

Wiersz smogowy - raport techniczny

Leszek Onak, Jacek Olczyk

Grudzień 2019

UBU-19-03

Raport techniczny z
UBU lab
Uniwersytet Jagielloński
Wydział Zarządzania i Komunikacji Społecznej
Instytut Kultury
ul. Łojasiewicza 4, p.2.120, 30-348 Kraków
piotr.marecki@uj.edu.pl
www.ubulab.edu.pl

Praca naukowa finansowana
w ramach programu Ministra Nauki
i Szkolnictwa Wyższego pod nazwą
„Narodowy Program Rozwoju Humanistyki”
w latach 2016-2019

Copyright © by the Authors, 2019

Spis treści

- Wstęp · 3
- 1. Opis utworu · 3
- 2. Koncepcja utworu · 4
- 3. Zanieczyszczenie powietrza - dane, którymi oddychamy · 5
 - 3.1. Źródło danych · 5
 - 3.2. Parser · 6
 - 3.3. Przyjęta skala zanieczyszczenia · 6
- 4. Wiersz smogowy jako utwór potencjalny · 7
- 5. Przymus smogowy - metoda pojedynczego cięcia zdań z prawdopodobieństwem warunkowym zależnym od zewnętrznych danych · 8
 - 5.1. Korpus · 8
 - 5.2. PM 10 · 9
 - 5.3. PM 2.5 · 9
- 6. Historia powstania, platforma i ograniczenia · 11
 - 6.1. Historia powstania · 11
 - 6.2. Platforma · 11
 - 6.3. Ograniczenia · 12
 - 6.3.1. Ekstrakcja zdań z kodu źródłowego · 12
 - 6.3.2. Czas ładowania się strony z uruchomioną wtyczką · 12
 - 6.3.3. Waga wtyczki · 12
 - 6.3.4. Pamięć podręczna przeglądarki · 12
- 7. Mechanizm utworu · 13
 - 7.1. Budowanie korpusu · 13
 - 7.1.1. Pobranie tekstu ze strony internetowej · 13
 - 7.1.2. Ekstrakcja zdań · 14
 - 7.1.3. Zapisanie zdań w pamięci przeglądarki · 14
 - 7.1.4. Przygotowanie N-gramowego modelu bazy tekstowej · 15
 - 7.2. Zanieczyszczenie strony · 15
 - 7.2.1. Wytypowanie zdań do przetworzenia · 15
 - 7.2.2. Skażenie tekstu · 16
- 8. Konteksty kulturowe · 19
 - 8.1. Smog w sztuce i smog jako problem społeczny · 19
 - 8.2. Smog w kulturze polskiej · 21

Data Dramatization, as opposed to data visualization presents a data set with not only legibility or clarity but in such a way as to provoke an empathetic or emotive response in its audience.

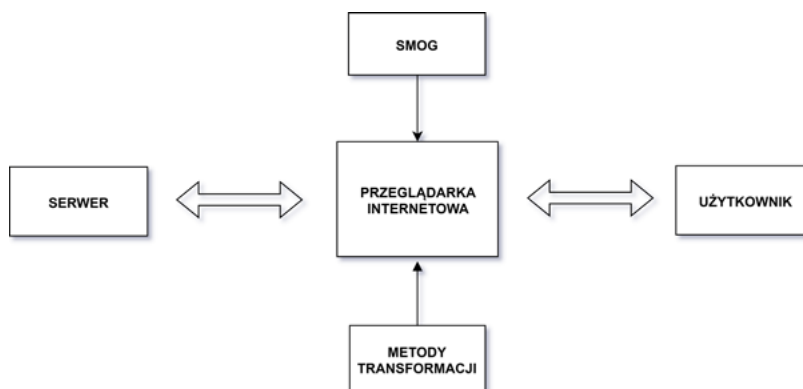
Liam Young

Wstęp

Jednym z zadań w ramach grantu naukowego „Twórcze programowanie - Laboratorium” w latach 2016-2019, przyznanego przez Narodowy Program Rozwoju Humanistyki, było stworzenie przez Leszka Onaka, jednego z uczestników grantu, utworu potencjalnego pt. *Wiersz smogowy*. W niniejszym raporcie przedstawiamy założenia, mechanizm działania, historię powstania oraz opis techniczny aplikacji, a także wpisujemy ją kontekst teorii przymusu grupy Oulipo.

1. Opis utworu

Wiersz smogowy jest wtyczką do przeglądarki internetowej wykorzystującą dane o skażeniu powietrza w Polsce i - zależnie od jego stężenia - zanieczyszczającą odwiedzane przez internautów strony. Utwór składa się z dwóch głównych silników: mechanizmu zaciągającego informacje o aktualnym skażeniu powietrza oraz mechanizmu odpowiadającego za wczytanie materiałów ze stron internetowych i ich modyfikację. Po przekroczeniu norm pyłu zawieszonego PM 10 w tkankę przeglądanej witryny „wstrzykiwane” są treści ze stron, które użytkownik kiedyś odwiedził. Metoda transformacji opiera się na połączeniu ze sobą dwóch ciągów tekstowych: zdania z aktualnie przeglądanej strony ze zdaniem archiwalnym znajdującym się w bazie tekstowej. Poziom pyłu zawieszonego PM 10 w powietrzu odpowiada za liczbę zarażonych jednostek tekstowych na stronie, natomiast stężenie PM 2.5 wpływa na stopień ich deformacji. Im większe zanieczyszczenie, tym zestawiane ze sobą elementy tworzą coraz mniej prawdopodobne związki znaczeniowe.



