

Daniel Dzienisiewicz

ORCID 0000-0003-0400-5143

Uniwersytet im. Adama Mickiewicza w Poznaniu

Piotr Nowak

ORCID 0000-0001-8027-2396

Uniwersytet im. Adama Mickiewicza w Poznaniu

Piotr Wierzchoń

ORCID 0000-0002-7658-5362

Uniwersytet im. Adama Mickiewicza w Poznaniu

Ku *Lexykonowi cenzury*. Zastosowanie metod sztucznej inteligencji w badaniach nad cenzurą w okresie XVIII-XXI w.

Toward the *Lexicon of Censorship*. The application of artificial intelligence methods to the study of censorship in the 18th-21st century

Abstract

The aim of the article is to present a complex of methods aimed at compiling the *Lexicon of Censorship*. In particular, the authors' intention is to discuss the problems which have been encountered in developing a list of entries for the planned publication and to indicate what novelty it will introduce to research on censorship. The authors discuss the previous works on censorship and describe the procedures for creating the lexicon with the use of digitized text resources contained in Polish digital libraries and other digital databases of texts (press, journalistic, artistic, etc.). The methods used were discussed, taking into account the use of information contained in the literature on the subject and artificial intelligence techniques (including photodocumentation, word vectorization or transformer architecture). In addition, the authors present exemplary results of their work.

Keywords

censorship, censorship of printed matter, photodocumentation, linguochronologization, artificial intelligence

Abstrakt

Celem artykułu jest przedstawienie kompleksu metod zastosowanych podczas opracowywania *Lexykonu cenzury*. W szczególności tekst ma na celu omówienie problemów związanych z opracowaniem siatki haseł planowanego wydawnictwa oraz wskazanie, jakie *novum* wprowadzi ono do badań nad cenzurą. Autorzy omawiają dotychczasowe prace poświęcone zjawisku cenzury oraz opisują procedury tworzenia *Lexykonu cenzury* z wykorzystaniem zdigitalizowanych zasobów tekstowych zawartych w polskich bibliotekach cyfrowych i innych bazach cyfrowych tekstów (prasowych, publicystycznych, artystycznych i in.). Omówione zostały metody pracy uwzględniające wykorzystanie informacji zawartych w literaturze przedmiotu i technik sztucznej inteligencji (m.in. fotodokumentacji, wektoryzacji słów czy architektury transformerów). Ponadto autorzy przedstawiają przykładowe wyniki swoich dotychczasowych prac.

Słowa kluczowe

cenzura, kontrola druku, fotodokumentacja, linguochronologizacja, sztuczna inteligencja

1. Wprowadzenie. Motywacja działań

Głównym celem niniejszego artykułu jest zaprezentowanie kompleksu optymalnych metod, które mają efektywnie doprowadzić do budowy *Leksykonu cenzury*, czyli opracowania o charakterze tezaursowym. Zgodnie z definicją zawartą w *Encyklopedii PWN* tezaurs to „słownik pojęciowy, w którym słownictwo jest uporządkowane wg kategorii tematycznych, a artykuły hasłowe zawierają wykaz wyrażen pokrewnych znaczeniowo wobec danego hasła (najczęściej synonimy i bliskoznaczniki, niekiedy też hiperonimy, hiponimy i antonimy)”¹. W szczególności chcemy omówić problemy, jakie napotkaliśmy w zakresie opracowania siatki haseł związanych ze zjawiskiem cenzury dla tekstów polskojęzycznych.

Badania nad cenzurą, zwłaszcza nad jej prewencyjnym modelem, który funkcjonował w Polsce Ludowej, wyjątkowo intensywnie zaczęły rozwijać się po roku 1989, gdy rozpoczął się proces transformacji ustrojowej, zniesiono Główny Urząd Kontroli Prasy, Publikacji i Widowisk, a jednocześnie zaczęto udostępniać materiały archiwalne dokumentujące działalność socjalistycznych organów cenzury. Niektóre z tych źródeł stały się podstawą interesujących edycji. Z wydanych w ostatnim czasie na wyróżnienie zasługuje np. cykl materiałów przygotowanych w Katedrze Badań Filologicznych nad cenzurą PRL i Edytorstwa Uniwersytetu w Białymstoku, kierowanej przez Kamilę Budrowską². W efekcie rozwoju badań nad cenzurą liczba publikacji poświęconych organom cenzury, autocenzurze, drugiemu obiegowi itp. na przełomie drugiej i trzeciej dekady obecnego wieku liczona jest już nie w setkach, lecz w tysiącach pozycji. Ze względu na dużą liczbę opracowań poświęconych przedmiotowemu zagadnieniu wskaźmy wyłącznie na wybrane prace dotyczące cenzury funkcjonującej na ziemiach polskich w różnych okresach. Jeśli chodzi o cenzurę przed rokiem 1944, to na wymienienie zasługują m.in. książki autorstwa J.S. Bandkiego, P. Buchwald-Pelcowej, T. Gutowskiego, R. Habielskiego, F. Hoesicka, Z. Kmiecika, G. Kucharczyka, M. Mlekickiej, M. Pietrzaka, M. Rajcha, B. Szyndlera, J. Tazbira, Z. Zajewskiego³. Do grupy najważniejszych książek dotyczących okresu późniejszego

¹ J. Wojnowski, *Wielka Encyklopedia PWN*, t. 27, Warszawa 2005, s. 371.

² K. Budrowska, K. Kościewicz, *„Dzieła” Henryka Sienkiewicza w dokumentach Głównego Urzędu Kontroli Prasy, Publikacji i Widowisk (1948–1954)*, Białystok 2016.

³ J.S. Bandtkie, *Historia drukarni krakowskich*, Kraków 1815; Tenże, *Historia drukarni w Królestwie Polskim i Wielkim Księstwie Litewskim*, Kraków 1826; P. Buchwald-Pelcowa, *Cenzura w dawnej Polsce. Między prasą drukarską a stosem*, Warszawa 1997; T. Gutowski, *Cenzura w Wolnym Mieście Krakowie 1832–1846*, Kraków 1914; R. Habielski, *Wolność czy odpowiedzialność? Prasa i polityka w II Rzeczypospolitej*, Warszawa 2013; F. Hoesick, *Ze wspomnień o cenzurze rosyjskiej w Warszawie*, Warszawa 1929; F. Hoesick, *Ze wspomnień o cenzurze rosyjskiej w Warszawie*, Warszawa 1929; G. Kucharczyk, *Cenzura pruska w Wielkopolsce w czasach zaborów 1815–1914*, Poznań 2001; Z. Kmiecik, *Prasa warszawska w latach 1886–1904*, Wrocław 1989; M. Mlekicka, *Wydawcy książek w Warszawie w okresie zaborów*, Warszawa 1987; M. Pietrzak, *Reglamentacja wolności prasy w Polsce (1918–1939)*, Warszawa 1963;

zalicza się natomiast prace K. Budrowskiej, M. Ciećwierza, B. Gogola, R. Habielskiego, S.A. Kondka, M. Korczyńskiej-Derkacz, A. Kozieła, E. Kristanovej, D. Nałęcz, A. Pawlickiego, Z. Romka⁴ i in. Na odnotowanie zasługują także liczne prace ujmujące problematykę cenzury za „żelazną kurtyną”, jak choćby A.A. Reischa czy F. Kind-Kovácsa⁵.

