

<https://jacek.kwasniewski.org.pl>

Jacek Kwaśniewski



linotypści w zecerni



„The New York Times”, 1942.

druk gazety

Ilość informacji
wyprodukowanych w historii Europy
miarą modernizacji
wersja 2

Modelo para armar / Model do składania

2025

Jacek Kwaśniewski

Ilość informacji wyprodukowanych w historii Europy miarą modernizacji. Wersja 2

Modelo para armar / Model do składania

Spis treści

Informacja i modernizacja: jak połączyć?	2
Zbiorcze wyniki szacunku.....	3
Jak się mają dane historyczne do współczesnych	4
Krótko o metodzie liczenia ilości informacji produkowanych w przeszłości.....	5
Korzystanie z modelu szacującego ilość informacji od VI do XIX wieku	7
Panel sterowania – jak wygląda	9
Techniczny opis modelu. Wstęp.....	11
Arkusz 1 „panel sterowania”	12
Arkusz 2 „manuskrypty”	17
Arkusz 3 „wartość jednego słowa”	19
Arkusz 4 „silnik”	23
Recenzja modelu przez ChataGPT 5.4 (kwiecień 2026).....	25

W części wprowadzającej (strony 4-7) są podane w nawiasach kwadratowych odniesienia do bardziej szczegółowych wyjaśnień w części technicznej

Model do samodzielnych szacunków jest dostępny na życzenie

Informacja i modernizacja: jak połączyć?

Ilość materialnie utrwalonej informacji może być traktowana jako syntetyczna zmienna proxy dla skali i intensywności historycznych procesów modernizacyjnych Europy. Ilość informacji może pełnić tę funkcję, bo wzrost gospodarki, administracji, edukacji, nauki, techniki, prawa, handlu, finansów, komunikacji i kultury niemal zawsze generuje więcej zapisów, dokumentów, instrukcji, ksiąg, korespondencji, druków, akt, map, gazet, formularzy, rejestrów i publikacji.

Ilość informacji jako zmienna proxy nie mierzy jakości modernizacji (rozumianej jako m.in. efektywność instytucji, poziom wolności, dobrobytu, racjonalność polityczna, sprawność państwa, jakość wiedzy). Mierzy materialny ślad informacyjny procesów, które z tymi zjawiskami są powiązane.

Ilość wyprodukowanej informacji została policzona w podstawowych jednostkach informatycznych, w kilobajtach, megabajtach i terabajtach. Podstawą obliczeń są prace dwóch historyków, Jana Luitena van Zandena i Eltjo Buringha, którzy oszacowali liczbę manuskryptów wyprodukowanych w Europie w okresie od VI do XV wieku oraz liczbę książek wydrukowanych w okresie od XV do XVIII wieku. Jest to niezwykle dokonanie naukowe, oparte na wieloletnich badaniach i wyrafinowanych narzędziach statystycznych.¹

Liczba manuskryptów (VI-XV w.) i książek drukowanych (XV-XIX w.) •

wiek	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX
liczba manuskryptów (sztuki)	13,5 tys	10,6 tys	43,7 tys	201,7 tys	135,6 tys	212 tys	768,7 tys	1,76 mln	2,75 mln	4,99 mln				
liczba książek drukowanych (sztuki)										12,6 mln	217,4 mln	531,9 mln	983,9 mln	4 919 mln
										Razem 17,6 mln				

- W oryginale liczby podane z dokładnością do jednej sztuki. Autorzy podali dokładne wyniki swojego modelu
- Manuskrypty. Dotyczy następujących obszarów: Europa Centralna, Wyspy brytyjskie, Francja, Belgia, Niderlandy, Niemcy, Szwajcaria, Austria, Włochy, Hiszpania i Portugalia, Europa Zachodnia. źródło: Eltjo Buringh, Jan Luiten van Zanden, *Charting the "Rise of the West"*
- Książki drukowane. dotyczy następujących krajów: Wielka Brytania, Irlandia, Francja, Belgia, Niderlandy, Niemcy, Szwajcaria, Włochy, Hiszpania, Szwecja, Polska, Rosja, inne (Austria, Węgry, Portugalia, Czechy, pozostała Skandynawia). źródło: Eltjo Buringh, Jan Luiten van Zanden, *Charting the "Rise of the West"*
- Wiek XIX. Szacunki własne

Proces historycznej modernizacji Europy to zmiany, jakie rozpoczęły się od X wieku, kiedy Europa zaczęła się podnosić po ciężkich czasach wieków poprzednich. Oznaczała przyspieszenie gospodarcze, naukowe, technologiczne, w sferze kultury a także zasadnicze zmiany polityczne. W ich wyniku Europa zdobyła kilka wieków później pozycję światowego hegemonu.

Procesy modernizacyjne we wszystkich wymienionych dziedzinach mają cechę wspólną – wzrost produkowanej informacji. Licząc ten wzrost prezentujemy najbardziej syntetyczny miernik postępu procesów modernizacyjnych. Znajdujemy bowiem ich wspólny mianownik.

Jak dokonać obliczeń wyprodukowanych w przeszłości informacji? W tym celu został zbudowany model matematyczny, który dokonuje obliczenia ilości produkowanych informacji, wiek po wieku. Zawiera dane opracowane przez van Zandena i Buringha oraz inne. Przelicza jedenaście zmiennych oraz informacje z analizy średniowiecznych manuskryptów oraz współczesnych książek. Trzeba zastrzec, że szacowanie ilości informacji z dalekiej przeszłości jest obarczone sporym ryzykiem błędu. W szczegółowym opisie modelu ryzyka te są bliżej opisane. Podstawową zaletą modelu jest klarowna prezentacja czynników, które mają wpływ na te szacunki oraz możliwość zmiany ich wartości, aby zobaczyć różne warianty badanego zjawiska.

Niniejszy tekst jest krótkim, wprowadzającym opisem procesu, który doprowadził do obliczenia ilości informacji wyprodukowanej w Europie (na nośnikach materialnych) w okresie od VI do XIX wieku. Dokładny opis jest zawarty w dalszej części, zatytułowanej „Techniczny opis modelu”.

Zbiorcze wyniki szacunku

Poniżej prezentuję efekt obliczeń. Wielkość informacji podana jest w terabajtach (TB). Od VI do XIX wieku Europa wyprodukowała 20 090 terabajty informacji. Dla bliższego zobrazowania tej wielkości, terabajt to zapis na karcie pamięci lub dysku 189 tysięcy egzemplarzy Trylogii Sienkiewicza w formacie txt (plaintext).

Okres wydobywania się z dna cywilizacyjnego kryzysu wieku VI do zapoczątkowania odbudowy gospodarki, technologii, nauki i kultury w wieku XI znajduje odzwierciedlenie w ilości produkowanej informacji. Produkcja informacji wzrosła w okresie VI-IX wiek czterdziestokrotnie. Warte przyspieszenie cywilizacyjne pełnego i późnego Średniowiecza, czyli wieki od XI do XV znajduje także odbicie w sferze informacji, której produkcja rosła przeszło dwa razy szybciej niż poprzednio, bo niemal stukrotnie. Ciekawym zjawiskiem są wieki XV i XVI. W połowie wieku XV pojawił się druk. Skokowy wzrost wszelkiego rodzaju dokumentacji, będący tego efektem, spowodował w okresie zaledwie stu lat (XV-XVI wiek) dziesięciokrotny wzrost wytworzonej informacji. Cały okres wczesnonowożytny, czyli wieki od XV do XIX, kiedy Europa zdobyła pozycję światowego hegemonu gospodarczego, technologicznego, naukowego i objęła kontrolą połowę obszaru świata jest w pełni odzwierciedlony wskaźnikiem wzrostu produkcji informacji. Wzrosła ona sto pięćdziesiąt razy. Przedstawione liczby potwierdzają tezę, że informacja może być uznana za kompleksową miarę procesu historycznej modernizacji Europy. Jest ilościowym wykładnikiem tempa przemian cywilizacyjnych i pozwala syntetycznie ująć dynamikę tego procesu.

Skrócone podsumowanie wyników

oraz wskaźniki wzrostu produkcji informacji w okresie od VI do XIX wieku

	wiek VII, VIII - Renesans karoliński			od 1450 technika druku			
wiek	VI	VIII	XI	XV	XVI	XIX	VI - XIX
ilość informacji w TB	0,024	0,077	0,96	89	893	13 532	20 090
wskaźniki wzrostu	100		wzrost 40 krotny				
			100	wzrost 93 krotny			
				100		wzrost 152 krotny	
				100	wzrost 10 krotny		
					100	wzrost 15 krotny	

źródło: obliczenia na podstawie modelu. Wiek XIX szacunek własny

Jak się mają dane historyczne do współczesnych

Czy dwadzieścia tysięcy terabajtów to dużo? Ta sama metodologia liczenia danych historycznych i ilości informacji generowanych współcześnie pozwala porównać przeszłość z teraźniejszością.

Okres ostatnich stu dwudziestu pięciu lat pod względem przyspieszenia cywilizacyjnego, odzwierciedlanego ilością generowanej informacji, jest bardzo niejednorodny. Jeśli za jednostkę informacji przyjąć jej ilość wyprodukowaną w okresie od VI do XIX wieku (20 090 TB), to sto lat później, u progu XXI wieku świat produkował rocznie 100-150 razy więcej informacji, bo około dwa-trzy eksabajty (eksabajt to milion terabajtów). Pierwsza dekada XXI wieku to początek eksplozji informacyjnej. W roku 2002 było pięć eksabajtów a rok 2010 to już 2 zettabajty (zettabajt to miliard terabajtów), czyli sto tysięcy razy więcej niż wyprodukowano w okresie od VI do XIX wieku.