Rysunek 1. Schemat komunikacyjny *Wiersza smogowego*

Wtyczka jest wirusem, którego użytkownik dobrowolnie wpuszcza do komputera, by zanieczyścić przeglądane przez niego witryny - „niechciany” oprogramowaniem, które instaluje się w BIOS-ie przeglądarki. Jego zadaniem nie jest jednak wykradanie czy kasowanie danych, a informowanie w nietypowy sposób o skażeniu powietrza.

2. Koncepcja utworu

W 1948 roku Claude E. Shannon opublikował pracę *Matematyczna teoria komunikacji*¹, w której zaproponował podstawy współczesnej teorii informacji i opisał systemem binarnym, czyli ciągiem zer i jedynek, zależności warunkowe między literami i słowami w języku angielskim. Przy okazji rozważań nad pojęciem entropii przedstawił model generatora tekstowego budującego nowe teksty na podstawie korpusu, ale opierając się nie na niezależnych zmiennych losowych, a na prawdopodobieństwie warunkowym. Wykorzystał do tego koncepcję łańcuchów Markova, czyli ciągu zdarzeń, w którym prawdopodobieństwo każdego zdarzenia zależy tylko od zdarzenia poprzedniego.

5. First-order word approximation. Rather than continue with tetragram, . . . , n -gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.

REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERI GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.

6. Second-order word approximation. The word transition probabilities are correct but no further structure is included.

THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

Rysunek 2. Zdania wygenerowane przy pomocy łańcuchów Markova przy uwzględnieniu następstwa jednego wyrazu (punkt 5) i sekwencji dwóch wyrazów (punkt 6)²

Rozważania teoretyczne Shannona dały podwaliny pod współczesne modele informacji, a zastosowanie językowych modeli łańcuchów Markova przyczyniło się do gwałtownego rozwoju dziedziny komputerowego przetwarzania tekstów.

Minęło siedemdziesiąt lat od publikacji artykułu i łańcuchy Markova możemy znaleźć wszędzie: w telefonach ułatwiają pisanie wiadomości prywatnych, podpowiadając kolejne słowo na podstawie wbudowanego słownika i historii dotychczasowych konwersacji; usprawniają procesy automatycznego rozpoznawania mowy i maszynowego tłumaczenia tekstów. W stylometrii zwiększają skuteczność analizy statystycznej korpusów i wykrywania autorstwa książek. Połączone z maszynowym uczeniem - wspomagają budowanie między innymi czatbotów.

Ta rewolucja ma również swoje ciemne strony. Zaawansowane technologie przetwarzania dużych korpusów tekstów z profilowaniem użytkowników, zbieraniem danych o ich aktywności, umożliwiło - na wcześniej niespotykaną skalę - manipulowanie społeczeństwem. Rewolucja Web 2.0, która z jednej strony miała zdecentralizować ośrodki nadawcze, a z drugiej dać głos każdemu mieszkańcowi cyfrowego świata, została wykorzystana do: tworzenia behawioralnych modeli statystycznych i na ich podstawie automatycznego dostosowywania komunikatów reklamowych czy generowania sztucznych bytów, fake'owych profili i botów rozpowszechniających nieprawdziwe informacje w celu wpłynięcia na głosujących.

Wykorzystanie technologii językowych do szerzenia dezinformacji było jedną z inspiracji do odzwierciedlenia zanieczyszczenia powietrza na przestrzeń internetową. W utworze przejęto mechanizmy, którymi posługują się podmioty generujące fake'owe treści (oczywiście w bardzo okrojonej wersji). Aplikacja, po zainstalowaniu, zbiera informacje o użytkowniku: z każdej przeglądanej strony kopiuje zdania, a następnie posiłkuje się nimi do modyfikowania aktualnie otwartej witryny przy użyciu prawdopodobieństwa warunkowego - matematycznego modelu obliczającego prawdopodobieństwo wystąpienia zdarzenia pod warunkiem zajścia zdarzenia poprzedniego.

¹ C.E. Shannon, *A Mathematical Theory of Communication*, „The Bell System Technical Journal” 1948, No. 27, s. 379.

² Źródło: Claude E. Shannon, *A Mathematical Theory...*, s. 379.

Szczególną uwagę przywiązano do takiego przygotowania metod generowania nowych treści, by skażenie tekstu było przede wszystkim czytelne i nie zwracało uwagi na błędy gramatyczne (choć nie do końca udało się je wyeliminować).