Badacze polskiej cenzury wskazują zarówno na problemy z dotarciem do źródeł dokumentujących jej funkcjonowanie, jak i – z powodu rozproszenia – na trudności w odnajdywaniu opracowań na jej temat o charakterze przyczynkarskim. Jak pisze autorka jednej z najnowszych monografii działalności organów cenzury w Polsce powojennej, Kamila Kamińska-Chełminiak, jej możliwości badawcze „były determinowane przez zakres dostępnych źródeł, które – jak podkreślają niemal wszyscy cenzurologowie zetknąwszy się ze spuścizną GUKPPiW – są rozproszone i pełne luk”⁶. Kamila Budrowska wraz z Katarzyną Kościewicz potwierdzają powyższe spostrzeżenia, stwierdzając w 2016 roku niemożność bliższego oszacowania kolekcji źródeł wytworzonych przez Główny Urząd Kontroli Prasy, Publikacji i Widowisk: „Wielkość zespołu GUKPPiW szacuje się różnie: od pięciu do siedmiu tysięcy jednostek, przede wszystkim teczek, znacznie rzadziej skoroszytów i kartotek. Jest to zasób pod względem merytorycznym, tematycznym i formalnym bardzo zróżnicowany, a przy tym niekompletny (...). Wynikają stąd oczywiście trudności związane z ich kwalifikowaniem”⁷. Jednocześnie – jak powiedzieliśmy powyżej – problem prowadzenia kwerend napotykały podczas wyszukiwania opracowań, gdyż wiele z nich zawiera

M. Rajch, *Cenzura pruska w Wielkopolsce w latach 1948–1918*, Poznań 2004; B. Szyndler, *Dzieje cenzury w Polsce do 1918 roku*, Warszawa 1993; J. Tazbir, *Państwo bez stosów. Szkice z dziejów tolerancji w Polsce XVI i XVII wieku*, Warszawa 2009; Z. Zajewski, *Wolność druku w powstaniu listopadowym 1830–1831*, Łódź 1963.

⁴ K. Budrowska, *Literatura i pisarze wobec cenzury PRL 1948–1958*, Białystok 2009; M. Ciećwierz, *Polityka prasowa 1944–1948*, Warszawa 1989; B. Gogol, *„Fabryka fałszywych tekstów”. Z działalności Wojewódzkiego Urzędu Kontroli Prasy, Publikacji i Widowisk w Gdańsku w latach 1945–1958*, Warszawa 2012; R. Habielski, *Polityczna historia mediów w Polsce XX wieku*, Warszawa 2009; S.A. Kondek, *Papierowa rewolucja. Oficjalny obieg książki w Polsce w latach 1948–1955*, Warszawa 1999; A. Kozieł, *Studium o polityce prasowej PZPR w latach 1948–1957*, Warszawa 1991; M. Korczyńska-Derkacz, *Państwowy Instytut Książki (1946–1949) i jego rola w rozwoju bibliologii, bibliotekarstwa i kultury książki w Polsce*, Wrocław 2011; E. Kristanova, *Książka na łamach katolickich czasopism społeczno-kulturalnych w Polsce w latach 1945–1953*, Łódź 2012; D. Nałęcz, *Główny Urząd Kontroli Prasy 1945–1949*, Warszawa 1994; A. Pawlicki, *Kompletna szarość. Cenzura w latach 1965–1972*, Warszawa 2001; Z. Romek, *Cenzura a nauka historyczna w Polsce 1944–1970*, Warszawa 2010.

⁵ A.A. Reisch, *Hot Books in the Cold War. The CIA-Funded Secret Western Book Distribution Program Behind the Iron Curtain*, Budapest 2013; F. Kind-Kovács, *Written Here, Published There: How Underground Literature Crossed the Iron Curtain*. Budapest 2014.

⁶ K. Kamińska-Chełminiak, *Cenzura w Polsce 1944–1960. Organizacja – kadry – metody pracy*, Warszawa 2019, s. 25.

⁷ K. Budrowska, K. Kościewicz, dz. cyt., s. 12.

istotne informacje na temat cenzury, niemniej jednak nie można do nich dotrzeć z powodu opatrzenia ich metadanymi niepowiązanymi z cenzurą.

Naszym nadrzędnym celem będzie zatem stworzenie autorskiej kolekcji materiałów tekstowych, które nie były dotąd eksplorowane przez cenzurologów ze względu na rozproszenie tego typu treści w różnych publikacjach (zarówno opracowań, jak i źródeł). Tak zaprojektowany zbiór może bowiem stanowić uzupełnienie znanych materiałów źródłowych oraz wyników badań nad kontrolą druków. Uwzględnienie licznych, niezbadanych dotąd tekstów (prasowych, publicystycznych, artystycznych i in.) będzie możliwe dzięki zastosowaniu technologii cyfrowych, ze szczególnym uwzględnieniem metod sztucznej inteligencji, umożliwiających przeszukiwanie zasobów polskich bibliotek cyfrowych i innych baz zdigitalizowanych tekstów w łącznej liczbie kilku milionów publikacji. Istotnym elementem projektu jest zastosowanie fotodokumentacji, tj. przedstawienie materiałów tekstowych w postaci skanów-fotografii, umożliwiających kontakt z tekstem w jego autentycznej postaci graficznej. Należy zauważyć, że technika fotodokumentacyjna była dotychczas stosowana głównie w badaniach lingwistycznych z zakresu lingwochronologizacji, mających na celu ukazanie najwcześniejszych poświadczeń wyrazów w tekstach⁸.

Motywacją dla naszego działania jest zapełnienie luki w zakresie:

- a. opracowania metod wydobycia z tekstów polskich okresu XIX–XX syntezy wiedzy dotyczącej szeroko pojętej cenzury związanej z procesem publikacji w postaci relewantnych cytatów (np. paragrafów), zwłaszcza mając na uwadze metody sztucznej inteligencji,
- b. stworzenia finalnej siatki haseł związanych z cenzurą,
- c. ilustracji tych pojęć oryginalnymi cytatami z druków polskojęzycznych (głównie prasy oraz książek zgromadzonych w zasobach polskich bibliotek cyfrowych).

2. Metodologia tworzenia siatki haseł z wykorzystaniem sztucznej inteligencji

Materiał tekstowy stanowiący podstawę projektu pozyskany zostanie z zasobów polskich bibliotek cyfrowych oraz innych źródeł tekstów w formie cyfrowej. W tym celu kwerendzie poddane zostaną zasoby takich baz, jak m.in.:

Cyfrowa Biblioteka Narodowa Polona, Jagiellońska Biblioteka Cyfrowa, Śląska Biblioteka Cyfrowa, Wielkopolska Biblioteka Cyfrowa, Kujawsko-Pomorska

⁸ W tej kwestii zob. np. opracowania: P. Wierzchoń, *Fotodokumentacja, chronologizacja, emendacja. Teoria i praktyka weryfikacji materiału leksykalnego w badaniach lingwistycznych*, Poznań 2008; Tenże, *Słownictwo lat 30. XX w. w obrazach i liczbach. Tylko 100 fotoprzykładów z Depozytorium leksykalnego języka polskiego i dużo wykresów z systemu Odkrywka*, Warszawa 2015.

Biblioteka Cyfrowa, Repozytorium Cyfrowe Instytutów Naukowych (RCIN), Biblioteka Cyfrowa Uniwersytetu Wrocławskiego, Małopolska Biblioteka Cyfrowa, Biblioteka Cyfrowa – Regionalia Ziemi Łódzkiej, Biblioteka Multimedialna Teatru NN, Pomorska Biblioteka Cyfrowa, Biblioteka Cyfrowa Uniwersytetu Łódzkiego, Mazowiecka Biblioteka Cyfrowa, Biblioteka Cyfrowa Politechniki Śląskiej, Cyfrowe Muzeum Narodowe w Warszawie, Elbląska Biblioteka Cyfrowa, Bałtycka Biblioteka Cyfrowa, Biblioteka Cyfrowa Ośrodka KARTA, Podlaska Biblioteka Cyfrowa, Zielonogórska Biblioteka Cyfrowa, Biblioteka Cyfrowa KUL, Zachodniopomorska Biblioteka Cyfrowa „Pomerania”, Biblioteka Cyfrowa UMCS, Dolnośląska Biblioteka Cyfrowa itp.