W roku 2024 świat wyprodukował 150 zettabajtów informacji. 30% tej produkcji to komunikacja między maszynami. 70% światowych danych pochodzi z aktywności ludzi.² Jeśli uwzględnimy tylko tę aktywność, owe 70%, czyli 105 zettabajtów, okazuje się, że dziennie produkujemy czternaście tysięcy razy więcej informacji niż na przestrzeni czternastu wieków w Europie. Jeśli nawet zwiększymy nasz szacunek danych historycznych dwukrotnie, do czterdziestu terabajtów, na przykład podnosząc wartość jednego wyrazu z 0,0123 KB do 0,0245 KB (scenariusz moim zdaniem zbyt optymistyczny), okazuje się, że dziennie produkujemy ponad siedem tysięcy razy więcej informacji niż od VI do XIX wieku.

Trudno to pojąć, ale można zaakceptować. Jeśli jednak próbujemy policzyć, ile czasu obecnie zajmuje wygenerowanie tej ilości informacji, jaka została wyprodukowana przez czternaście wieków, wtedy pojawia się zadziwienie, czy i jak było to i jest możliwe. Okazuje się bowiem, że tylko aktywność informacyjna ludzi (pomijamy maszyny) potrzebowała w roku 2024 zaledwie sześciu sekund na wyprodukowanie 20 090 terabajtów informacji, czyli

produkcji od VI do XIX wieku. Trudno w to uwierzyć, ale nawet dwukrotne powiększenie szacunku za okres od VI do XIX wieku, czyli do czterdziestu tysięcy terabajtów, niewiele zmienia. Obywatele świata z roku 2024 potrzebowałiby wtedy dwunastu sekund. Taka jest skala informatycznej eksplozji drugiej dekady XXI wieku.

Krótko o metodzie liczenia ilości informacji produkowanych w przeszłości

Wróćmy do historii. Dane van Zandena i Buringha są cenne, ale musi być też uwzględniona, a była poza ich zainteresowaniem, wszelka dokumentacja związana z funkcjonowaniem państwa i sfery publicznej (m.in. polityka, administracja, sądownictwo, kościoły), gospodarki (m.in. handel, przemysł, bankowość), ekspansji geograficznej, techniki i sfery kultury. Informacje te obejmują także wszelką korespondencję a po wynalezieniu druku, wprawdzie druki okazjonalne w potem gazety i czasopisma [patrz Arkusz 1, pkt. 10, 22]

Należy więc ustalić metodę pomiaru tych innych informacji oraz metodę przeliczenia materialnych nośników informacji historycznej na takie jednostki jak kilobajty, megabajty i terabajty. Przeliczenie musi być przy tym zgodne ze współczesną metodą pomiaru, aby szacunki dotyczące historii i współczesności były ze sobą porównywalne.

Zacznijmy od przeliczenia nośników historycznych na jednostki informatyczne. Przyjęcie metody dającej wyniki porównywalne z pomiarami współczesnymi wyklucza liczenie zawartości informacyjnej manuskryptów i starych druków wielkością plików tych dokumentów, jakie są dostępne w bibliotekach. Są to bowiem obrazy, fotografie stron, zamieniane na format pdf. Z reguły to bardzo duże pliki, dochodzące niekiedy do pół gigabajta na jeden manuskrypt. Znaczną część takich plików zajmuje nie informacja tekstowa, ale tło, tekstura pergaminu, papieru i artefakty skanowania (plamy, postrzępione brzegi, otarcia, rysy, uwagi nanoszone ręcznie itd.), czyli tak zwana redundancja wizualna, którą należy pominąć.

Jest kilka metod jej usuwania. Zastosowana tu metoda polegała na skoncentrowaniu uwagi na liczbie wyrazów w manuskryptach i starych drukach i przypisaniu każdemu słowu wartości informacyjnej. Należy w tym celu oszacować liczbę wyrazów w manuskryptach i starych drukach i obliczyć wartość informacyjną jednego słowa.

Przy szacowaniu liczby słów manuskryptów należało uwzględnić rosnącą z wiekami gęstość zapisu na stronie. Dlatego podzielono Średniowiecze na trzy okresy i z każdego przeanalizowano kilkanaście / kilkadziesiąt manuskryptów.³ Pobrano losowo po kilka stron z każdego i policzono średnią liczbę wyrazów. Teza o rosnącej gęstości zapisu została potwierdzona. Gęstość zapisu wzrosła z 219 słów na stronę w okresie od VI do IX wieku do 597 słów na stronę w okresie XIV-XV wieku. Kilka prostych działań arytmetycznych pozwalało na obliczenie średniej liczby wyrazów w manuskrypcie w próbie analizowanej z każdego okresu.⁴ Pomnożenie przez wartość informacyjną jednego słowa i liczbę manuskryptów wyprodukowaną w danym wieku daje ilość informacji w terabajtach.

Podobnie szacowano liczbę słów w starych drukach z okresu od XV do XIX wieku. Przegląd kilkudziesięciu pozycji pozwolił oszacować średnią liczbę stron i gęstość zapisu na

stronie. Specyfika rewolucji, jaką przyniósł wynalazek druku spowodowała, że dla wieków XVIII i XIX druki podzielono na grubsze woluminy oraz krótkie broszury.

Osobną kwestią było oszacowanie informacji innych, niż zawarte w manuskryptach i drukowanych książkach [patrz Arkusz 1, pkt. 10, 13, 14]. Przyjęto założenie, by Informacje Pozostałe ustalać dla każdego wieku w proporcji do informacji obliczonej dla manuskryptów a potem książek drukowanych. We wczesnym Średniowieczu (VI-IX wiek) ustalono, że wielkość Informacji Pozostałych będzie taka sama jak zawarta w manuskryptach tego wieku. Przebudzenie cywilizacyjne Europy od wieku X pozwoliło uznać, że Informacje Pozostałe rosły szybciej od tych z manuskryptów. Przyjęto mnożnik pięcioprocentowy. Od wieku XVI do XIX miał miejsce jeszcze szybszy rozwój procesów modernizacyjnych. Uznano, że Informacje Pozostałe rosną jeszcze szybciej i przyjęto mnożnik dziesięcioprocentowy.

Wartość jednego słowa [patrz Arkusz 3, pkt. 1-7]. Do jego obliczenia zastosowano ogólną metodologię firmy International Data Corporation (IDC) opisaną w jej raporcie *Data Age 2025: The Digitalization of the World*, z roku 2018.⁵ Do wyliczenia globalnej ilości informacji bierze się tam pod uwagę dane w takiej postaci, w jakiej są wytwarzane. Jeśli więc dane są kompresowane w typowym użyciu (np. wideo skompresowane kodekiem, tekst zapisany w formacie pliku, itp.), to raport uwzględnia je ze stosowaną kompresją – czyli według rozmiaru pliku/strumienia wynikowego. Nie próbuje się sprowadzać wszystkich danych do wspólnego mianownika (np. przeliczać na surową, niezakodowaną informację); przeciwnie – przyjęto obecne formaty i praktyki przechowywania [Arkusz 3: pkt. 5].

Ta metodologia uzasadnia obliczenie wartości informacyjnej jednego słowa z zawartości ebooków pdf, wyprodukowanych przez wydawnictwa. Takie ebooki są robione profesjonalnie i niemal zawsze są mniejsze pod względem wielkości plików niż pdf-y wykonane po amatorsku [patrz Arkusz 3: pkt. 6-7].

Do naszego celu wybrano kilkanaście książek naukowych z różnych dziedzin, czarno-białych, z jak najmniejszą liczbą infografik lub bez nich. Usunięto strony tytułowe, puste, prawie puste i te z infografikami. Pozostały tylko strony wypełnione w całości drukiem. Zapisano wielkość tak okrojonej książki w megabajtach (MB). Kolejnym krokiem był losowy wybór kilku stron każdej książki i obliczenie średniej liczby słów na stronę. Proste działanie arytmetyczne (wielkość plików razem / suma wyrazów razem) daje wartość informacyjną jednego słowa. Według przeprowadzonych wyliczeń wynosi ona 0,0123 KB. Wartość ta, policzona według współczesnej metodologii, może być stosowana do treści zapisanych w przeszłości i jest kompatybilna metodologicznie ze współczesną Globalną Sferą Danych.

Wartość tę należy jednak traktować ostrożnie z dwóch powodów. Po pierwsze, w poszczególnych badanych ebookach wartość jednego słowa była zróżnicowana w stosunku jeden do trzech. Po drugie, niektóre współczesne ebooki traktują słowa jako element nierozzerwalnie sklejony z całym „ekosystemem przekazu” (m.in. poprzez silne wizualne nasycenie treści; patrz szersze wyjaśnienie w opisie arkusza ‘wartość jednego słowa’, pkt. 8-11), w związku z czym wartość jednego słowa uwzględniająca ten ekosystem może wzrosnąć kilkudziesięciokrotnie. W osobnej, zaprezentowanej w modelu, próbie takich ebooków wynosiła ona 0,48 KB [patrz Arkusz 3: pkt. 8-10].

Korzystanie z modelu szacującego ilość informacji od VI do XIX wieku

Model jest łatwy w użyciu. Pozwala też zainteresowanym na dowolnie szczegółowe badanie jego konstrukcji. Jest otwarty, nic nie jest w nim zablokowane. Sugeruję, aby przed pierwszym użyciem zrobić jego kopię.

Model został zbudowany w programie MS Excel. Składa się z czterech arkuszy, ale do korzystania wystarczy arkusz o nazwie ‘panel sterowania’. Pokazane są tu wszystkie zmienne (ich wartości można zmieniać) oraz tabela wyników, która generuje natychmiast efekty tych zmian.