3. Zanieczyszczenie powietrza - dane, którymi oddychamy

Głównym aktorem biorącym udział w utworze jest smog - mieszanina mgły, dymu i spalin; chmura pyłów i związków chemicznych szkodliwych dla zdrowia człowieka, zwierząt i środowiska naturalnego. W przeciwieństwie do bezbarwnego dwutlenku węgla współodpowiedzialnego za procesy globalnego ocieplenia smog jest widoczny. W sezonie grzewczym powietrze traci swoją przezroczystość. Staje się nośnikiem informacji o stężeniu szkodliwych substancji, wiadomością, której adresatem są wszyscy mieszkańcy miasta. Nawiązując do słów Jussi Parikka - przeistacza się w wizualne medium, którym oddychamy:

Air pollution is in many ways visual media you breathe. This is the elemental media condition across the aesthetic landscapes of contemporary industrial and post-industrial life. An urban screen is a residue of the transport cultures of automobiles as well as other sources of fossil fuel culture³.

Od wielu lat słyszymy, że za wysokie pomiary szkodliwych substancji odpowiedzialni są ludzie palący w piecach odpadami i złej jakości węglem, ale problem jest o wiele bardziej skomplikowany. Na stan jakości powietrza wpływ mają lata dzięki zabudowy przez deweloperów kanałów przewietrzania miasta, wycinanie naturalnych płuc metropolii: skupisk drzew, przy równoczesnym braku inicjatyw budowy parków. Za pięćdziesiąt procent skażenia lokalnej atmosfery odpowiada pokłosie rewolucji przemysłowej opartej na węglu i ropie oraz skutki masowej produkcji, między innymi transport drogowy, procesy spalania w przemyśle i transformacji energii, a także procesy produkcyjne. Są to lata systemowych zaniedbań i omijania problemu przez władze miast.

Pomimo że działanie utworu oparte jest na stężeniu pyłów zaliczających się do zjawiska smogu, to jego wydźwięk i koncepcja jest o wiele szersza i dotyczy w ogólności zanieczyszczenia powietrza, wpływu industrialnej i postindustrialnej rewolucji na zdrowie ludzi i środowisko naturalne planety.

3.1. Źródło danych

Jako źródło informacji o jakości powietrza wykorzystano udostępnione przez Główny Inspektorat Ochrony Środowiska bieżące pomiary zawartości substancji szkodliwych pobrane ze stacji umiejscowionych na terenie Polski. Dane przedstawione są w formacie JSON i aktualizowane co godzinę. W utworze spośród wszystkich miast pominięto te lokalizacje, w których stacje pomiarowe nie podają automatycznych odczytów dla pyłu zawieszono PM 10 i PM 2.5.

Po zainstalowaniu aplikacji użytkownik ma do wyboru pięćdziesiąt pięć miejscowości, których skażenie wpłynie na przeglądane przez niego strony internetowe: Augustów, Białystok (dwie stacje), Bielsko-Biała (dwie stacje), Borsukowizna, Duszniki-Zdrój, Elbląg, Gdańsk (cztery stacje), Goczałkowice-Zdrój, Gołdap (dwie stacje), Jarosław, Jawor, Jelenia Góra, Kalisz, Katowice, Kędzierzyn-Koźle, Kielce, Konstancin-Jeziorna, Kościerzyna, Kraków (osiem stacji), Kutno, Legionowo, Lublin, Łódź (trzy stacje), Mielec, Mińsk Mazowiecki, Mogilno, Nałęczów, Nowa Sól, Olsztyn, Otwock, Piastów, Płock, Połaniec, Poznań (dwie stacje), Prudnik, Przemyśl, Racibórz, Radom, Rymanów-Zdrój, Rzeszów (dwie stacje), Siedlce, Skarżysko-Kamienna,

³ J. Parikka, *The Sensed Smog: Smart Ubiquitous Cities and the Sensorial Body*, „The Fibreculture Journal” 2017, No. 29, <http://twenty-nine.fibreculturejournal.org/fcj-219-the-sensed-smog-smart-ubiquitous-cities-and-the-sensorial-body> [dostęp: 9.10.2019].

Starachowice, Szczecin (trzy stacje), Tarnów (dwie stacje), Toruń (trzy stacje), Warszawa (sześć stacji), Włocławek (dwie stacje), Wrocław, Wschowa, Zgierz, Zielona Góra, Żłoty Potok, Żary, Żyrardów.

3.2. Parser

Do pobrania i analizy danych zbudowano parser, który co godzinę łączy się z serwerem Głównego Inspektoratu Ochrony Środowiska i wczytuje aktualne stężenie substancji szkodliwych, dokonuje analizy udostępnionego pliku i kopiuje wartości dla poszczególnych wskaźników (PM 10, PM 2.5) z wszystkich stacji pomiarowych. Następnie aplikacja oblicza średnią arytmetyczną dla każdego zanieczyszczenia, dodając do siebie pomiary i dzieląc je przez liczbę wszystkich stacji badawczych, z których zostały zaciągnięte. Program został umieszczony na oddzielnym serwerze, na którym co godzinę publikuje wykstrahowane, uśrednione dane - przygotowane do pobrania przez mechanizm wtyczki.

3.3. Przyjęta skala zanieczyszczenia

Do wyliczenia stopnia skażenia powietrza zastosowano tzw. wspólny indeks jakości powietrza (Common Air Quality Index, CAQI), który wskazuje pięć poziomów wpływu skażenia atmosferycznego na zdrowie człowieka. Zdecydowano się na powyższą skalę ze względu na to, że opiera się na ogólnodostępnym wzorze matematycznym, stworzonym na podstawie wyników badań naukowych.