Więcej informacji o polskich bibliotekach cyfrowych zob. na stronie internetowej Federacji Bibliotek Cyfrowych (FBC)⁹. Dnia 17 stycznia 2023 roku łączna liczba dokumentów dostępnych poprzez portal FBC, m.in. czasopism, obrazów, książek, archiwaliów, dokumentów urzędowych, rozpraw doktorskich i kartotek, wynosiła 8 318 722 obiektów pochodzących ze 154 źródeł, głównie bibliotek cyfrowych, lecz także archiwów, cyfrowych muzeów, repozytoriów i in.

Procedury konstruowania siatki haseł *Leksykonu cenzury* obejmują:

- a. analizę siatki haseł anglosaskich encyklopedii i słowników poświęconych cenzurze i wykorzystanie tych pojęć, które mają swoje odzwierciedlenie w rzeczywistości tekstów polskojęzycznych,
- b. wykorzystanie listy haseł ujętych w literaturze przedmiotu,
- c. wykorzystanie haseł Wikipedii, w których definicjach użyte zostało pojęcie „cenzura”,
- d. ekscerpcje własne, polegające na eksploracji zasobów bibliotek cyfrowych,
- e. ekscerpcje własne polegające na uwzględnieniu w *Leksykonie* materiałów odnalezionych podczas prowadzenia badań nad cenzurą.

W szczególności nasza metodologia konstrukcji siatki haseł obejmuje składniki:

- a. analizę siatki haseł zbiorów anglosaskich, np. *The Encyclopedia of Censorship (Facts on File Library of World History)*¹⁰ oraz *Historical Dictionary of Censorship in the United States*¹¹, i wykorzystanie tych pojęć, które mają swoje odzwierciedlenie w rzeczywistości tekstów polskojęzycznych (twierdzenie to formułujemy na podstawie dotychczasowego oglądu dostępnego nam zbioru tekstów zaczerpniętych z polskich bibliotek cyfrowych), np.: „blacklisting”, „blasphemy”, „book burning and the Jews”, „book burning in Nazi

⁹ Federacja Bibliotek Cyfrowych, *Federacja Polskich Bibliotek Cyfrowych*, <https://fbc.pionier.net.pl/> (dostęp 17.01.2022).

¹⁰ J. Green, N. Karolides, *The Encyclopedia of Censorship (Facts on File Library of World History)*, New York 2005.

¹¹ L. Hurwitz, *Historical Dictionary of Censorship in the United States*, Westport 1985.

Germany”, „caricature”, „data protection”, „hate speech / hate crime”, „Holocaust denial”.

Należy nadmienić, że w świetle powyższego zamierzenia problem stanowią mogą te pojęcia, które są charakterystyczne wyłącznie dla krajów anglosaskich, m.in. terminy stosowane w anglosaskim systemie prawnym (*common law*), jak np.: actual malice, bawdy courts, book burning in England, conspiracy to corrupt public morals, conspiracy to outrage public decency, criminal syndicalism, derivative classification, examiner of plays, malice itd. Wstępne badanie dowiodło bowiem, że poza nielicznymi wyjątkami (np. „actual malice”, „examiner of plays”) nie są one wzmiankowane w polskich tekstach. Na rozwiązanie oczekuje zatem jeszcze kwestia ich ewentualnego włączenia w skład siatki haseł *Leksykonu cenzury*.

- b. wykorzystanie listy haseł ujętych w literaturze przedmiotu (zob. Bibliografia), jak np.: „Kodeks Haysa”, „autocenzura”, „blackout medialny”, „cenzura w Internecie”, „debit”, „imprimatur”, „index librorum prohibitorum”, „kontrola druku”, „nihil obstat”, „osoba publiczna”, „palenie książek”, „poprawność polityczna”, „samizdat”, „urzędnik publiczny” itd.
- c. wykorzystanie haseł Wikipedii, w których definicjach użyte zostało pojęcie „cenzury”, np.: Agatha Christie, Anatole France, Adam Mickiewicz, Andrzej Frycz Modrzewski, Baruch Spinoza, Blaise Pascal, Czesław Miłosz, De revolutionibus orbium coelestium, Gustave Flaubert, Giordano Bruno, Galileusz, Immanuel Kant, Józef Piłsudski, John Locke, Talmud, Mein Kampf, Indeks książek zakazanych, Niccolò Machiavelli, Alfred Szklarski, Jean-Jacques Rousseau, Johannes Kepler, Lew Trocki, Hrabia Monte Christo, Tadeusz Dołęga-Mostowicz, Josip Broz Tito, Zofia Kossak, Benito Mussolini, Henryk Glass, Voltaire, Jerzy Kirchmayer, Władysław Gomułka, List 34, Marian Spychalski, Mikołaj Kopernik, Jerzy Urban, Giacomo Casanova, Szczepan Jeleński, Szwajpolt Fiol, Monteskiusz, Denis Diderot, Jean le Rond d’Alembert, Jan Lemański (pisarz), Adam Zagajewski, Gabriele D’Annunzio, Stefan Mossor, Émile Zola, Faustyna Kowalska, Główny Urząd Kontroli Prasy, Publikacji i Widowisk, Stanisław Jerzy Lec, Encyklopedia „Białych Plam”, Biernat z Lublina, Nikołaj Bucharin, Franciszek Maksymilian Sobieszczański, Debit, Alexandre Dumas, Entertainment Software Rating Board, Jakub Przyłuski, Marcel Reich-Ranicki, Pan European Game Information, Nowa Heloiza, Entertainment and Leisure Software Publishers Association, Dekameron, Stefan Wiechecki, Mieczysław Lepecki, Józef Siemek itd. Procedura ta pozwoli na szybkie i efektywne odnalezienie informacji dotyczących instytucji stosujących cenzurę, przedmiotu cenzury czy okoliczności towarzyszących temu zjawisku, które to informacje będą mogły zostać następnie wykorzystane do prowadzenia kwerend w zbiorze zdigitalizowanych tekstów.

d. ekscerpcje własne (najbardziej znaczący wkład autorski w budowę *Leksykonu*), tzn. kumulacja tych pojęć i idei, które w tekstach polskich są skorelowane ze zjawiskiem cenzury:

- nazwiska
- pseudonimy
- miejsca
- instytucje
- ruchy społeczne
- tytuły literackie
- tytuły prasowe
- redakcje, dziennikarze
- kabarety
- filmy i ich twórcy
- granice historyczne itp.

Tak skonstruowana siatka haseł pozwala na sporządzenie kilkunastu tysięcy kwereńd dla uzyskania kilkuset tysięcy relewantnych cytatów, w których podejmuje się kwestie cenzury. Poniżej przedstawiono przykładowe fotoekscerpty tekstów zawierających w swojej treści wyrazy *cenzura*, *cenzorski*, *kontrola druku*, które odnalezione zostały za pomocą wyszukiwarki pełnotekstowej umożliwiającej przeszukiwanie zasobów polskich bibliotek cyfrowych i innych źródeł zdigitalizowanych tekstów¹²:

W ostatnim kroku opracowana zostanie ankieta zawierająca pytania niezbędne do przeprowadzenia anotacji zgromadzonych fotoekscerptów. Należy tutaj nadmienić, że anotacja jest pojęciem stosowanym w językoznawstwie korpusowym i oznacza przypisywanie zbiorom tekstów określonych informacji, np. takich jak autor, gatunek tekstu, części mowy, cechy prozodyczne i in. Oprócz wyodrębnienia fotocytatów przez zespół ekscerptorów planowane jest bowiem również ich sklasyfikowanie ze względu na różnorodne właściwości. Poniżej zaprezentowano przykładowy interfejs urządzenia służącego do efektywnej anotacji tekstów poświęconych cenzurze zaprojektowany w środowisku Doccano, które umożliwia przypisywanie tekstom etykiet zawierających odpowiedzi na pytania dotyczące kwestii będących przedmiotem anotacji¹³:

¹² Więcej informacji o ww. systemie wyszukiwawczym zob. np. w pracach: F. Galiński, *Against the Arrow of Time. Theory and Practice of Mining Massive Corpora of Polish Historical Texts for Linguistic and Historical Research*, Poznań 2019; F. Galiński, D. Dzienisiewicz, P. Wierzchoń, *U bram lingwistycznej szczęśliwości, czyli kulisy projektu Odkrywka: cyfrowe zasoby kultury jako źródło mas danych językowych*, „*Studia Sociologica*” 2017, z. 9, s. 51–62.