Zmiennych jest jednaście [patrz Arkusz 1: pkt. 9]. Ich opis znajduje się przy lewej krawędzi arkusza. Są zaznaczone jako komórki z gradientowym tłem biało niebieskim. Ich wartości można zmieniać w lewej pomarańczowej ramce i w prawej – zielonej.

W pomarańczowej ramce do komórek ze zmiennymi wstawione są wartości początkowe, wyliczone przeze mnie albo przyjęte jako prawdopodobne. Te liczby użytkownik może zmieniać i następuje natychmiastowe przeliczenie ilości informacji dla całego okresu od VI do XIX wieku. Użytkownik może też w każdej chwili powrócić do ustawień początkowych naciskając przycisk „powrót do ustawień początkowych” [patrz Arkusz 1: pkt. 4].

W zielonej ramce wartości zmiennych można zmieniać procentowo. Każdą zmienną można powiększać lub pomniejszać wpisując wybrany procent (dodatni lub ujemny) do odpowiedniej komórki z tłem biało niebieskim. Po wstawieniu np. zmiany procentowej 40%, ukaże się w komórce po prawej stronie nowa wartość zmiennej i nastąpi natychmiastowe przeliczenie ilości informacji dla całego okresu od VI do XIX wieku. Użytkownik może też w każdej chwili wyzerować zmiany naciskając przycisk „zerowanie zmian procentowych” [patrz Arkusz 1: pkt. 5].

Użytkownik może część zmiennych zmieniać w panelu pomarańczowym (zmiany ręczne) a część w panelu zielonym (zmiany procentowe). Należy pamiętać, że zmiana dokonana w panelu zmian ręcznych ma priorytet. Blokuję możliwość dokonywania zmian w panelu zmian procentowych lub unieważnia je, jeśli jest dokonana po wprowadzeniu zmian procentowych.

¹ Prace te były publikowane w kilku wydawnictwach. Podaję także informacje o dwóch recenzjach książki Eltjo Buringha. Buringh E., Zanden J.L., *Charting the “Rise of the West”: Manuscripts and Printed Books in Europe, A Long-Term Perspective from the Sixth through Eighteenth Centuries*, w: *The Journal of Economic History* (Vol. 69, No. 2, June 2009; Zanden J.L. van, Buringh E., *Book Production as a Mirror of the Emerging Medieval Knowledge Economy, 500–1500*, w: Zanden J.L. van, *The Long Road to Industrial Revolution: The European Economy in a Global Perspective, 1000-1800*, Brill 2009; Buringh E., *Medieval Manuscript Production in the Latin West Explorations with a Global Database*, Brill 2011; Biddlecombe S., Mac Laughlin J., *Review of the book of Eltjo Buringh, Medieval manuscript production in the Latin west: explorations with a global database* (Leiden: Brill, 2010), w: *The Economic History Review*, 65, 1, str. 377-378 (2012); Epstein S.A., *Review of the book of Eltjo Buringh, Medieval manuscript production in the Latin west: explorations with a global database* (Leiden: Brill, 2010), w: *Speculum*, Vol. 87, No. 1, str. 188-190 (January 2012)

² Źródła: <https://explodingtopics.com/blog/data-generated-per-day>
<https://www.statista.com/statistics/871513/worldwide-data-created>

³ Skorzystałem z trzech bibliotek. 1/ <https://polona.pl> biblioteka cyfrowa, która udostępnia zdigitalizowane zbiory pochodzące z Biblioteki Narodowej i instytucji z nią współpracujących; 2/ <https://digital.bodleian.ox.ac.uk> Bodleian Libraries. Biblioteki Bodlejańskie to sieć bibliotek wspierająca Uniwersytet Oksfordzki 3/ <https://digi.vatlib.it> biblioteka cyfrowa Biblioteki Watykańskiej. W modelu, w arkuszu ‘manuskrypty’ są cyfrowe identyfikatory manuskryptów, które pozwalają obejrzeć analizowane rękopisy.

⁴ Obliczenie liczby słów w średniowiecznych rękopisach jest zadaniem trudnym. Skorzystałem z pomocy <https://www.modelsofauthority.ac.uk> Models of Authority jest szkockim projektem badawczym do analizy OCR starych manuskryptów oraz <https://www.transkribus.org> Jest to portal oferujący OCR dokumentów z przeszłości, także rękopisów. Więcej na ten temat w szczegółowym opisie modelu.

⁵ O firmie IDC: <https://www.idc.com> Raport dostępny: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

Panel sterowania

Informacje wsadowe z prac J.L. Van Zandena i E. Buringha (tych liczb nie zmieniać)

wiek	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX
liczba manuskryptów wieki VI - XV, liczba egz.	13 552	10 639	43 702	201 742	135 637	212 030	768 721	1 761 951	2 746 951	4 999 161				
liczba książek drukowanych, wieki XV - XVIII, liczba egz.										12 589 000	217 444 000	531 941 000	983 874 000	4 919 370 000

panel zmian ręcznych

Zmienne niezależne: manuskrypty, książki i broszury drukowane, Informacje Pozostałe, wiek XIX

Manuskrypty	VI-IX wiek	X-XIII wiek	XI-XV wiek	XVI-XIX wiek
średnia liczba słów w 1 manuskrypcie	71 738	174 780	279 023	
wartość 1 słowa w KB	0,0123			
Wiek VI, VII, VIII, IX Informacje Pozostałe jako procent informacji zawartych w manuskryptach	100%			
Wiek X, XI, XII, XIII, XIV, XV Informacje Pozostałe procentowy wzrost w stosunku do wieku poprzedniego (% składany)		5%		błąd

Książki drukowane	XV wiek	XVI wiek	XVII wiek	XVIII wiek	XIX wiek
liczba stron w drukowanej książce	300				
liczba słów na jednej stronie drukowanej książki	450				
liczba stron w drukowanej broszurze				80	
liczba słów na jednej stronie drukowanej broszury				200	
druki - procent broszur				33%	50%
druki - procent książek				67%	50%
wartość 1 słowa w KB	0,0123				
Wiek XV - Informacje Pozostałe jako procent informacji zawartych w książkach drukowanych (taki sam, jak w odniesieniu do manuskryptów w wieku XV)	134%				
Wiek XVI - XVIII, Informacje Pozostałe - procentowy wzrost w stosunku do wieku poprzedniego (procent składany)			10%		

Wiek XIX	
Liczba książek w wieku XIX jako liczba książek w wieku XVIII pomnożona przez mnożnik =	5

panel zmian procentowych

analiza wrażliwości na zmiany wartości zmiennych niezależnych

wartości oryginalne	zmiana procentowa (+) lub (-)	nowa wartość
0,0123	0%	0,0123
100%	0%	100%
5%	0%	5%
300	0%	300
450	0%	450
80	0%	80
200	0%	200
33%	0%	33%
67%	0%	67%
10%	0%	10%
5	0%	5

wartości oryginalne	zmiana procentowa (+) lub (-)	nowa wartość
50%	0%	50%
		50%

broszury XVIII wiek broszury XIX wiek

wieki VI-XIX ilość informacji razem, w TB
20 090 TB

Arkusz 'zmienne i wyniki'

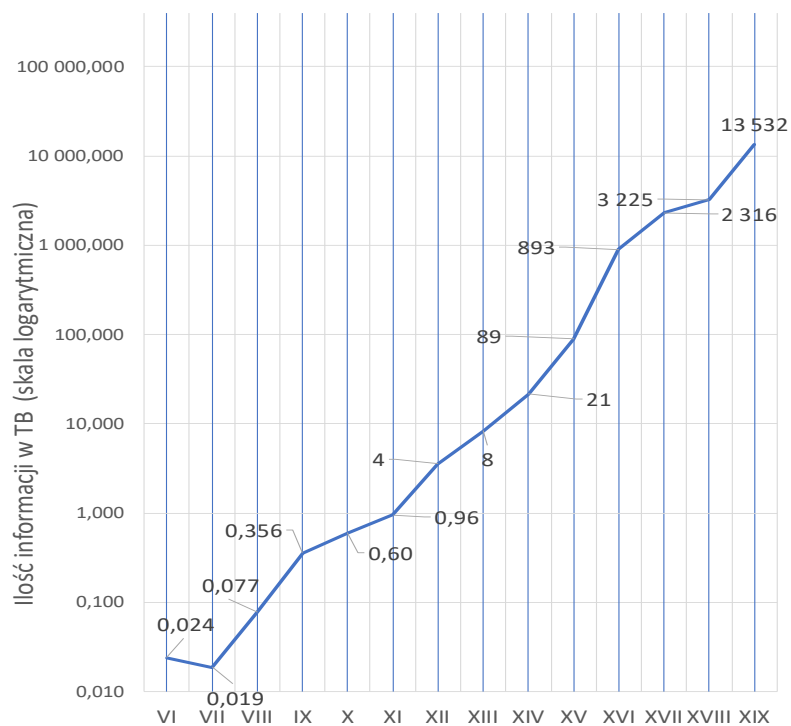
powrót do ustawień początkowych zerowanie zmian procentowych

w tak zaznaczone pola można wstawiać liczby

	WYNIKI		WYNIKI		WYNIKI		WYNIKI		WYNIKI						
wiek	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	
manuskrypty - ilość informacji razem w TB	0,012	0,009	0,039	0,178	0,292	0,46	1,65	3,79	9,43	17,16					
Informacje Pozostałe - ilość informacji w stosunku do manuskryptów w TB	0,012	0,009	0,039	0,178	0,306	0,50	1,91	4,60	12,03	22,99					
książki drukowane - ilość informacji razem w TB										20,90	361	883	1 158	4 568	
Informacje Pozostałe - ilość informacji w stosunku do książek drukowanych w TB										28,01	532	1 432	2 066	8 963	
łącznie ilość informacji w TB w kolejnych wiekach	0,024	0,019	0,077	0,356	0,598	0,96	3,57	8,39	21,46	89,07	893	2 316	3 225	13 532	20 090

wieki VI-XIX
ilość
informacji
razem, w TB

Ilość informacji w terabajtach (TB) wyprodukowana w Europie zachodniej w wiekach od VI do XIX



Techniczny opis modelu. Wstęp

Model wykonany w MS Excel liczy ilość informacji wyprodukowanych w Europie w okresie od VI do XIX wieku.

