effects of using data mining techniques to predict tie strength, “Computer in Human Behavior” 60 (2016).
 - Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining. Report from the expert group, zespół pod kierunkiem I. Hargreavesa, Unia Europejska 2014.
 - Stanisławska-Kloc S., w: Prawo autorskie i prawa pokrewne. Komentarz, red. D. Flisak, Warszawa 2015.
 - Suber P., Knowledge as a Public Good, w: P. Suber, „Knowledge unbound. Selected writings on open access 2002-2011”, Cambridge, Massachusetts, London, 2016, <https://mitpress.mit.edu/books/knowledge-unbound>
 - Suber P., Otwarty dostęp, tłum. Warszawa 2014, <https://depot.ceon.pl/handle/123456789/6598>.
 - Swanson D., Undiscovered public knowledge, The Library Quarterly, 1986/56(2).
 - Trappey Ch., Trapey A., Collective intelligence applied to legal e-discovery: A ten-year case study of Australia franchise and trademark litigation, “Advanced Engineering Informatics” 29 (2015).
 - Triaille J-P. i inni, Study on the legal framework of text and data mining (TDM), 2014.
 - Truyens M., Van Eecke P., Legal aspects of text mining, LREC Conference 2014, Proceedings, <http://www.lrec-conf.org/proceedings/lrec2014/index.html>.
 - Traille J.-P., J. de Meeûs d'Argenteuil, A. Francquen, Study on the legal framework of text and data mining (TDM), 2014.
 - Targosz T., w: Prawo autorskie i prawa pokrewne. Komentarz, red. D. Flisak, Warszawa 2015.
 - Traple E., w: Prawo autorskie i prawa pokrewne. Komentarz, red. J. Barta, R. Markiewicz, wyd. 5, Warszawa 2011.
 - Understanding Knowledge as a Commons. From Theory to Practice, red. Ch. Hess, E. Ostrom, Cambridge, Massachusetts, London, 2007.

Strony internetowe:

- <https://blogs.ucl.ac.uk/copyright/2015/12/23/eu-reform-plans-text-and-data-mining-tdm/>
- <http://www.communia-association.org/2016/01/08/what-the-diary-of-anne-frank-can-tell-us-about-text-and-data-mining/>.
- <https://creativecommons.pl/poznaj-licencje-creative-commons/>.
- http://europa.eu/rapid/press-release_IP-15-6261_en.htm.
- <http://home.cern/about/computing>
- <http://project.futuretdm.eu/project-overview/>.
- <http://sparcopen.org/news/2015/the-right-to-read-is-the-right-to-mine/>
- <https://wiki.surfnet.nl/display/OSCFA/2.+Facilitate+text+and+data+mining+of+content>.
- http://www.wipo.int/wipo_magazine/en/2016/02/article_0001.html

■ https://www.whitehouse.gov/omb/fedreg_a110-finalnotice.

■ <https://wordnet.princeton.edu/>