	PM 10	PM 2.5
Bardzo wysoki	> 180	> 110
Wysoki	180	110
	90	55
Średni	90	55
	50	30
Niski	50	30
	25	15
Bardzo niski	25	15
	0	0

Tabela 1. Wspólny Indeks Jakości Powietrza dla odczytów aktualizowanych co godzinę ($\mu\text{g}/\text{m}^3$)⁴

W utworze nie wykorzystano poziomów informowania i alarmowania ustanowionych przez ministra środowiska. Polskie normy w porównaniu z normami wielu innych państw europejskich są niekiedy nawet czterokrotnie wyższe. W przypadku pyłu zawieszzonego PM 10 poziomy informowania (w mikrogramach na metr sześcienny, dalej: $\mu\text{g}/\text{m}^3$) to: Szwajcaria - 75, Słowacja - 100, Finlandia - 50, Włochy - 50, podczas gdy w Polsce jest to 200. Z kolei poziomy alarmowe to: Szwajcaria - 100, Słowacja - 150, Finlandia - 80, Włochy - 75, a Polska - 200. Unia Europejska nie ujednoliciła poziomów zanieczyszczenia powietrza, które wymagają od rządów poinformowania społeczeństwa lub wprowadzenia nadzwyczajnych działań chroniących mieszkańców. Każdy kraj może takie normy określić sam.

4 Źródło: https://www.airqualitynow.eu/pl/about_indices_definition.php [dostęp: 12.02.2019].

4. Wiersz smogowy jako utwór potencjalny

Wiersz smogowy jest utworem potencjalnym, opartym na wykorzystaniu przymusu smogowego ilustrującego stopień zanieczyszczenia powietrza. Przymus wykorzystuje dwie najbardziej szkodliwe substancje emitowane do atmosfery. Każdej opisanej substancji przypisany jest odpowiedni mechanizm „zanieczyszczania” stron internetowych.

Przy opisie i analizie utworu zasadne wydaje się zapożyczenie kilku pojęć z teorii i praktyki francuskiej grupy Oulipo. Pojęcie przymusu (fr. *contrainte*) zostało na początku lat 60. zaproponowane przez François Le Lionnais, założyciela Oulipo, jako konkurencja dla innych pojęć, takich jak: struktura (fr. *structure*), forma (fr. *forme*), reguła (fr. *règle*) czy rygor (fr. *rigueur*). Oulipijczycy szczególnie chcieli zastąpić słowo „struktura” - które nie było zbyt wyszukane w dobie szalejącego we francuskiej nauce strukturalizmu - innym terminem, pozbawionym takiego balastu znaczeń, a z drugiej strony lepiej opisującym oulipijską koncepcję. Tradycja, do której Oulipijczycy często nawiązują, to między innymi działalność tzw. wielkich retyryków na przełomie XV i XVI wieku, od których bezpośrednio zaczerpnęli pojęcia *ouvroir* (warsztat) i *contrainte* (przymus).

Przymus oznacza dla pisarzy Oulipo pewną potencjalną strukturę, która organizuje język, temat i jest wyrazem dobrowolnie przyjętego, a nie narzuconego ograniczenia. Jacques Jouet pisał, że:

(...) aby zaistniał oulipijski przymus, musi zostać wykorzystana pewna jawna procedura, aksjomat porządku formalnego, którego implikacje i następstwa logiczne utworzą tekst. Przymus jest problematem, tekst jego rozwiązaniem. Lub też, by tak rzec, przymus jest sformułowaniem zagadki, a tekst odpowiedzią czy raczej jedną z wielu możliwych odpowiedzi. Przymus jest więc czymś innym niż zwykłe sposoby organizowania pracy literackiej. Te z pewnością są bardzo ważne, ale nie na tym polega przymus. Przymus jest systematyczny. Poza tym przymus oulipijski musi być użyteczny dla innych pisarzy, dlatego powinien być sformułowany w jasny sposób (formalizacja). (...) Pisanie wedle przymusów to szczególny sposób uprawiania literatury. Nie jedyny. Jeden z wielu⁵.

Dobrowolne używanie przymusu ma gwarantować pisarzowi świadomość ograniczeń, jakie stawia język, a jednocześnie wyzwalać inwencję twórczą.

W wielu przypadkach przymus staje się także rodzajem maszyny, która raz wprawiona w ruch, pozwala wygenerować nieskończoną liczbę tekstów, czego przykładem są liczne generatory tekstowe analogowe oraz cyfrowe. Do generatorów analogowych zalicza się na przykład pozornie tradycyjnie wydaną książkę *Sto tysięcy miliardów wierszy* Queneau.

Oulipo przygląda się temu etapowi pisania, w którym tekst jeszcze nie istnieje i poprzedzony jest pewną przemyślaną konstrukcją, planem, strukturą, „jakością obranych reguł”, w jak największym stopniu potencjalnych, czyli zdolnych do wytworzenia jak największej liczby (różnych) tekstów.