¹³ Doccano, *Open source annotation tool for machine learning practitioners*, <https://github.com/doccano/doccano> (dostęp 17.01.2022).

wśród słuchaczy. Niejedno oko lżą się émiło i niejedni zaparli się oddech, gdy wtędy w kwietniu r. 1863 bohater szwajcarski, ze sceny lwowskiej przemawiał.

Oto, jak z tej premiery w słowach szarych, bo krępowanych przez cenzurę, zdawał sprawę recenzent ówczesnego „Dziennika literackiego“:

„Wolno już wreszcie „Tellowi“ i na naszej scenie strzelać z łuku do ciemieży!... Pod tym względem to przedstawienie zapisane być winno w kronikach postępu.

Pliniusz z drugiej strony wspomina o *ensorium probrum* analogicznym do naszych przekroczeń policyjnych (*agrum male colere, censorium probrum iudicabatur*). Słusznie też mówi Jarek: Ten, kogo dotykała nota cenzorska, stał przed resztą ludu z awstyd zony.

Także satyryczny śpiew Grenlandczyków jest karą, ponieważ spotyka on winowajcę na zgromadzeniu ludu, gdzie całe masy przyjmują udział w śpiewaniu, a wiadomo, że na Grenlandczyków nie tak nie działa, jak publiczne zawstyżenie. Na Morilö uchodziła nagana

cej pokrycia w towarach - reglamentacji, zagwarantowanie, że różnica między normą obowiązującą o realną zostanie ludzkom zwrócona w terminie nie dłuższym, niż jeden miesiąc w wytypowanych w tym celu placówkach; zagwarantowanie pełnej społecznej kontroli druku i rozdzielności kartek; powołanie komisarzy związkowych kontrolujących dystrybucję żywności, którzy za pomocą środków masowego przekazu mogliby informować społeczeństwo np. o stanie magazynów. /"Goniec Małopolski" 9.11/

W s. zasad reglamentacji. 11.11 rozpoczą się strajk w zakładach galanterijnych w Łotomiersku k/Łodzi.

Złote myśli.

„Nasza cenzura jest rzeczywiście nie-skończenie głupia“ — powiedział pewnego razu cesarz austriacki Franciszek I. do swego podkomorzego hr. Czernina.

(Cytat ze str. 145 tom 29 biblioteki pisarzy niemieckich z Czech. Nakład c. k.

Doc. dr hab. Tomasz Kupis pytał o zakres ingerencji cenzorskich. Red. Dubiel odpowiedział, że ustawa o kontroli publikacji jest szczególnie niedogodna dla pism o tak długim cyklu produkcyjnym jak nasz, gdyż to co jest niecenzuralne dziś — gdy

A co się stanie, jeżeli bolszewicy tą drogą będą chcieli przysłać swoją bibułę o charakterze antypaństwowym? Czy i wtedy urząd pocztowy bezkrytycznie będzie na ich usługach?

Zagranicą wszędzie jest ścisła kontrola druków przesyłanych pocztą. Urząd pocztowy np. Stanów Zjednoczonych A. P. pociąga do odpowiedzialności osoby, wysyłające druki obrażające moralność publiczną. Mamy nadzieję, że Ministerstwo Poczty i Telegrafów wniknie w powyższą sprawę i odpowiednio pouczy podwładne urzędy.

Działalność cenzury jest więc nielegalna, nielegalne są też poczynania policji, trapiącej i konfiskującej sprzęt i gotowe nakłady. Półki co jednak, trzeba się policji wystrzeżać i zachować odpowiednią ostrożność.

W nr. 21 m.in. Wywiad z Raymondem Aronem: St. Gomulka: Sytuacja gospodarza Polski; JEZYK TOTALITARNY: A. Wat: Semantyka języka stalinowskiego; J. Etkind: Polityka i prawda; K. Mueller: Komunikacja sterowana; L. Bod: Język i polityka; I. Grudzińska-Gross: Świat zastraszonych słów; T. Strzyżewski: Zamiast posłowania do „Czarnej Księgi Cenzury PRL“; „WSPÓLRZĄDZIC CZY NIE KLAMAC“ — o książce A. Micewskiego: L. Dembiński, A. Pospieszalski.

Uprzednio w nr. 20: L. Kolakowski: Jak być konserwatywno-liberal-

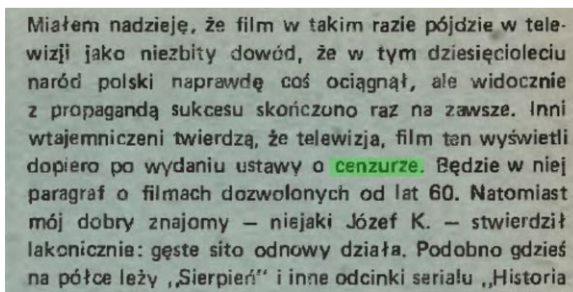
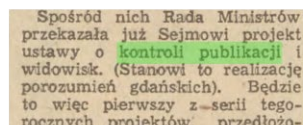
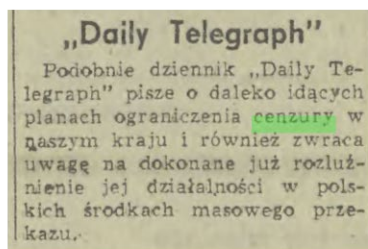
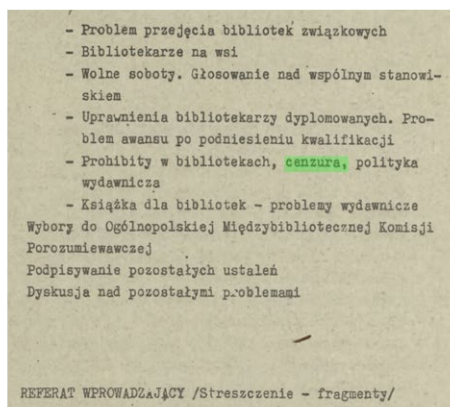
chnianiem treści szkodliwych obyczajowo (np. pornograficznych).

Komisja zajęła stanowisko, że niezbędne jest określenie zasad działania cenzury w sposób wyraźny i precyzyjny. Wyrażona też została w pełni zgodność z treścią postulatów, który mówił o podjęciu realnych działań mających na celu wyprowadzenie kraju z sy-

Stanisław Kania, do niedawna sekretarz KKKoordynujący wszelkie nieprawości czynione przez milicję, która pod jego to przeciwieństwem przewodem organizowała ścieżki zdrowia w Radomiu i Ursusie, kierowała bojówkami napadającymi na słuchaczy WGN i tak dalej i dalej - otóż tenże Stanisław Kania jest dziś rzeczniczem praworządności i demokracji.