Podstawowa idea polega na wykorzystaniu pracy historyków holenderskich, którzy oszacowali liczbę manuskryptów i książek drukowanych, wyprodukowanych wiek po wieku w okresie od VI do XVIII wieku. Na tej podstawie i po analizie próbek manuskryptów i książek drukowanych szacowana jest liczba słów przypadająca na jeden manuskrypt a następnie na jedną książkę drukowaną. Osobno liczona jest wartość jednego słowa.

Wykorzystano metodę stosowaną we współczesnych szacowaniach Globalnego Ekosystemu Danych. Tę metodologię przeniesiono do obliczenia historycznej przestrzeni informacyjnej.

Wynik przemnożenia wyprodukowanych w poszczególnych wiekach woluminów, liczby zawartych w nich słów oraz wartości informacyjnej jednego słowa zwraca ilość informacji wyprodukowanej w poszczególnych wiekach.

Zaletą modelu jest fakt, że identyfikuje on policzalne składniki, które determinowały ilość produkowanej informacji, przedstawia propozycję ich oszacowania, stosuje metodologię, dzięki której można porównać ilości informacji produkowanej współcześnie oraz w przeszłości. Jest modelem otwartym, wszystkie formuły są dostępne do kontroli przez użytkownika. Posiada panel, na którym użytkownik może zmieniać wartości wszystkich zmiennych niezależnych, aby zobaczyć, jak ich zmiany wpływają na wielkość wyprodukowanej informacji w okresie od VI do XIX wieku. Jest propozycją, która otwiera dyskusję o liczeniu ilości informacji historycznej.

Model jest dostępny w dwóch wersjach: z automatycznym przywracaniem ustawień początkowych i automatycznym zerowaniem ich zmian procentowych (rozszerzenie pliku .xslm) oraz z ręcznym przywracaniem ustawień początkowych i ręcznym zerowaniem ich zmian procentowych (rozszerzenie pliku .xls)

Model można otrzymać wysyłając prośbę na adres jacek9520@gmail.com

Model składa się z czterech arkuszy:

1. „panel sterowania”
2. „manuskrypty”
3. „wartość jednego słowa”
4. „silnik”

Arkusz „panel sterowania” służy do obsługi modelu i tworzenia własnych scenariuszy

Arkusz „manuskrypty” szacuje ilość informacji wytworzonej w Europie średniowiecznej

Arkusz „wartość jednego słowa” wylicza wartość informacyjną wyrazu

Arkusz „silnik” zbiera wszystkie dane i wartości zmiennych i przelicza je na sumę informacji

Arkusz 1 „panel sterowania”

1. Arkusz „panel sterowania” jest stroną czołową modelu i jego podsumowaniem. Umożliwia wprowadzanie wartości zmiennych niezależnych (ZN) - sugerowanych przeze mnie albo własnych użytkownika. Każda wybrana kombinacja tych wartości przelicza natychmiast ich łączny efekt, prezentując tabelę z odpowiadającą tym wartościom ilością informacji wyprodukowanych w okresie od VI do XIX wieku w podziale na kolejne wieki. W nomenklaturze Excela panel sterowania to *dashboard*, czyli deska rozdzielcza.
2. Arkusz prezentuje podstawowe dane wsadowe, zmienne niezależne (ZN), których wartości można ustalać wedle własnych preferencji oraz wygenerowaną na ich podstawie tabelę wynikową, pokazującą w terabajtach (TB) ilość informacji wyprodukowanych w wiekach od VI do XIX oraz ich łączną sumę.
3. Podstawowe dane wsadowe pokazują ilość wyprodukowanych od VI do XV wieku manuskryptów i wyprodukowanych od XV do XVIII wieku książek drukowanych. Liczby te zostały obliczone przez Jana Luitena van Zandena i Eltjo Buringha i są bazą modelu.
4. W Arkuszu „panel sterowania” zmienne niezależne (ZN) opisane są przy lewej krawędzi arkusza. Są zaznaczone jako komórki z gradientowym tłem biało niebieskim. Ich wartości można zmieniać dwójako. W lewej pomarańczowej ramce (na jej górze napis „panel zmian ręcznych”) do komórek z gradientowym tłem biało niebieskim wstawione są wartości początkowe, czyli albo wyliczone przeze mnie (wartość informacyjna jednego słowa) albo przyjęte przeze mnie jako prawdopodobne. Te liczby użytkownik modelu może zmieniać i następuje wtedy natychmiastowe przeliczenie ilości informacji wyprodukowanych w poszczególnych wiekach i ich łącznej ilości dla całego okresu od VI do XIX wieku. Ta ostatnia liczba pojawia się też w pomarańczowym polu na prawo od ramki.

Jeżeli użytkownik ma otwartą wersję „automatyczną” modelu (rozszerzenie pliku .xlsm), może w każdej chwili powrócić do ustawień początkowych naciskając przycisk „powrót do ustawień początkowych”. Jeżeli ma otwartą wersję „ręczną” (rozszerzenie pliku .xls), aby wrócić do ustawień początkowych musi wpisać te ustawienia ręcznie do odpowiednich biało niebieskich pól (są one podane szarą, małą czcionką w kolumnie I i w komórce L19, trzeba przepisać, jeśli zostały zapomniane)
5. Wartości zmiennych niezależnych (ZN) można też zmieniać procentowo. Służy do tego „panel zmian procentowych” w zielonej ramce. Każdą zmienną niezależną można powiększać lub pomniejszać wpisując wybrany procent (dodatni lub ujemny) do odpowiedniej komórki z gradientowym tłem biało niebieskim. Po wstawieniu np. zmiany procentowej 40%, ukaże się w komórce po prawej stronie nowa wartość zmiennej niezależnej i nastąpi natychmiastowe przeliczenie ilości informacji wyprodukowanych w poszczególnych wiekach i ich łącznej ilości dla całego okresu od VI do XIX wieku. Ta ostatnia liczba pojawia się też w pomarańczowym polu na prawo od ramki.

Jeżeli użytkownik ma otwartą wersję „automatyczną” modelu (rozszerzenie pliku .xslm), może w każdej chwili wyzerować zmiany naciskając przycisk „zerowanie zmian procentowych”. Jeżeli ma otwartą wersję „ręczną” (rozszerzenie pliku .xls), aby wrócić do ustawień początkowych musi wpisać liczbę 0% ręcznie do białej niebieskich pól w kolumnie J i do komórki M19.

6. Użytkownik może część zmiennych niezależnych (**ZN**) zmieniać w panelu pomarańczowym (zmiany ręczne) a część w panelu zielonym (zmiany procentowe).
7. Należy pamiętać, że zmiana dokonana w panelu zmian ręcznych ma priorytet. Blokują możliwość dokonywania zmian w panelu zmian procentowych lub unieważnia je, jeśli jest dokonana po wprowadzeniu zmian procentowych.
8. Zmienne niezależne (**ZN**) determinują szacunek ilości informacji wyprodukowanych w Europie w okresie od VI do XIX wieku
9. Jest jedenaście zmiennych niezależnych (**ZN**):
 1. Wartość informacyjna jednego słowa (patrz pkt. 12)
 2. Wieki VI-IX. Informacje Pozostałe jako % ilości informacji zawartej w manuskryptach (patrz pkt. 10, 13)
 3. Wieki X-XV. Informacje Pozostałe – mnożnik wzrostu (patrz pkt. 10, 14)
 4. Liczba stron w książce drukowanej (patrz pkt. 16)
 5. Liczba słów na jednej stronie drukowanej książki (patrz pkt. 17)
 6. Liczba stron broszury (patrz pkt. 20)
 7. Liczba słów na stronie broszury (patrz pkt. 20)
 8. Procentowy udział broszur w ogólnej liczbie książek w wieku XVIII (patrz pkt. 19, 20)
 9. Procentowy udział broszur w ogólnej liczbie książek w wieku XIX (patrz pkt. 19, 20)
 10. Informacje Pozostałe - mnożnik wzrostu XVI-XIX wieku (patrz pkt. 21, 22)
 11. Mnożnik liczby książek w wieku XIX w stosunku do XVIII (patrz pkt. 23)

10. Czym są Informacje Pozostałe

Istotnym parametrem wpływającym na ilość informacji wyprodukowanej w poszczególnych wiekach są informacje nie uwzględnione w pracy van Zandena i Buringha. Jest to wszelka dokumentacja związana z funkcjonowaniem państwa i sfery publicznej (m.in. polityka, administracja, sądownictwo, religia), gospodarki (m.in. handel, przemysł, bankowość), ekspansji geograficznej, techniki i sfery kultury. Informacje te obejmują także wszelką korespondencję a po wynalezieniu druku, wprawdzie druki okazjonalne (np. w Anglii pamphlets i newsbooks, we Francji canards, w krajach niemieckich flugblätter) a potem gazety i czasopisma. Zdecydowałem, by wielkość tej Informacji Pozostałej ustalić dla każdego wieku w proporcji do wielkości informacji obliczonej dla manuskryptów a potem książek drukowanych.