W przypadku *Wiersza smogowego* to użytkownik, decydując się na zainstalowanie utworu na swoim komputerze, dobrowolnie zgadza się na modyfikowanie stron internetowych, które przegląda. W przeciwieństwie jednak do wtyczki, którą można odinstalować, zanieczyszczenie powietrza pozostanie.

Niniejszy raport techniczny jest opisem maszynerii, według której działa potencjalna struktura *Wiersza smogowego*.

5 J. Jouet, *Z przymusami (ale i bez)*, tłum. J. Franczak, „Nowy Wiek” 2006, nr 10, s. 18.

5. Przymus smogowy - metoda pojedynczego cięcia zdań z prawdopodobieństwem warunkowym zależnym od zewnętrznych danych

W momencie przekroczenia norm stężenia PM 10 mechanizm wstrzykuje w otwartą stronę fragmenty treści zaciągniętych z wcześniej odwiedzonych przez użytkownika witryn. Metoda transformacji opiera się na połączeniu ze sobą dwóch ciągów tekstowych: zdania z aktualnie przeglądanej strony ze zdaniem archiwalnym znajdującym się w bazie tekstowej. Lokalizacja cięcia wyznaczana jest przez miejsce wspólne - sekwencję dwóch wyrazów, która występuje zarówno w pierwszym, jak i w drugim zdaniu.

Zdanie źródłowe:

Z każdym oddechem nowe drobinki pyłu [trafiają do] płuc, powodując zwiększenie zachorowalności na choroby układu oddechowego.

Zdanie archiwalne:

Wirusy, wykorzystując określone luki w niezaktualizowanych programach, [trafiają do] pamięci operacyjnej komputera, skąd nadpisują pliki z danymi.

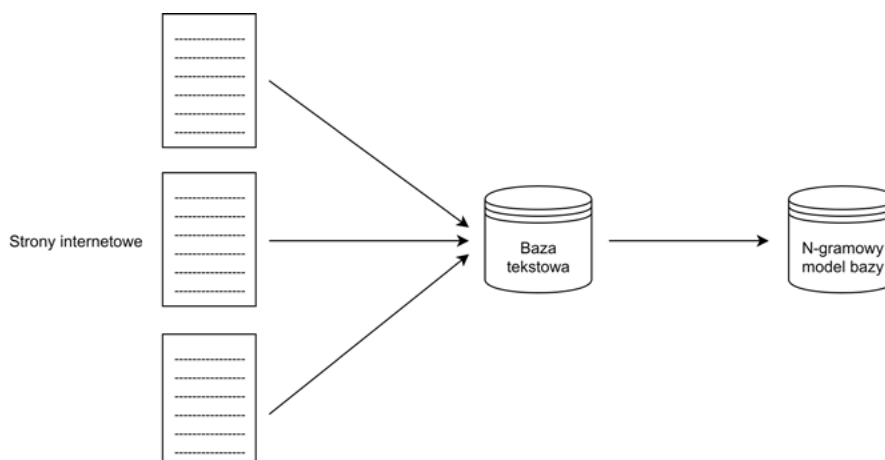
Zdanie wygenerowane:

Z każdym oddechem nowe drobinki pyłu [trafiają do] pamięci operacyjnej komputera, skąd nadpisują pliki z danymi.

Odzwierciedlenie zanieczyszczenia odbywa się w dwóch trybach: ilościowym i jakościowym. Poziom pyłu zawieszonego PM 10 w powietrzu odpowiada za liczbę zarażonych jednostek tekstowych na stronie, natomiast stężenie PM 2.5 wpływa na dobór zdań z korpusu.

5.1. Korpus

Wtyczka po zainstalowaniu towarzyszy użytkownikowi przy przeglądaniu stron internetowych i zapisuje w bazie danych zdania z odwiedzanych witryn. Po zgromadzeniu 30 tys. ciągów tekstowych aplikacja przygotowuje n-gramowy model bazy, w którym zlicza wystąpienia wszystkich dwuwyrzowych sekwencji (bigramów) w tekstach i przypisuje im następujące po nich słowa (następniki). Umieszczenie danych w nowej strukturze ułatwia dostęp do zawartości zmagazynowanych treści i pozwala na przeprowadzenie ich statystycznej analizy. Przy wykorzystaniu niewielkich zasobów pamięci operacyjnej sprawdzimy, czy w korpusie znajduje się określona grupa słów, oraz uzyskamy listę wszystkich wyrazów, które pojawiły się po szukanej sekwencji.