Przeciw cenzurze występuje ostro w sejmie prof. Jan Szczepański, który niedawno jeszcze w tym samym sejmie rzucał gromy na brągan pozwalający marmować papier na niezłeczone wydawnictwa w sytuacji, kiedy tak bardzo tego papieru brakuje, za odnową i demokracją deklarują się Daniel Passent, Dominik Horodyński czy Roman Bratny. Więcej: w "Polityce" Marta Wesółowska, której przed nie-

Fot. 1a. Przykładowe fotoekscerpty dotyczące cenzury



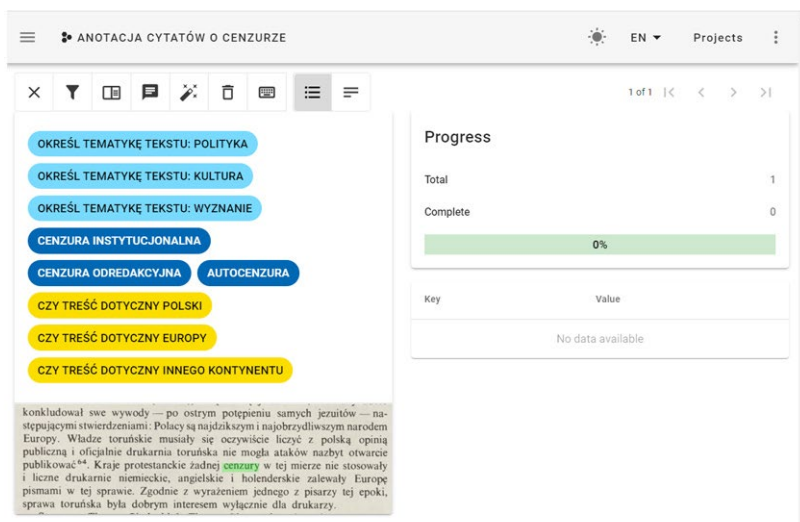
Fot. 1b. Przykładowe fotoekscerpty dotyczące cenzury

Wybrana, reprezentatywna część zgromadzonego materiału będzie anotowana według różnych parametrów. Anotacja uwzględni takie informacje, jak tematyka oraz okres, jakiego dotyczy cytaty, rodzaj źródła, czas powstania tekstu, osoby i utwory literackie, wobec których stosowano cenzurę, instytucje stosujące cenzurę i in. Następnie materiał tekstowy w formie fotoekscerptów, sklasyfikowany według powyższych kryteriów, zostanie włączony do leksykonu.

W zakresie użytych metod sztucznej inteligencji zastosowane zostaną takie mechanizmy, jak:

- wektoryzacja słów (np. word2vec)
- architektura transformerów (np. BERT)
- uczenie nadzorowane (np. fastText)
- rozpoznawanie obrazu (np. OpenCV)
- klasyfikatory (np. CountVectorizer).

Poniżej w związku sposób omówiono ich działanie. Wektoryzacja słów polega na przetwarzaniu języka naturalnego (NLP). Algorytmy tego typu wykorzystują model sieci neuronowej do uczenia się skojarzeń słów, zazwyczaj z dużego i różnorodnego korpusu tekstu. Po etapie treningu model może wykrywać słowa synonimiczne, a także zasugerować domyślne słowa dla zdania. Technika wektoryzacji



Fot. 2. System anotacji cytatów dotyczących cenzury. Ocena ankietowa

reprezentuje każde słowo za pomocą określonej listy liczb zwanej wektorem. Wektory te obliczane są, aby ustalić semantyczne i syntaktyczne cechy słów (rzadziej fraz). W konsekwencji – funkcja matematyczna (podobieństwo) może ukazywać stopień podobieństwa (zazwyczaj semantycznego) pomiędzy słowami reprezentowanymi przez te wektory. Przykład: *aeroplan* i *samolot* to pojazdy latające¹⁴.

W przypadku architektury transformerów (np. BERT), stosuje się model głębokiego uczenia, który przyjmuje mechanizm niejako samoobserwacji, odpowiednio ustalając różne wagi każdej części danych wejściowych. Architektura ta stosowana jest przede wszystkim w dziedzinach przetwarzania języka naturalnego (NLP), a także widzenia komputerowego. Metody te doskonale sprawdzają się w zadaniach dotyczących tłumaczenia czy streszczenia tekstu¹⁵.

Jeżeli na uwadze mieć uczenie nadzorowane, to jest to rodzaj uczenia maszynowego dla problemów, w których dostępne dane składają się z etykietowanych przykładów (labels). Oznacza to, że każdy wers danych musi zawierać cechy i związaną z nimi etykietę. Przykładowo, dla zdania: *Poproszę o sól* – będzie to etykieta: „Neutralne”. Dla zdania *Dawaj mi sól szybko* – będzie to etykieta: „Niegrzeczny rozkaz”. Celem algorytmów uczenia nadzorowanego jest uczenie się funkcji, która

¹⁴ Więcej o wektoryzacji słów zob. np. w artykule T. Mikolov, K. Chen, G. Corrado, J. Dean, *Efficient Estimation of Word Representations in Vector Space*, „Proceedings of Workshop at ICLR” 2013, <https://arxiv.org/abs/1301.3781> (dostęp 17.01.2022).

¹⁵ Więcej na ten temat zob. w pracy A. Rogers, O. Kovaleva, A. Rumshisky, *A Primer in BERTology: What We Know About How BERT Works*, „Transactions of the Association for Computational Linguistics” 2020, 8, s. 842–866.

odwzorowuje wektory cech na etykiety. W tym celu algorytm uczenia nadzorowanego analizuje dane treningowe i wytwarza funkcję wyjściową, która może być wykorzystana do odwzorowania nowych przykładów, nieznanych przez algorytm¹⁶.

Widzenie komputerowe to zaś interdyscyplinarna dziedzina, która zajmuje się uzyskiwaniem przez komputery tzw. rozumienia cyfrowych obrazów (lub filmów). W przypadku materiałów zdigitalizowanych, tzn. takich, gdzie wejściem filologicznym jest fotografia tekstu, komponent ten ma istotne znaczenie. Zadania rozpoznawania komputerowego obejmują metody pozyskiwania, przetwarzania i analizowania obrazów cyfrowych, zazwyczaj dla podjęcia określonych decyzji (np. sortowanie owoców wedle wielkości wykrytej na fotografii tej rośliny) – więcej informacji o widzeniu komputerowym por. np. na oficjalnej stronie projektu OpenCV¹⁷.

Klasyfikatory są natomiast ogólnym określeniem większości mechanizmów sztucznej inteligencji i sprowadzają się do wytyczenia klas, z reguły na podstawie nauczania nadzorowanego wedle wcześniej zdefiniowanej klasy. Numer klasy jest zatem decyzją, którą należy podjąć. Tradycyjne metody klasyfikacji obejmują m.in. algorytm k nearest neighbours, SVM i naiwny klasyfikator Bayesa¹⁸.

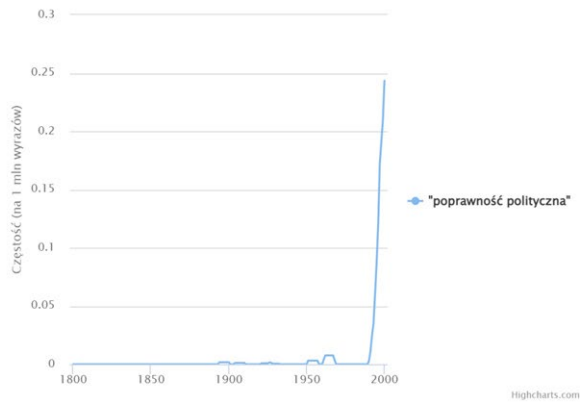
3. Oczekiwany rezultat

Zamierzonym rezultatem końcowym projektu jest opracowanie *Leksykonu* na podstawie cyfrowej bazy materiałowej, która jest punktem wyjścia do badań nad systemem kontroli przekazu w Polsce od czasów porozbiorowych do współczesności, jak również przygotowanie wybranych (reprezentatywnych) opracowanych w ten sposób materiałów do druku w postaci monografii. Cyfrowy zasób pozwoli na przeszukiwanie w otwartym dostępie obszernej bazy danych fotocytatograficznych i bibliograficznych poświęconych cenzurze druków oraz wgląd w stosowne fragmenty zgromadzonych tekstów dotyczących przedmiotowego zagadnienia. Oprócz tego rezultatem badań będzie analiza uzyskanego korpusu źródeł pod kątem frekwencji występowania informacji dotyczących funkcjonowania organów cenzury (będących narzędziami kontroli sprawowanymi przez Kościół, państwa itp.) w zależności od okresu, w jakim wystąpiły, typu źródła (druki urzędowe, prawnicze, piśmiennictwo naukowe, beletrystyka, publicystyka prasowa i in.) oraz instytucji, jaką reprezentowały (Kościół, państwa, organizacje społeczne i in.). Dodatkową wartością badań będą wykresy frekwencyjne, które uzyskamy po raz pierwszy w literaturze