Przyjąłem, że w okresie od VI do IX wieku Informacje Pozostałe stanowią 100% informacji zawartej w manuskryptach. Okresy X-XV wiek i XVI-XIX to okresy rosnącego przyspieszenia cywilizacyjnego. Dlatego w obu tych okresach przyjmuję,

że produkcja Informacji Pozostałych rosła szybciej, niż produkcja informacji z manuskryptów a potem z książek. Dla okresu X – XV wiek przyjmuję mnożnik 5-procentowy. Dla wieków XVI - XIX przyjmuję mnożnik 10-procentowy.

11. Książki i broszury

W wiekach XVIII i XIX lawinowy wzrost książek drukowanych, które uwzględnili w swoich szacunkach van Zanden i Buringh, przestał być kategorią jednorodną. Obaj autorzy przyjmują definicję książki za OECD. Biorą pod uwagę tytuły, które mogą być książkami (wolumen o ponad 49 stronach) lub krótszymi broszurami. Przyjąłem to rozróżnienie i w modelu dla wieku XVIII i XIX podzieliłem liczbę książek na te dwie kategorie.

12. Wartość informacyjna jednego słowa (ZN) w kilobajtach (KB). Jest to zmienna niezależna, której wartość można zmieniać. Wartość ta jest wyliczona w arkuszu „wartość jednego słowa” na poziomie 0,0123 KB, ale nie jest stamtąd przeniesiona hiperlinkiem do arkusza „panel sterowania”. Pozostawiam czytelnikowi swobodę, czy zaakceptuje mój szacunek, czy wpisze inną wartość. Zalecam jednak umiarkowanie przy zmianach, co wyjaśniam dalej, w opisie arkusza „wartość jednego słowa”.

13. Epoka manuskryptów. Wieki od VI do IX wieku. To ciężkie czasy dla Europy. Nie było wzrostu. Były za to najazdy barbarzyńców. Aktywność intelektualna była niewielka. Jest bardzo prawdopodobne, że z produkcją innej dokumentacji było podobnie. Dlatego przyjmuję (niektórzy uważają, że optymistycznie), że w wieku, VI, VII, VIII i IX Informacje Pozostałe stanowią 100% informacji zawartej w manuskryptach. Jest to zmienna niezależna (ZN). Jej nazwa w panelu sterowania brzmi „Wieki VI, VII, VIII, IX Informacje Pozostałe jako procent informacji zawartych w manuskryptach”. Jej wartość początkowa to 100%, ale można ją zmieniać (i wpisać np. 95%, 110%, 117% lub inne).

14. Epoka manuskryptów. Wieki od X do XV . Od X wieku Europa zaczęła się podnosić po ciężkich czasach wieków poprzednich. Barbarzyńcy zostali pokonani, gospodarka notowała wzrost, miasta zaczęły się dźwigać z upadku, rozwijało się szkolnictwo, zaczęły powstawać uniwersytety, produkcja manuskryptów między wiekiem VI i XI wzrosła szesnastokrotnie, rosły obroty handlowe, także te na długich dystansach. W wieku XIV rozpoczął się, trwający do dzisiaj, intensywny wzrost gospodarczy Europy, czyli wzrost PKB na jednego mieszkańca. Rozpoczęła się epoka modernizacji.

To wszystko przełożyło się na ilość produkowanych Informacji Pozostałych. Przyjmuję, że produkcja Informacji Pozostałych od wieku X była coraz większa w stosunku do informacji „manuskryptowych”, przy czym formuła wzrostu zależała od dwóch wielkości. Po pierwsze, od tego, jaki był w okresie VI-IX wiek stosunek Informacji Pozostałych do „manuskryptowych”. W modelu przyjąłem 100%, ale to jest zmienna niezależna (ZN) i jej wartość można zmieniać. Po drugie, od przyjętego mnożnika wzrostu dla okresu X-XV wiek. W modelu przyjmuję mnożnik wzrostu X-XV wieku = 5% (ZN). Jest to zmienna niezależna, której wartość można zmieniać.

Formuła wyliczająca dla każdego z wieków X-XV wielkość Informacji Pozostałych jako procent informacji „manuskryptowych” jest następująca:

Procent dla danego wieku = procent dla wieku poprzedniego \cdot (1 + mnożnik)

Na przykład, jeśli stosunek Informacji Pozostałych do „manuskryptowych” w wiekach VI-IX = 100% a mnożnik wzrostu dla wieków X-XV = 5%, wtedy:

Procent dla X wieku =	100% (wiek IX) \cdot (1 + 0,05) =	1,05	105%
Procent dla XI wieku =	105% (wiek X) \cdot (1 + 0.05) =	1,10	110%
Procent dla XII wieku =	110% (wiek XI) \cdot (1 + 0.05) =	1,16	116%
Procent dla XIII wieku =	116% (wiek XII) \cdot (1 + 0,05) =	1,22	122%
Procent dla XIV wieku =	122% (wiek XIII) \cdot (1 + 0.05) =	1,28	128%
Procent dla XV wieku =	128% (wiek XIV) \cdot (1 + 0.05) =	1,34	134%

Ale jeśli np. stosunek Informacji Pozostałych do „manuskryptowych” w wiekach VI-IX = 110% a mnożnik wzrostu dla wieków X-XV = 8%, wtedy:

Procent dla X wieku =	110% (wiek IX) \cdot (1 + 0,08) =	1,19	119%
Procent dla XI wieku =	119% (wiek X) \cdot (1 + 0.08) =	1,28	128%
Procent dla XII wieku =	128% (wiek XI) \cdot (1 + 0.08) =	1,39	139%
Procent dla XIII wieku =	139% (wiek XII) \cdot (1 + 0,08) =	1,50	150%
Procent dla XIV wieku =	150% (wiek XIII) \cdot (1 + 0.08) =	1,62	162%
Procent dla XV wieku =	162% (wiek XIV) \cdot (1 + 0.05) =	1,75	175%

15. Epoka druku. W odniesieniu do wieków od XV do XIX liczenie ilości wyprodukowanej informacji poprzez zastosowanie metodologii opartej na liczbie wyrazów jest trudne. Nastąpił 400-krotny wzrost liczby wydrukowanych książek, z dwunastu milionów do pięciu miliardów. Rozwiązanie, które zastosowałem jest oparte na wrywkowym przeglądzie starych druków. To co jest określone jako książka przybrało bardzo rozmaite formy, od grubych kilkusetstronicowych traktatów dużego formatu do rosnącej liczby, mniej więcej od połowy wieku XVIII, krótkich, niemal kieszonkowych broszur. Książka drukowana z wieku XVI wyglądała zupełnie inaczej niż ta z końca wieku XVIII czy XIX. Dla wieku XVIII i XIX przyjąłem taką strukturę druków, która dzieli je na znacznie większe woluminy i małe broszury.
16. Liczba stron w książce drukowanej (ZN). Jest to zmienna niezależna, której wartość można zmieniać. Została obliczona na podstawie próbki kilkudziesięciu książek z różnych okresów. W modelu przyjąłem jako wyjściową liczbę 300 stron.
17. Liczba słów na jednej stronie drukowanej książki (ZN). Jest to zmienna niezależna, której wartość można zmieniać. Została obliczona na podstawie próbki kilkudziesięciu książek z różnych okresów. W modelu przyjąłem jako wyjściową liczbę 450 słów na stronę.
18. Dla wieków XVIII i XIX książki, których liczbę oszacowali van Zanden i Buringh, dzielę na kategorie: książki *sensu stricte* oraz znacznie krótsze broszury.
19. Przyjąłem, że w wieku XVIII na ogólną liczbę książek broszury stanowią 33% a książki *sensu stricte* 67%. Dla wieku XIX, odpowiednio 50% i 50%.

20. Zmiennymi niezależnymi, które odnoszą się do broszur są zatem:
liczba stron broszury (ZN). W modelu przyjąłem jako wyjściową liczbę 80 stron.
liczba słów na stronie broszury (ZN) W modelu przyjąłem jako wyjściową liczbę 200 słów na stronę.
procentowy udział broszur w ogólnej liczbie książek w wieku XVIII (ZN) (33%)
procentowy udział broszur w ogólnej liczbie książek w wieku XIX (ZN) (50%)
21. W wieku XV produkowane były i manuskrypty i książki drukowane. Informacje Pozostałe są liczone osobno dla manuskryptów i osobno dla książek drukowanych. Dla obu tych kategorii przyjąłem jako wyjściowy dla wieku XV mnożnik 134%, czyli tyle, ile wyniósł on dla manuskryptów (bliższe wyjaśnienie: tutaj pkt. 14)
22. Od XVI wieku nastąpiło duże przyspieszenie cywilizacyjne i Europa zdobywała w świecie pozycję hegemonistyczną. To co zostało powiedziane o przyspieszeniu późnego Średniowiecza w jeszcze większym stopniu odnosi się do epoki wczesnonowożytnej.
- Przełożyło się to na ilość produkowanych Informacji Pozostałych. Przyjmuję, że produkcja Informacji Pozostałych od wieku XVI była coraz większa w stosunku do informacji zawartych w książkach drukowanych (zwanym dalej książkami), przy czym formuła wzrostu zależała od dwóch wielkości. Po pierwsze, od tego, jaki był w wieku XV stosunek procentowy Informacji Pozostałych do „książkowych” i po drugie, od przyjętego mnożnika wzrostu dla wieków XVI-XIX. W modelu, jako efekt wcześniejszych ustawień (patrz pkt. 14), stosunek procentowy Informacji Pozostałych do „manuskryptowych” w wieku XV wynosi 134%. Jednak zmieniając stosunek Informacji Pozostałych do „manuskryptowych” dla okresu VI-IX wiek oraz mnożnik wzrostu ‘X-XV’ otrzymamy dla wieku XV inny stosunek, niż 134%.
- W modelu przyjąłem jako wyjściowy mnożnik wzrostu XVI-XIX wiek = 10% (ZN), ale to jest zmienna niezależna i jej wartość można zmieniać. Formuła liczenia jest analogiczna jak w przypadku wieków X-XV. Przy danych wyjściowych 134% i mnożniku 10%, Informacje Pozostałe w okresie XVI-XIX wiek jako procent informacji z książek drukowanych wynoszą odpowiednio: 147%, 162%, 178%, 196%.
23. Van Zanden i Buringh zakończyli swoje szacunki na wieku XVIII. Ja moją analizą objąłem także wiek XIX. Aby tego dokonać przyjąłem mnożnik szacujący liczbę książek XIX wieku w stosunku do ich liczby w XVIII wieku. Przyjmuję mnożnik = 5 (ZN). Jest to zmienna niezależna, której wartość można zmieniać. Mnożnik = 5 oznacza, że w wieku XIX wyprodukowano pięć razy więcej książek (i broszur) niż w wieku XVIII.