Rysunek 3. Schemat budowania korpusu

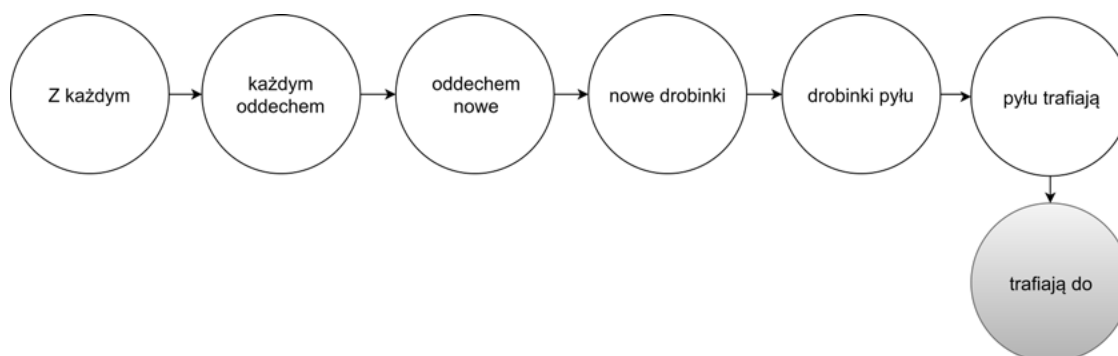
5.2. PM 10

Zależnie od stężenia pyłu zawieszonego PM 10 aplikacja modyfikuje proporcjonalną liczbę zdań na aktualnie przeglądanej stronie. Im większe zanieczyszczenie, tym więcej treści zostaje zniekształcone. Do wyznaczenia liczby elementów, które mają ulec zmianie, wykorzystano dolną (25 ug/m^3) i górną granicę (180 ug/m^3) wspólnego indeksu jakości powietrza dla odczytu jednogodzinnego. Przykładowo: przy stężeniu równym lub przekraczającym 180 ug/m^3 mechanizm zainfekuje wszystkie ciągi tekstowe znajdujące się na stronie, przy pomiarze wynoszącym 89 ug/m^3 : z szesnastu dostępnych zdań aplikacja przekonstruuje siedem z nich.

5.3. PM 2.5

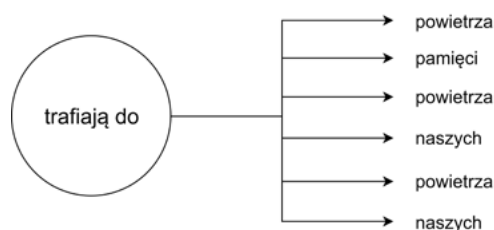
Metoda cięcia zdań oparta jest na prawdopodobieństwie warunkowym, czyli matematycznym modelu obliczającym prawdopodobieństwo zajścia zdarzenia A pod warunkiem zajścia zdarzenia B. Zdanie źródłowe pobrane z aktualnie otwartej strony zostanie zmodyfikowane, pod warunkiem że zawiera sekwencję dwóch wyrazów, która występuje w korpusie. W przypadku odnalezienia miejsca wspólnego następuje cięcie przetwarzanego ciągu tekstowego, mechanizm oblicza rozkład prawdopodobieństwa dla słów, które pojawiły się w korpusie po szukanej dwuwyrzowej sekwencji, a poziom pyłu zawieszonego PM 2.5 odpowiada za dobór treści, które zanieczyszczą zdanie źródłowe, wedle zasady: im większe skażenie powietrza, tym mniej prawdopodobne elementy są wybierane.

Krok 1. Przy przekroczeniu stężenia 25 ug/m^3 pyłu zawieszonego PM 10 aplikacja rozpoczyna właściwe działanie: pobiera zdanie źródłowe z aktualnie otwartej strony. Mechanizm iteruje po kolejnych dwuelementowych sekwencjach łańcucha i sprawdza, czy w n-gramowej bazie tekstowej znajduje się szukana grupa słów.



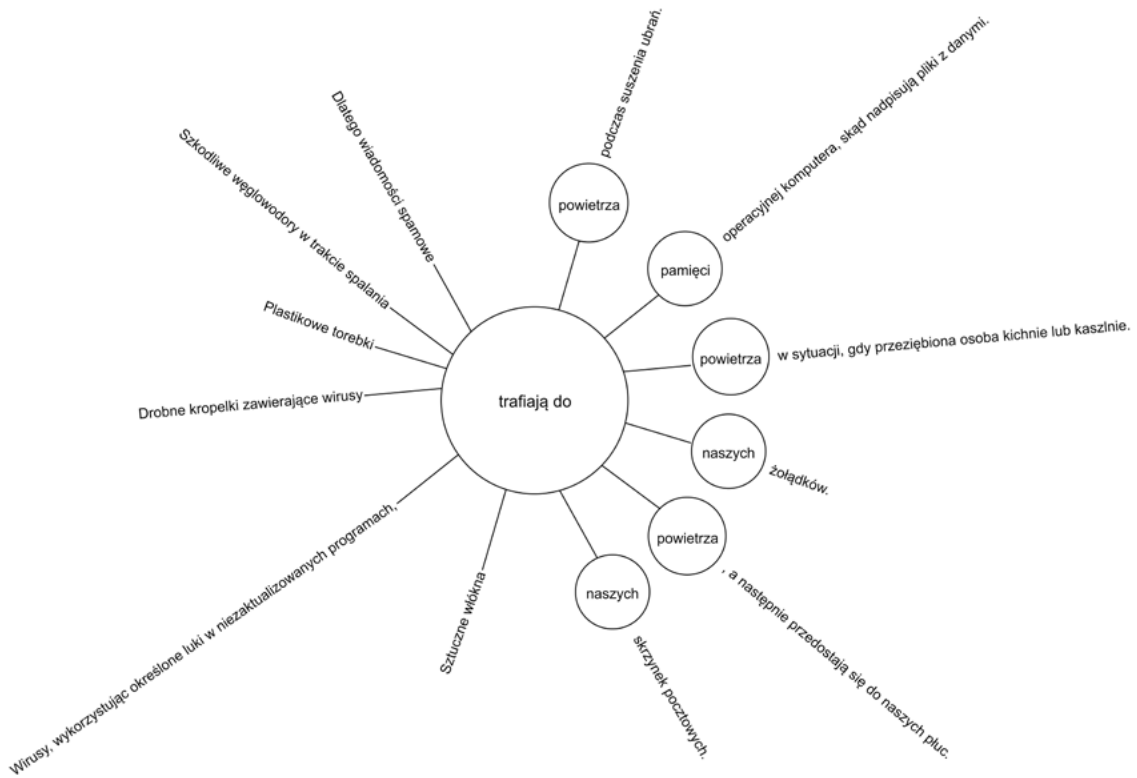
Rysunek 4. Zdanie źródłowe przedstawione w formie łańcucha bigramów