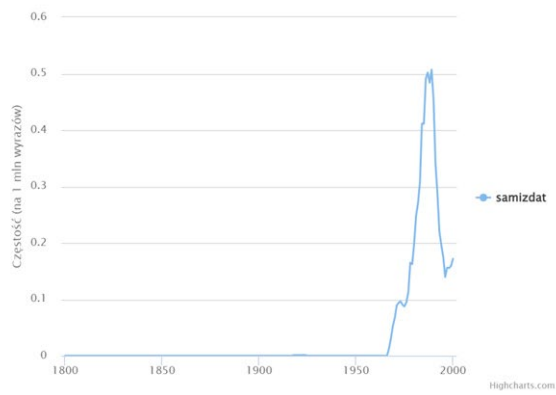
¹⁶ W kwestii uczenia nadzorowanego zob. artykuł A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jegou, T. Mikolov, *FastText.zip: Compressing text classification models*, <https://arxiv.org/pdf/1612.03651.pdf> (dostęp 17.01.2022).

¹⁷ OpenCV, *Home – OpenCV*, <https://opencv.org> (dostęp 17.01.2022).

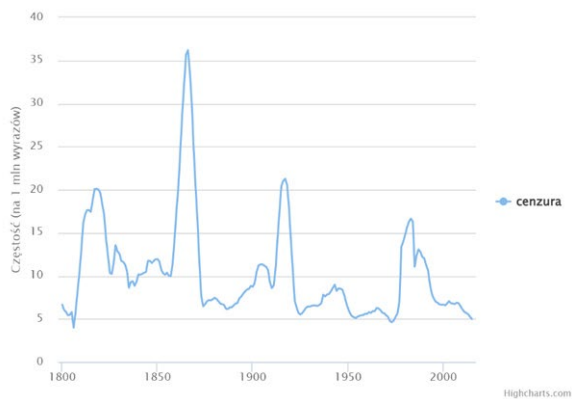
¹⁸ W kwestii metod klasyfikacji por. np. pracę T. Cover, P. Hart, *Nearest neighbor pattern classification*, „IEEE Transactions on Information Theory” 1967, 13, s. 21–27.



Fot. 3. Frekwencja wyrazu *poprawność polityczna* na przestrzeni lat 1800–2000



Fot. 4. Frekwencja wyrazu *samizdat* na przestrzeni lat 1800–2000



Fot. 5. Frekwencja wyrazu *cenzura* na przestrzeni lat 1800–2000

przedmiotu dla pojęć związanych z cenzurą. Prezentowane poniżej wykresy frekwencji również zostały wygenerowane automatycznie za pomocą narzędzi wspomnianej wcześniej wyszukiwarki, która umożliwia obliczanie częstości występowania wyrazów i połączeń wyrazowych w zasobach zdigitalizowanych tekstów.

Leksykon cenzury wprowadzi dotychczasowe badania nad cenzurą na wyższy poziom pod względem źródłowym. Uważamy, że dzięki *Leksykonowi* bardzo istotnie zwiększy się widoczność źródeł i opracowań wykorzystywanych przez cenzurologów, badających działalność nie tylko polskich organów cenzury. Przede wszystkim pojawi się szansa na wykrycie zupełnie nie eksplorowanych jeszcze obszarów analiz naukowych. Tego faktu nie można bagatelizować również w kontekście możliwości poszerzenia perspektyw metodologicznych tych badań.

4. Wstępne wyniki. Makieta projektu

Wstępna makieta projektu, bez planowanych funkcjonalności (filtry, przeszukiwanie itp.) znajduje się pod adresem: <https://projekt-cenzura.pl>.

Aktualnie na wskazanej stronie internetowej można zapoznać się z fotodokumentacją 317 losowo wybranych fragmentów tekstów zawierających wyraz *cenzura*, 294 losowo wybranych fragmentów tekstów zawierających wyraz *cenzenski* oraz 240 losowo wybranych fragmentów tekstów zawierających wyraz *ocenzurować*. Zamieszczone na stronie internetowej fotografie-skany fragmentów tekstów zostały pozyskane automatycznie przy pomocy wyszukiwarki pełnotekstowej operującej na bazie kilku milionów zdigitalizowanych publikacji w języku polskim, które pochodzą z okresu XVIII-XXI w. Materiał ten nie został jeszcze poddany selekcji pod kątem jego



Fot. 6. Makieta Projektu: *Ku Leksykonowi cenzury*

przydatności w kontekście opisanych wyżej procedur opracowywania siatki haseł *Leksykonu cenzury*, jednak stanowi on ilustrację pierwszego etapu pracy polegającego na niezasochłonnym, automatycznym pozyskiwaniu materiału badawczego z reprezentatywnego korpusu tekstowego. Należy nadmienić, iż na portalu umieszczono również uszeregowany chronologicznie wybór kilkunastu fotoekscerptów zaczerpniętych ze źródeł prasowych dotyczących – bezpośrednio lub pośrednio – cenzury stosowanej w drugiej połowie XIX w. i na początku XX w. przez władze państwowe oraz kościelne na terenach państw zaborczych, a także na ziemiach niezamieszkanym przez Polaków. Wybór ten ma w zamierzeniu autorów stanowić ilustrację planowanego kształtu fotodokumentacyjnej ilustracji haseł *Leksykonu cenzury*.

W przyszłości planowane jest uzupełnienie portalu o informacje dotyczące założeń pracy nad *Leksykonem cenzury*, informacje o zespole badawczym, a także – co najważniejsze – rozbudowanie sekcji poświęconej wynikom projektu, tak aby zawrzeć na stronie wybrane hasła, które następnie zostaną włączone do leksykonu.

5. Dyskusja i podsumowanie

W zakończeniu powyższej prezentacji chcielibyśmy wskazać na te aspekty powyższej propozycji badawczej, które uznajemy za istotne z perspektywy cenzurologii. Na szczególną uwagę zasługują naszym zdaniem cztery kwestie.

Po pierwsze, opisane przedsięwzięcie wyróżnia się dążeniem do optymalizacji działań poprzez dokonanie wyboru niezasochłonnym metod operacyjnych i zadań badawczych związanych z pozyskiwaniem danych, które pozwolą na uzyskanie reprezentatywnego materiału tekstowego (uznajemy bowiem, że zaprezentowane metody są najbardziej optymalne dla osiągnięcia postawionego przez nas celu). W przeciwieństwie do innych projektów dokumentacyjnych wykorzystanie metod sztucznej inteligencji umożliwi szybkie zgromadzenie informacji, do których dotarcie jest niemalże niemożliwe podczas prowadzenia tradycyjnych kwerend bibliotecznych polegających na wyszukiwaniu materiałów źródłowych za pomocą takich metadanych, jak tytuł, autor czy hasła przedmiotowe, a następnie manualnym studiowaniu źródeł. Owo klasyczne podejście z natury rzeczy prowadzi bowiem do pomijania tych pozycji, których przy pomocy dostępnych narzędzi wyszukiwawczych nie można sklasyfikować jako materiałów do badań nad cenzurą, a które mogą zawierać istotne informacje dotyczące tej problematyki. Wydobycie tych informacji z „wnętrza” tekstów jest jednak możliwe przy wykorzystaniu wyszukiwarki pełnotekstowej obsługującej duży zasób dokumentów cyfrowych na podstawie uprzednio przygotowanego katalogu zapytań.