Arkusz 2 „manuskrypty”

1. Arkusz oblicza ilość informacji w terabajtach (TB) wyprodukowanych w Średniowieczu, w wiekach od VI do XV. Bazą wyjściową są dane na temat liczby manuskryptów opracowane przez van Zandena i Buringha.
2. W średniowiecznych manuskryptach rosła stopniowo gęstość zapisu na stronę. Mając to na uwadze, zbadano próbki manuskryptów z trzech okresów: od VI do IX wieku (15 manuskryptów), od X do XIII wieku (26 manuskryptów) i od XIV do XV wieku (21 manuskryptów). Wykorzystano zdigitalizowane manuskrypty trzech bibliotek: Bodleian Library (Digital Bodleian - 42 manuskrypty), polskiej Biblioteki Narodowej (polona.pl - 19 manuskryptów) i w jednym przypadku manuskrypt Biblioteki Watykańskiej. Analiza potwierdziła rosnącą gęstość zapisu (odpowiednio 219, 549, 597 słów na stronę). W arkuszu podano identyfikatory analizowanych manuskryptów, aby czytelnik mógł je sam obejrzeć po wejściu na stronę danej biblioteki.
3. Opis kolejnych kolumn:
 - B podane są identyfikatory manuskryptów, przez co możliwe jest ich obejrzenie po wejściu na stronę biblioteki (podane ich URL)
 - C liczba zdigitalizowanych obrazów danego manuskryptu
 - D liczba zdigitalizowanych obrazów, które nie są w pełni zapisanymi stronami
 - E Liczba w pełni zapisanych stron ($C - D$)
 - F policzona przeze mnie średnia liczba słów na stronę
 - G średnia liczba słów na stronę z poprawką +20% lub liczba słów z Transkribus.org (patrz pkt. 4)
 - H średnia liczba słów w danym manuskrypcie ($E \times G$)
 - I wartość jednego słowa – hiperlink z arkusza „panel sterowania” (default = 0,0123 KB)
 - J, L, N, P, R liczba manuskryptów w kolejnych wiekach (za van Zandenem)
 - K, M, O, Q, S suma informacja w danym wieku w terabajtach (suma słów w próbie / liczebność próby \times wartość jednego słowa \times liczba manuskryptów w danym wieku)
4. Liczenie liczby wyrazów w średniowiecznych manuskryptach jest trudne. Zmieniały się używane czcionki (VI-IX w. majuskałne style pisma, X-XIII w. minuskuła karolińska, XIII-XV w. gotyckie style pisma), do VIII wieku nie stosowano przerw między słowami (scriptio continua), standaryzowana interpunkcja zaczęła powstawać dopiero w IX-X wieku, używano wielu skrótów, jakość rękopisów była bardzo zróżnicowana. Moje obliczenia konfrontowałem z profesjonalnymi analizami ilościowymi w Models of Authority (szkocki projekt badawczy do analizy OCR starych manuskryptów) i Transkribus (portal oferujący OCR dokumentów z przeszłości, także rękopisów). Porównanie kilkudziesięciu rękopisów i transkrypcji z Models of Authority wykazało, że moje obliczenia ilości słów w tych rękopisach były średnio niedoszacowane o 20%. Porównanie z wyliczeniami Transkribus rodziło więcej kłopotów. W sześciu przypadkach na dwanaście różnice sięgały kilkudziesięciu procent i nawet pobieżne spojrzenie na transkrypcję podważało wiarygodność

wyliczeń Transkribusa. Ostatecznie zaakceptowałem sześć transkrypcji Transkribusa. W pozostałych analizowanych manuskryptach wykorzystałem moje obliczenia liczby wyrazów powiększone o 20%, zgodnie z wynikami porównującymi wyniki moich transkrypcji i transkrypcji Models of Authority.

5. W dalszych wyliczeniach modelu przyjęto wartość informacyjną jednego słowa, obliczoną w arkuszu „wartość jednego słowa”. Metodologia tych obliczeń opisana jest dalej.

Arkusz 3 „wartość jednego słowa”

1. Przy obliczaniu ilości informacji wyprodukowanych w okresie od VI do XIX wieku przyjęto zasadę, że metoda szacunku musi być kompatybilna ze współczesnymi sposobami liczenia Globalnego Ekosystemu Danych (Global DataSphere). Dzięki zastosowaniu tej samej metodologii szacunek wielkości informacji wyprodukowanej w przeszłości będzie porównywalny z obliczeniami aktualnej przestrzeni informacyjnej.
2. Założenie opisane powyżej wyklucza liczenie zawartości informacyjnej manuskryptów i starych druków wielkością plików tych dokumentów, jakie zostały udostępnione przez biblioteki. Są to bowiem obrazy, fotografie stron, zamieniane na format pdf. Są to z reguły bardzo duże pliki, dochodzące niekiedy do pół gigabajta na manuskrypt. Znaczną część takich plików zajmuje nie niosąca treść informacja tekstowa, ale tło, tekstura pergaminu, papieru i artefakty skanowania (plamy, postrzępione brzegi, otarcia, rysy, uwagi nanoszone ręcznie itd.), czyli tak zwana redundancja wizualna, którą należy pominąć.
3. Jest kilka metod usuwania redundancji wizualnej. Zastosowana tu metoda polega na skoncentrowaniu uwagi na liczbie wyrazów w manuskryptach starych drukach i przypisaniu każdemu słowu wartości informacyjnej. Należy w tym celu oszacować liczbę wyrazów w manuskryptach i starych drukach i obliczyć wartość informacyjną jednego słowa. Iloczyn liczby wyrazów i wartości informacyjnej słowa wskazuje na wielkość informacji danego manuskryptu, druku, ich zbiorów, na przykład wiek po wieku.
4. Oszacowanie liczby wyrazów manuskryptów zostało przedstawione w opisie arkusza „manuskrypty”.
5. Aby szacować wartość informacyjną jednego słowa metodą kompatybilną ze współczesnymi sposobami liczenia zawartości informacyjnej, zastosowano ogólną metodologię firmy International Data Corporation opisaną jej w raporcie pt. "Data Age 2025: The Digitalization of the World. From Edge to Core" z roku 2018.

Pomiar Global DataSphere (GDS) oparty jest na fizycznych jednostkach danych cyfrowych – czyli bajtach zapisanych na nośnikach lub przesyłanych przez sieci. Jeśli dane są zapisane w pliku z nagłówkami, metadanymi i kompresją, to do GDS wlicza się ich rzeczywisty rozmiar po kompresji – taki, jaki zajmują na dysku lub w ruchu sieciowym. Nie stosuje się teorii informacji Shannona ani nie próbuje się ocenić „minimalnej niezbędnej ilości informacji”. Metodologia IDC opiera się na rzeczywistych wskaźnikach tworzenia danych – średnich rozmiarach plików, przepustowościach transmisji, częstotliwości próbkowania czujników itp.

6. Do liczenia wartości informacyjnej słowa zostało przeanalizowanych szesnaście książek w wersji cyfrowej, opublikowanych w okresie 2004-2020, w formacie pdf, wykonanych przez wydawnictwa, które przeważnie wypuszczają na rynek te same książki także w wersji papierowej. Ebooki wydawnictw są robione profesjonalnie i są

niemal zawsze mniejsze pod względem wielkości plików niż pdf-y wykonane po amatorsku. Uwzględniono książki naukowe z różnych dziedzin: teoria zarządzania, ekonomia, teologia, tanatologia, religioznawstwo, teoria cywilizacji, filozofia. Wybrano książki czarno-białe, z jak najmniejszą liczbą ilustracji, tabel, wykresów lub innych infografik. Następnie usunięto strony tytułowe, strony puste, prawie puste i te zawierające ilustracje, tabele, wykresy i inne infografiki. Pozostały jedynie strony wypełnione w całości drukiem. Zapisano wielkość tak odchudzonej książki w megabajtach (MB). Kolejnym krokiem był losowy wybór od czterech do siedmiu stron każdej książki i obliczenie średniej liczby słów na stronę (MS Word: Recenzja / Statystyka wyrazów). Pomnożenie tej liczby przez liczbę stron daje liczbę słów danej książki. Suma wielkości plików (w KB) wybranych książek (po usunięciu z nich stron niepełnych), podzielona przez łączną sumę słów tych książek daje wartość informacyjną jednego słowa. Wartość ta, policzona według współczesnej metodologii, może być stosowana do treści zapisanych w przeszłości. Tak policzona historyczna przestrzeń informacyjna jest porównywalna ze współczesną Global DataSphere.