Krok 2. W przypadku odnalezienia miejsca wspólnego pętla kończy działanie - otrzymujemy listę wszystkich wyrazów (następników), które pojawiły się w n-gramowej bazie po szukanych wyrażeniu kluczowym (bigramie).



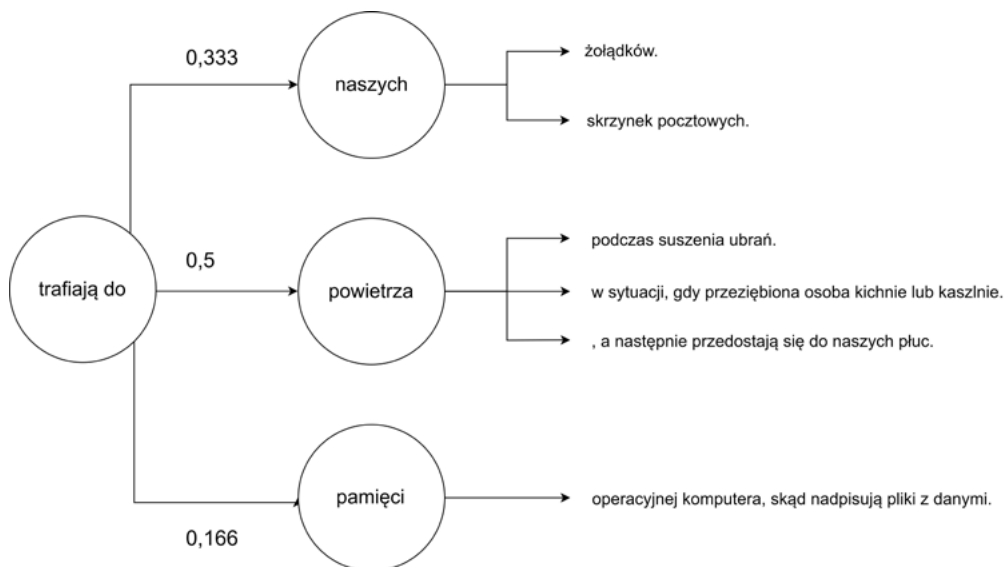
Rysunek 5. Lista następników bigramu „trafiają do”

Odzwierciedlają one zawartość bazy tekstowej:



Rysunek 6. Przykładowe zdania z bazy tekstowej zawierające bigram „trafiają do”

Krok 3. Dalej obliczany jest rozkład prawdopodobieństwa - każdemu następnikowi przyporządkowana jest procentowa wartość określająca prawdopodobieństwo wylosowania w niezależnym rzucie kostką. Przykładowo: prawdopodobieństwo wylosowania słowa „powietrza” wynosi 50 proc., gdyż w zbiorze sześćoelementowym występuje ono trzy razy.



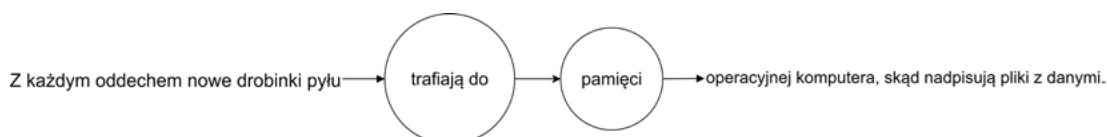
Rysunek 7. Rozkład prawdopodobieństwa dla bigramu „trafiają do”

Krok 4. W kolejnym kroku następuje rzutowanie dolnej (15 ug/m^3) i górnej (110 ug/m^3) granicy wspólnego indeksu jakości powietrza dla pyłu zawieszonego PM 2.5 na rozkład prawdopodobieństwa. Wartościom rozkładu przypisywane są zakresy zanieczyszczenia

wedle zasady: im większe prawdopodobieństwo wystąpienia wyrazu, tym niższe stężenie szkodliwej substancji jest mu przydzielane.