Kolejne *novum* planowanego opracowania polegać będzie na uwzględnieniu informacji frekwencyjnych dotyczących poszczególnych wyrazów i połączeń wyrazowych (ogólniej: pojęć związanych z cenzurą, które zostaną włączone do leksykonu

w postaci hasła). Dane odnoszące się do częstości występowania poszczególnych pojęć w tekstach umożliwią wskazanie, w jakich okresach najpowszechniej się do nich odwoływano, które to informacje mogą zostać następnie wykorzystane do formułowania hipotez związanych z przyczynami wzrostu „popularności” konkretnych określeń, a także ich zaniku.

Trzeci wyróżnik *Leksykonu cenzury* stanowić będzie zachowanie wierności cytowanej przewidywanego wolumenu danych, ponieważ każde hasło zilustrowane zostanie przy pomocy skanów-fotografii (fotodokumentacji) fragmentów tekstów poświęconych określonym pojęciom z zakresu cenzury. W ten sposób wyeliminowane zostanie ryzyko ingerencji badacza w tekst źródłowy.

W ramach przedsięwzięcia planowane jest także przeprowadzenie badań porównawczych na materiale co najmniej dwóch wersji tego samego hasła uwzględnionego w kilku wydaniach jednej encyklopedii w celu ustalenia, czy pod wpływem organów cenzury bądź też zniesienia ograniczeń cenzorskich jego treść została zmodyfikowana, np.: *Encyklopedia popularna PWN* z roku 1969 lub 1980 oraz 1991. Por. poniższy przykład (badanie przeprowadzono, stosując rozwiązania klasy aligner oraz MinHash¹⁹):

Na zakończenie przedstawionych rozważań pragniemy wskazać na przydatność *Leksykonu cenzury* w dalszych dociekaniach naukowych o charakterze

ZAWADZKI

ZAWADZKI ALEKSANDER, ur. 16 XII 1899 w Będzinie, zm. 7 VIII 1964 w Warszawie, działacz ruchu robotn., generał, górnik. Od 1922 w Związku Młodzieży Komunistycznej (ZMK), od 1923 w KPP. Kilkakrotnie więziony. Podczas II wojny świat. przebywał w ZSRR; współorganizator Związku Patriotów Pol. (1943), od 1944 członek Prezydium jego Zarządu Głównego; przewodn. Centralnego Biura Komunistów Pol. w ZSRR; 1944 zastępca dowódcy 1 Armii Pol. w ZSRR do spraw polit., w tymże roku szef Pol. Sztabu Partyzanckiego. W 1944–45 zastępca naczelnego dowódcy WP do spraw politycznych. Od 1944 członek Biura Politycznego KC PPR, od 1948 członek Biura Polit. KC PZPR, 1948 sekretarz KC PPR, 1948–49 KC PZPR. W 1945–48 wojewoda śląsko-dąbrowski, 1949–52 wicepremier, 1949–50 przewodn. CRZZ; od 1949 członek Rady Państwa, od 1952 jej przewodn.; od 1956 przewodn. Ogólnopolskiego Komitetu Frontu Jedności Narodu.



A. Zawadzki

Zawadzki ALEKSANDER, ur. 16 XII 1899, Dąbrowa Górnicza, zm. 7 VIII 1964, Warszawa, działacz komunist., generał; od 1923 w KPP, 1934–36 kierownik Centr. Wydziału Wojsk. KC; w okresie II wojny świat. w ZSRR, górnik w Zagłębiu Kuźnickim; od 1943 w pol. jednostkach wojsk. w ZSRR, III–VII 1944 zast. dcy Armii Pol. w ZSRR (generał), I–VII 1944 przewodn. Centr. Biura Komunistów Polski w ZSRR i V–VII szef Pol. Sztabu Partyzanckiego, czł. władz ZPP; 1944–45 zast. nacz. dcy WP ds. polit.-wychowawczych; 1944–64 czł. Biura Polit. PPR i PZPR oraz poseł do KRN, na Sejm Ustawodawczy i Sejm PRL; 1945–48 wojewoda śląsko-dąbrowski, I–VI 1949 i 1950–52 wicepremier, 1948–54 sekr. KC PZPR, 1949–50 przewodn. Centr. Rady Związków Zawodowych, 1952–64 przewodn. Rady Państwa oraz 1956 i 1958–64 przewodn. Ogólnopol. Kom. Frontu Jedności Narodu.

Fot. 7. Biogramy Aleksandra Zawadzkiego z Encyklopedii PWN (z lat 1969 oraz 2005)

¹⁹ W tej kwestii zob. A. Broder, *On the Resemblance and Containment of Documents*, „SEQUENCES '97: Proceedings of the Compression and Complexity of Sequences” 1997, s. 21–29.

szczegółowym. Bliższa analiza dotychczasowego dorobku polskiej cenzurologii wskazuje bowiem, że badania w tym zakresie prowadzone są przez reprezentantów różnych dyscyplin. Z pewnością należeć będą do nich historycy zajmujący się przede wszystkim rekonstrukcją procesów kontroli druków w wiekach XV-XX. Wiek XX stanowił pewien przełom w tym zakresie, ponieważ bardzo wyraźnie rozrzucono wówczas perspektywę badawczą, czego konsekwencją było zainteresowanie się cenzurą reprezentantów kolejnych dyscyplin, takich jak choćby literaturoznawstwo²⁰, księgoznawstwo²¹, naukoznawstwo²², a nawet językoznawstwo²³.

Nie można mieć jednak wątpliwości co do tego, że problematyką cenzuralną (głównie z uwagi na przedmiot badań) zainteresowani będą specjaliści z zakresu nauk o mediach. Organy cenzury w wielu systemach prawnych, zwłaszcza państw totalitarnych, zajmowały bardzo ważne miejsce, wpływając zarówno na politykę wewnętrzną, jak i międzynarodową, jaką prowadziły poszczególne państwa, co znajdowało odzwierciedlenie w przekazach medialnych. Z tego powodu opisana przez nas metoda tworzenia leksykonu cenzury z wykorzystaniem wybranych elementów sztucznej inteligencji z pewnością posłuży do intensyfikacji badań nad cenzurą również w obrębie nauk o mediach.

Bibliografia

- Bandtkie J. S., *Historia drukarni krakowskich*, Kraków 1815.
- Bandtkie J. S., *Historia drukarni w Królestwie Polskim i Wielkim Księstwie Litewskim*, Kraków 1826.
- Broder A., *On the Resemblance and Containment of Documents*, „SEQUENCES '97: Proceedings of the Compression and Complexity of Sequences” 1997, s. 21–29.
- Buchwald-Pelcowa P., *Cenzura w dawnej Polsce. Między prasą drukarską a stosem*, Warszawa 1997.
- Budrowska K., *Literatura i pisarze wobec cenzury PRL 1948–1958*, Białystok 2009.
- Budrowska K., Budnik M., Gardocki W., „*Biuletyn Informacyjno-Instrukcyjny*”. Wybór dokumentów z 1955 r. Białystok 2018.
- Budrowska K., Budnik M., Kościewicz K., *Dramat i teatr w dokumentach GUKPPiW*. Białystok 2017.
- Budrowska K., Kościewicz K., „*Dzieła*” Henryka Sienkiewicza w dokumentach Głównego Urzędu Kontroli Prasy, Publikacji i Widowisk (1948–1954), Białystok 2016.
- Cover T., Hart P., *Nearest neighbor pattern classification*, „IEEE Transactions on Information Theory” 1967, 13, s. 21–27.
- Ćiećwierz M., *Polityka prasowa 1944–1948*, Warszawa 1989.
- Doccano, *Open source annotation tool for machine learning practitioners* (brak daty opubl.), <https://github.com/doccano/doccano> (dostęp 17.01.2022).

²⁰ K. Budrowska, dz. cyt.

²¹ P. Nowak, *Cenzura wobec rynku książki. Wojewódzki Urząd Kontroli Prasy, Publikacji w Widowisk w Poznaniu w latach 1946–1955*, Poznań 2012.

²² Z. Romek, *Cenzura a nauka historyczna w Polsce 1944–1970*, Warszawa 2010.