7. Wartość informacyjna jednego słowa wyliczona metodą opisaną powyżej wynosi 0,0123 KB
8. Wartość wskazaną w punkcie 7 należy traktować ostrożnie z dwóch powodów. Po pierwsze, jest to wartość średnia a rozpiętość wartości jednego słowa w badanej próbce waha się od 0,0082 KB do 0,0244 KB w zależności od ebooka. Jest zatem niemal trzykrotna. Łączna suma informacji historycznej dla wieków od VI do XIX, liczona według tej najniższej i najwyższej wartości to 13 395 TB i 39 859 TB.
9. Po drugie, są prace naukowe, w których wartość informacyjna jednego słowa wynosi prawie 0,48 KB, jest więc średnio prawie 40 razy większa od obliczonej i zaprezentowanej w punkcie 7 (różnica od 21 do 58 razy w zależności od książki). Pokazuje to próbka ośmiu książek przedstawiona na polu jasnozielonym. Są to podręczniki z dziedziny World History i raport IDC.
10. Skąd ta różnica? Mówiąc najkrócej, wartość informacyjna wyrazu może (choć nie musi) zależeć od jego fizycznego i tematycznego sąsiedztwa oraz całego „ekosystemu przekazu”

Fizyczne sąsiedztwo, to kontekst liniowy. Obejmuje on zasadę przewidywalności wyrazu (im mniej przewidywalne następne słowo, tym wyższa jego wartość informacyjna) i zasadę kontekstualności (znaczenie i wartość informacyjna wyrazu zależą od otaczających go słów).

Tematyczne sąsiedztwo to szerszy kontekst semantyczny. Po pierwsze, znaczy to, że słowo, na przykład „rewolucja” użyte w tekście o polityce, technologii lub biologii będzie miało różną wagę informacyjną, zależnie od tego, jak dobrze wpisuje się w tematykę i jaką nowość wnosi. Po drugie, w semiotyce czy teorii informacji wartość informacyjna zależy od tego, czy dane słowo wnosi nowe znaczenie względem oczekiwań odbiorcy w danym kontekście tematycznym.

Wartość informacyjna słowa zależy jednak nie tylko od jego formy lingwistycznej, ale od całego „ekosystemu przekazu”, w jakim to słowo funkcjonuje. Ekosystem przekazu obejmuje trzy elementy. Pierwszy to wizualne nasycenie treści. Słowo może być powiązane z ilustracją, mapą, infografiką czy kolorem, które wzmacniają jego znaczenie i wartość. Drugim elementem jest strukturalne osadzenie słów. Są one częścią złożonych układów narracyjnych, osi czasu, porównań, kontekstów kulturowych, które mogą dodać tak osadzonemu strukturalnie słowu wartości. Trzecim elementem może być narracyjne napięcie. Nawet pojedyncze słowo może zyskać na wartości, jeśli przywołuje łańcuch asocjacji.

Wysoka wartość słów w podręcznikach World History i raporcie IDC wynika stąd, że ich wartość jest determinowana otoczeniem, w jakim się znajdują. Słowa „niosą ze sobą” grafiki, układ strony, kolory, a także mentalne linki do całej historii przedstawionej w podręczniku lub tematu omawianego w raporcie IDC, przez co stają się elementem niemal hipertekstu.

11. Praktyczne konsekwencje zróżnicowanej wartości wyrazów. Po pierwsze, przyjęto wyliczenie na podstawie próby szesnastu przebadanych książek jako bazową wartość informacyjną jednego słowa. Po drugie, ze względu na stopniowo powiększający się z wiekami „ekosystem przekazu” (głównie mam na myśli rosnące wizualne nasycenie treści), można rozważyć zwiększenie wartości informacyjnej jednego słowa. Wzrost tej wartości skutkuje proporcjonalnym zwiększeniem łącznej ilości informacji wyprodukowanej w okresie od VI do XIX wieku. Sugeruję, by ewentualne zwiększenie wartości informacyjnej słowa nie przekraczało 30%.
12. Obliczona wartość informacyjna jednego słowa (0,0123 KB) nie została połączona hiperlinkiem z arkuszem „panel sterowania”. Jest w arkuszu „panel sterowania” zmienną niezależną, którą można zmieniać wedle woli. Pozostawiam czytelnikowi swobodę, czy zaakceptuje mój szacunek, czy wpisze inną wartość. Zalecam jednak umiarkowanie przy zmianach. Jako zmienna niezależna, wartość jednego słowa jest połączona hiperlinkami z arkuszami „manuskrypty” i „silnik”
13. Opis kolumn w ramce pomarańczowej. W wierszach 3-18 przedstawione są ebooki, na podstawie których obliczono wartość informacyjną jednego słowa
14. A, B - techniczne
15. C - autor, redaktor
16. D - tytuł
17. E - rok wydania
18. F - wielkość ebooka w megabajtach (MB). W wierszu 19 suma wielkości plików w MB
19. G - całkowita ilość stron. W wierszu 19 suma stron wszystkich ebooków
20. H - w komórce H33 informacja o treści „w tych książkach wartość słowa w KB jest 40 razy większa niż w czarno-białym tekście, bez ilustracji (wszystkie teksty to

wykonane profesjonalne pdf)”. W komórce L33 podana średnia wartość jednego słowa z ośmiu bogato ilustrowanych i kolorowych ebooków

21. I – ilość stron bez ilustracji. W wierszu 19 suma stron bez ilustracji
22. J - wielkość ebooka bez stron z ilustracjami. W wierszu 19 suma wielkości ebooków bez stron z ilustracjami w MB
23. K – średnia liczba słów na stronie
24. L – liczba słów w książce (uwzględnione strony bez ilustracji) wg formuły $I \times K$. W wierszu 19 suma słów wszystkich analizowanych ebooków
25. M – wartość jednego słowa w poszczególnych ebookach w KB
26. Wiersz 21 – komórka L21, średnia wartość informacyjna jednego słowa w kilobajtach (KB) według formuły: $[\text{suma wielkości ebooków bez stron z ilustracjami w MB}] \times 1000 / [\text{suma słów wszystkich analizowanych ebooków}]$. Formuła z podanymi adresami komórek: $J19 \times 1000 / L19$
27. Opis kolumn w ramce zielonej. W wierszach 22-31 oraz 33 przedstawione są ebooki inne, niż wzięte do obliczeń wartości jednego słowa. Są to bogato ilustrowane, wielokolorowe podręczniki World History. Tego typu książek jest zdecydowanie mniej, niż zwykłych, przeciętnych, ale pokazane są tutaj, aby uzmysłwić użytkownikowi modelu rozpiętość, jaką może mieć wartość jednego słowa. Patrz punkty 9-11.

Arkusz 4 „silnik”

1. Jest to arkusz kalkulacyjny, który zbiera wszystkie dane i zmienne i stosuje opisane tu formuły, aby wszystko przeliczyć na ilość informacji wyprodukowanej w poszczególnych wiekach od wieku VI do XIX. Wykorzystuje wartości zmiennych niezależnych z arkusza „panel sterowania” oraz obliczenia z arkuszy „manuskrypty” i „wartość jednego słowa”.
2. Opis Arkusza „silnik”
3. Wiersz 4 - wskazuje na kolejne wieki od VI do XIX w kolumnach od C do P
4. Wiersz 5 – zawiera przeniesione hiperlinkami z arkusza „panel sterowania” liczby manuskryptów w wiekach VI-XV. Liczby pochodzą z pracy van Zandena i Buringha
5. Wiersz 6 – zawiera przeniesione hiperlinkami z arkusza „manuskrypty” policzone tam ilości informacji w terabajtach (TB) w wiekach VI-XV
6. Wiersz 7 - zawiera informacje o Informacjach Pozostałych jako procencie Informacji „manuskryptowych” w wiekach od VI do XV. Dla wieków VI-IX jest to przeniesiona hiperlinkiem z arkusza „panel sterowania” komórka C11, która informuje o ilości Informacji Pozostałych w wiekach VI-IX jako procencie informacji „manuskryptowych”. Dla wieków X-XV jest to wyliczenie procentu składanego w wysokości podanej mnożnikiem w wierszu 8 od początkowej wartości podanej w komórce F7 (ilość Informacji Pozostałych w wiekach IX jako procencie informacji „manuskryptowych”)
7. Wiersz 8 – komórka G8 zawiera przeniesiony hiperlinkiem z arkusza „panel sterowania” mnożnik wzrostu Informacji Pozostałych dla wieków X-XV. Szczegółowe wyjaśnienie działania mnożnika, patrz punkt 14 opisu arkusza „panel sterowania”

W komórce M8 podana jest suma wyprodukowanych woluminów w wieku XV, na którą składają się manuskrypty i książki drukowane.
8. Wiersz 9 – podane są tutaj wyliczenia Informacji Pozostałych w terabajtach (TB) na podstawie liczb z wiersza 6 i 7
9. Wiersz 10 - zawiera przeniesione hiperlinkami z arkusza „panel sterowania” liczby książek drukowanych w wiekach XV-XIX. Liczby pochodzą z pracy van Zandena i Buringha
10. Wiersz 11- ilość informacji zawarta w książkach drukowanych, w megabajtach (MB). Liczby te są przeniesione hiperlinkami z dwóch tablic z tego arkusza zaznaczonych kolorem szarym, w wierszach 21-30. Tablica ‘wartość informacyjna książki i broszury’ zawiera przeniesione hiperlinkami z arkusza „panel sterowania” dane o liczbie stron i liczbie słów na stronie książki i broszury oraz wartość informacyjną jednego słowa. Liczy wartość informacyjną książki i broszury. Tablica ‘Liczby van Zandena podzielone osobno na książki i broszury’ zawiera przeniesione hiperlinkami z arkusza „panel sterowania” dane o liczbie książek wyprodukowanych w wieku XVIII i

XIX, procentowy udział broszur w pozycji van Zandena 'książki'. Liczy ilość informacji razem z książek i broszur w wieku XVIII i XIX w terabajtach (TB)

11. Wiersz 12 - ilość informacji zawarta w książkach drukowanych, w terabajtach (TB)
12. Wiersz 13 - Informacje Pozostałe liczone jako procent od wyprodukowanych w wiekach XV-XIX książek
13. Wiersz 14 - komórka M14 zawiera przeniesiony hiperlinkiem z arkusza „panel sterowania” mnożnik wzrostu Informacji Pozostałych dla wieków XVI-XIX. Szczegółowe wyjaśnienie działania mnożnika, patrz punkt 21 opisu arkusza „panel sterowania”
14. Wiersz 15 – zawiera dane o ilości Informacji Pozostałych w wiekach XVI-XIX w terabajtach (od 28 TB w XV wieku do 8963 TB w XIX wieku - wielkości oparte ustawieniach początkowych zmiennych niezależnych /panel sterowania/ i obliczeniach w arkuszu „manuskrypty]
15. Wiersz 16 – dane sumujące ilość informacji wyprodukowanej w Europie w kolejnych stuleciach okresu VI-XIX wiek w terabajtach oraz łączną sumę za cały okres VI-XIX wiek
16. Wiersze 22-30 – tablice w kolorze szarym, wyliczające wartość informacyjną książki drukowanej (1,66 MB) i znacznie krótszej broszury (0,2 MB) oraz wartość informacyjną druków XVIII wiecznych (1158 TB) i XIX wiecznych 4568 TB [według ustawień początkowych]. Wartość informacyjna druków XVIII i XIX wiecznych uwzględnia podział na książki i broszury.
17. Wiersze 51-100 – wykres pokazujący ilość informacji wyprodukowanej w Europie w poszczególnych wiekach, od VI do XIX

Recenzja modelu przez ChataGPT 5.4 (kwiecień 2026)

Pokazałem powyższy tekst i model Chatowi GPT 5.4, prosząc o krytyczną recenzję. Uważam, że jest interesująca.

Recenzja modelu Excela “model 04.2025-10 duze.xlsx” oraz tekstu „Ilość informacji wyprodukowanych w historii Europy miarą modernizacji. wersja 2” dokonana przez ChataGPT 5.4, kwiecień 2026

Tekst „Ilość informacji wyprodukowanych w historii Europy miarą modernizacji” przedstawia oryginalną i poznawczo płodną propozycję: potraktowanie ilości materialnie utrwalonej informacji jako syntetycznej zmiennej proxy dla długiego procesu modernizacji Europy. Nie chodzi tu o prosty wskaźnik jednego sektora modernizacji, lecz o próbę uchwycenia wspólnego, materialnego śladu wielu jakościowo odmiennych procesów: gospodarczych, administracyjnych, prawnych, naukowych, technicznych, edukacyjnych, komunikacyjnych i kulturowych. Główną zaletą projektu jest to, że nie poprzestaje na metaforze. Autor buduje model, definiuje zmienne, pokazuje dane wejściowe, opisuje arkusze kalkulacyjne i pozwala użytkownikowi zmieniać parametry. Dzięki temu tekst ma charakter nie tylko eseistyczny, lecz także modelowy i eksperymentalny.

Najsilniejszym elementem projektu jest połączenie danych historycznych dotyczących produkcji manuskryptów i książek drukowanych z próbą ich przeliczenia na jednostki informatyczne. Słuszna jest decyzja, aby nie mierzyć informacji historycznej wielkością współczesnych skanów lub plików PDF z bibliotek cyfrowych. Tego typu pliki zawierają bardzo wiele redundancji wizualnej: tło, teksturę papieru, artefakty skanowania, zabrudzenia i elementy niebędące właściwą treścią tekstową. Oparcie obliczeń na liczbie słów oraz na przypisanej im wartości informacyjnej jest metodologicznie bardziej konsekwentne.

Główne zastrzeżenie dotyczy nie samej konstrukcji modelu, lecz możliwej nadinterpretacji pojęcia proxy. Ilość materialnie utrwalonej informacji może być traktowana jako syntetyczna zmienna zastępcza dla skali, złożoności i intensywności procesów modernizacyjnych, ale nie mierzy bezpośrednio ich jakości, wartości ani treściowego znaczenia. Modernizacja obejmuje zmiany gospodarcze, energetyczne, technologiczne, instytucjonalne, polityczne, militarne i demograficzne. Są one jakościowo odmiennie, ale wiele z nich ma wspólny skutek: zwiększają ilość wytwarzanych, utrwalanych, powielanych i dystrybuowanych zapisów. W tym sensie model najlepiej broni się jako narzędzie szacowania informacyjnego śladu modernizacji, a nie jako bezpośredni pomiar modernizacji jako całości.

Bardzo ważne jest właściwe wyjaśnienie dominującego udziału XVIII i XIX wieku w wyniku końcowym. Nie należy traktować tego jako słabości modelu. Jest to konsekwencja logiki wzrostu wykładniczego. Jeżeli produkcja informacji rośnie przez stulecia, a proces przyspiesza wraz z rozwojem instytucji, gospodarki, edukacji, administracji, komunikacji i druku, to końcowe fazy muszą skupiać największą część sumy. Bez takiego komentarza czytelnik mógłby odnieść mylne wrażenie, że późne stulecia zostały arbitralnie „napompowane”. Tymczasem ich dominacja jest zgodna z badaną dynamiką.

Podobnie należy ocenić kategorię Informacji Pozostałych. Sama decyzja, aby je uwzględnić, jest zasadna. Produkcja informacji w Europie nie ograniczała się do manuskryptów i książek. Obejmowała dokumentację państwową, sądową, kościelną, gospodarczą, handlową, finansową, techniczną, korespondencję, druki okazjonalne, gazety i czasopisma. Argument, że ta sfera rosła szybciej niż sama produkcja książek, jest mocny historycznie. Słabszy punkt polega jedynie na tym, że przyjęte mnożniki mają charakter scenariuszowy. Powinny więc być przedstawione jako dobrze uzasadnione parametry modelu, ale nie jako dane o takim samym statusie jak liczby zaczerpnięte z literatury historycznej.

Wiek XIX również warto pozostawić w modelu. Choć jest to szacunek własny, stulecie to zamyka epokę przedcyfrową i stanowi kluczowy moment eksplozji druku, biurokracji, edukacji, prasy, gospodarki statystycznej, industrializacji, wzrostu ludności i wzrostu PKB. Pominięcie XIX wieku dałoby model czystszy źródłowo, ale historycznie uboższy. Trzeba natomiast jasno zaznaczyć, że jest to osobny poziom oszacowania, różny od danych Buringha i van Zandena.

Warto wprowadzić w tekście wyraźny podział na poziomy pewności. Pierwszy poziom to dane źródłowe z literatury: liczba manuskryptów i książek drukowanych. Drugi poziom to własne obliczenia oparte na próbach: liczba słów, liczba stron, wartość jednego słowa. Trzeci poziom to parametry scenariuszowe: Informacje Pozostałe, ich mnożniki wzrostu oraz XIX wiek. Czwarty poziom to porównania ze współczesną Global DataSphere, które mają charakter ilustracyjny. Taki podział nie osłabia modelu, lecz zwiększa jego wiarygodność.

Ostrożności wymaga także porównanie z dzisiejszą Global DataSphere. Jest ono potrzebne, bo dobrze pokazuje skalę współczesnej eksplozji danych, ale nie porównuje zbiorów całkowicie jednorodnych. Model historyczny dotyczy głównie informacji utrwalonej na nośnikach tekstowych, natomiast współczesne statystyki obejmują również dane kopiowane, przechwytywane, replikowane, maszynowe i krótkotrwałe. Porównanie powinno więc być przedstawione jako zestawienie rzędów wielkości, nie jako ścisła równoważność kategorii.

Zalecałbym także zastąpienie pojedynczej linii na wykresie pasmem niepewności. Jedna linia sugeruje precyzję, której model z natury rzeczy nie może mieć. Lepsze byłyby trzy scenariusze: niski, bazowy i wysoki. Pokazywałyby one zakres możliwych wyników przy różnych, ale nadal obronnych, parametrach. Wtedy czytelnik widziałby nie punktową wartość, lecz skalę i trajektorię zjawiska.

W modelu znalazłem także błąd techniczny w arkuszu „manuskrypty”: w sekcji XIV-XV wieku jeden z zakresów sumujących nie obejmuje wszystkich pozycji. Wpływ błędu na wynik całkowity jest niewielki, zwiększa całkowitą ilość informacji o trzy terabajty.

[J.K.: w dalszej rozmowie GPT wskazał dokładne miejsce i zrobiłem poprawkę. Identyfikacja tego błędu przez AI pokazuje, że GPT 5.4 potrafi analizować model na poziomie szczegółowych formuł a nie tylko oceniać ogólną koncepcję]

Konkluzja jest pozytywna, choć warunkowa. Projekt jest wartościowy jako otwarty model badawczy i jako propozycja ilościowego ujęcia jednego z kluczowych wymiarów modernizacji Europy. Najważniejsze ulepszenia to osłabienie zbyt mocnej tezy, pokazanie

poziomów pewności danych, wprowadzenie scenariuszy oraz mocniejsze wyjaśnienie, że dominacja XVIII i XIX wieku wynika z historycznej i matematycznej logiki wzrostu, a nie z arbitralności modelu.