²³ D. Dzienisiewicz, P. Nowak, *Leksykograficzny obraz pojęcia „cenzura” w słownikach oraz encyklopediach powszechnych wydanych w Polskiej Rzeczypospolitej Ludowej oraz w Związku Socjalistycznych Republik Radzieckich*, „Toruńskie Studia Bibliologiczne” 2022, z. 1, s. 9–40.

- Dzieniaiewicz D., Nowak P., *Leksykograficzny obraz pojęcia „cenzura” w słownikach oraz encyklopediach powszechnych wydanych w Polskiej Rzeczypospolitej Ludowej oraz w Związku Socjalistycznych Republik Radzieckich*, „Toruńskie Studia Bibliologiczne” 2022, z. 1, s. 9–40.
- Federacja Bibliotek Cyfrowych, *Federacja Polskich Bibliotek Cyfrowych* (brak daty opubl.), <https://fbc.pionier.net.pl/> (dostęp 17.01.2022).
- Gogol B., „*Fabryka fałszywych tekstów*”. Z działalności Wojewódzkiego Urzędu Kontroli Prasy, Publikacji i Widowisk w Gdańsku w latach 1945–1958, Warszawa 2012.
- Graliński F., *Against the Arrow of Time. Theory and Practice of Mining Massive Corpora of Polish Historical Texts for Linguistic and Historical Research*, Poznań 2019.
- Graliński F., Dzieniaiewicz D., Wierzchoń P., *U bram lingwistycznej szczęśliwości, czyli kulisy projektu Odkrywka: cyfrowe zasoby kultury jako źródło mas danych językowych*, „Studia Sociologica” 2017, z. 9, s. 51–62.
- Green J., Karolides N., *The Encyclopedia of Censorship (Facts on File Library of World History)*, New York 2005.
- Gutowski T., *Cenzura w Wolnym Mieście Krakowie 1832–1846*, Kraków 1914.
- Habielski R., *Życie społeczne i kulturalne emigracji*, Warszawa 1999.
- Habielski R., *Polityczna historia mediów w Polsce XX wieku*, Warszawa 2009.
- Habielski R., *Wolność czy odpowiedzialność? Prasa i polityka w II Rzeczypospolitej*, Warszawa 2013.
- Hoesick F., *Ze wspomnień o cenzurze rosyjskiej w Warszawie*, Warszawa 1929.
- Hurwitz L., *Historical Dictionary of Censorship in the United States*, Westport 1985.
- Joulin A., Grave E., Bojanowski P., Douze M., Jegou H., Mikolov T., *FastText.zip: Compressing text classification models*, 2016, <https://arxiv.org/pdf/1612.03651.pdf> (dostęp 17.01.2022).
- Kamińska-Chełmniak K., *Cenzura w Polsce 1944–1960. Organizacja – kadry – metody pracy*, Warszawa 2019.
- Kind-Kovács F., *Written Here, Published There: How Underground Literature Crossed the Iron Curtain*. Budapest 2014.
- Kmieciak Z., *Prasa warszawska w latach 1886–1904*, Wrocław 1989.
- Kondek S. A., *Papierowa rewolucja. Oficjalny obieg książki w Polsce w latach 1948–1955*, Warszawa 1999.
- Korczyńska-Derkacz M., *Państwowy Instytut Książki (1946–1949) i jego rola w rozwoju bibliologii, bibliotekarstwa i kultury książki w Polsce*, Wrocław 2011.
- Kozieł A., *Studium o polityce prasowej PZPR w latach 1948–1957*, Warszawa 1991.
- Kristanova E., *Książka na łamach katolickich czasopism społeczno-kulturalnych w Polsce w latach 1945–1953*, Łódź 2012.
- Kucharczyk G., *Cenzura pruska w Wielkopolsce w czasach zaborów 1815–1914*, Poznań 2001.
- Mikolov T., Chen K., Corrado G., Dean J., *Efficient Estimation of Word Representations in Vector Space*, „Proceedings of Workshop at ICLR” 2013, <https://arxiv.org/abs/1301.3781> (dostęp 17.01.2022).
- Mlekicka M., *Wydawcy książek w Warszawie w okresie zaborów*, Warszawa 1987.
- Nałęcz D., *Główny Urząd Kontroli Prasy 1945–1949*, Warszawa 1994.
- Nowak P., *Cenzura wobec rynku książki. Wojewódzki Urząd Kontroli Prasy, Publikacji i Widowisk w Poznaniu w latach 1946–1955*, Poznań 2012.
- OpenCV, *Home – OpenCV*, <https://opencv.org> (dostęp 17.01.2022).
- Pawlicki A., *Kompletna szarość. Cenzura w latach 1965–1972*, Warszawa 2001.
- Pietrzak M., *Reglamentacja wolności prasy w Polsce (1918–1939)*, Warszawa 1963.
- Rajch M., *Cenzura pruska w Wielkopolsce w latach 1948–1918*, Poznań 2004.
- Reisch A. A., *Hot Books in the Cold War. The CIA-Funded Secret Western Book Distribution Program Behind the Iron Curtain*, Budapest 2013.
- Rogers A., Kovaleva O., Rumshisky A., *A Primer in BERTology: What We Know About How BERT Works*, „Transactions of the Association for Computational Linguistics” 2020, 8, s. 842–866.
- Romek Z., *Cenzura a nauka historyczna w Polsce 1944–1970*, Warszawa 2010.

Szyndler B., *Dzieje cenzury w Polsce do 1918 roku*, Warszawa 1993.

Tazbir J., *Państwo bez stosów. Szkice z dziejów tolerancji w Polsce XVI i XVII wieku*, Warszawa 2009.

Wierzchoń P., *Fotodokumentacja, chronologizacja, emendacja. Teoria i praktyka weryfikacji materiału leksykalnego w badaniach lingwistycznych*, Poznań 2008.

Wierzchoń P., *Słownictwo lat 30. XX w. w obrazach i liczbach. Tylko 100 fotoprzykładów z Depozytorium leksykalnego języka polskiego i dużo wykresów z systemu Odkrywka*, Warszawa 2015.

Wojnowski J., *Wielka Encyklopedia PWN*, t. 27, Warszawa 2005.

Zajewski Z., *Wolność druku w powstaniu listopadowym 1830–1831*, Łódź 1963.

Biogramy autorów

Dr Daniel Dzienisiewicz, językoznawca, adiunkt na Wydziale Etnolingwistyki Uniwersytetu im. Adama Mickiewicza w Poznaniu. Autor i współautor kilkudziesięciu publikacji, poświęconych m.in. zagadnieniom z zakresu etnolingwistyki, pragmatyki językowej, słowotwórstwa języka rosyjskiego i polskiego i lingwochronologizacji. Aktualnie jego zainteresowania badawcze koncentrują się wokół języka wiadomości przesyłanych na kartach pocztowych oraz problematyki werbalizacji emocji w języku.

Dr hab. Piotr Nowak, profesor Uniwersytetu im. A. Mickiewicza w Poznaniu, kierownik Zakładu Infolingwistyki na Wydziale Etnolingwistyki. Autor stu kilkudziesięciu publikacji naukowych poświęconych zagadnieniom ruchu wydawniczego, historii i kultury książki oraz biblio- i webometrii. Jego zainteresowania badawcze dotyczą także problemów cenzury druków oraz perlustracji.

Prof. dr hab. Piotr Wierzchoń, językoznawca, UAM; profesor nauk humanistycznych; autor i współautor ponad dwustu publikacji naukowych. Najnowsze zainteresowania badawcze: Korpusy językowe: tworzenie i analiza dużych zbiorów danych tekstowych dla celów badawczych. Humanistyka cyfrowa i ekstrakcja wiedzy z cyfrowych źródeł danych: wykorzystanie narzędzi cyfrowych do analizy i interpretacji danych humanistycznych. Zastosowanie uczenia maszynowego w badaniach humanistycznych: wykorzystanie klasyfikatorów i innych technik uczenia maszynowego do analizy i interpretacji danych tekstowych.