15-46 uq/m³ [„powietrza”, „powietrza”, „powietrza”]
 47-78 uq/m³ [„naszych”, „naszych”]
 79-110 uq/m³ [„pamięci”]

Krok 5. Aplikacja sprawdza aktualny pomiar pyłu zawieszzonego PM 2.5 i losuje następnik z przynależnego mu zakresu. Pobiera zdanie z bazy tekstowej, w którym wystąpiło wybrane słowo, ucina je w miejscu wspólnym i zestawia ze zdaniem źródłowym. Przykładowo: przy stężeniu 85 uq/m³ mechanizm wybrał ze zbioru wyraz „pamięci” i wygenerował przy jego użyciu nowy ciąg tekstowy.



Rysunek 8. Zdanie wygenerowane przy wykorzystaniu przymusu smogowego

Krok 6. Na stronie internetowej zdanie źródłowe podmieniane jest na wygenerowany ciąg tekstowy.

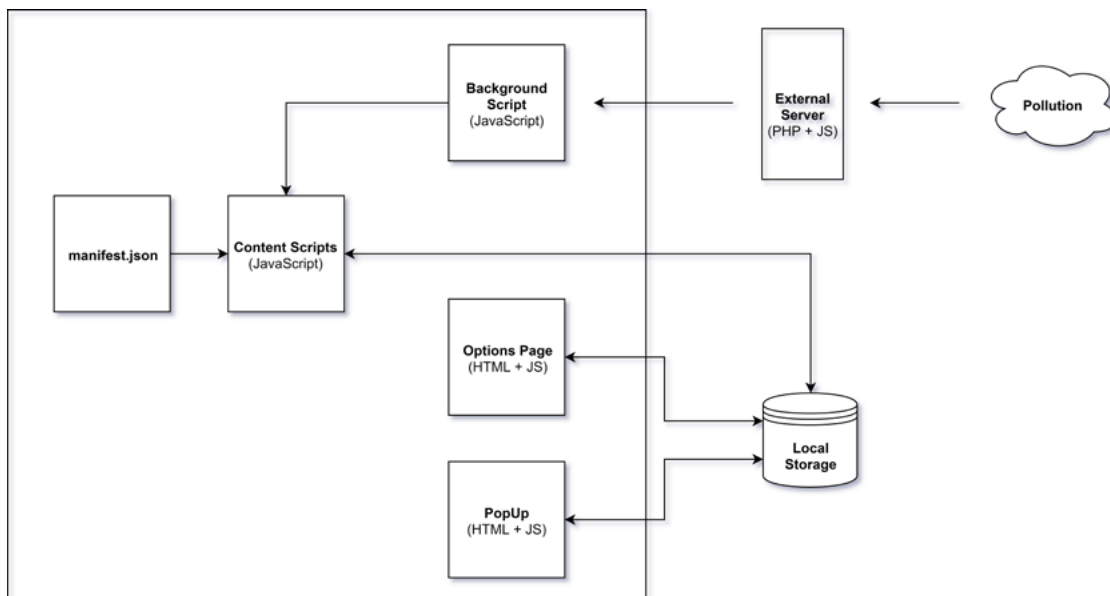
6. Historia powstania, platforma i ograniczenia

6.1. Historia powstania

Historia powstania utworu zawiera się w dwudziestu dziewięciu wersjach, które wyłoniły się w okresie między czerwcem 2016 roku a styczniem 2019 roku. Ostateczna odsłona aplikacji powstała w trakcie pisania niniejszego raportu technicznego. Utwór był tworzony początkowo chaotycznie, intuicyjnie, bez planu i harmonogramu. W pierwszym etapie opracowano moduł pobierający dane o zanieczyszczeniu powietrza z zewnętrznego serwera oraz mapowanie skażenia na tekst. W drugim - utwór z generatora tekstowego odtwarzanego na stronie HTML przeistoczył się we wtyczkę do przeglądarki. Trzeci etap upłynął na poszukiwaniach odpowiedniej metody transformacji treści - zajął on najwięcej czasu. W czwartym stadium przygotowano stronę opcji oraz panel zarządzający utworem (pomysł Jacka Olczyka). Kolejna faza poświęcona została optymalizacji kodu. Na końcu wysłano utwór wybranym beta testerom, by uruchomili generator na swoich komputerach.

6.2. Platforma

Aplikacja powstała na przeglądarkę Google Chrome i została napisana w języku JavaScript. Umiejscowiony na zewnętrznym serwerze parser do analizy, uśredniania i pobierania danych o zanieczyszczeniu powietrza powstał w językach PHP i JavaScript. Struktura utworu składa się z czterech elementów: *manifest.json*, *content scripts*, *background script* oraz strony z opcjami i popupem. W dokumencie *manifest.json* zdefiniowane są podstawowe informacje o programie, takie jak: nazwa, wersja, lista plików oraz pozwolenia na odczyt i zmianę wszystkich treści na odwiedzanych stronach. *Content scripts* odpowiadają za gromadzenie zdań do korpusu oraz generowanie tekstu. *Background script* łączy się z zewnętrznym serwerem i importuje wyekstrahowane dane o skażeniu powietrza z wybranego miasta. Na stronie z opcjami użytkownik ma do wyboru miasto oraz rodzaj zanieczyszczenia (aktualne lub archiwalne). Popup, którego ikonka znajduje się w prawym górnym rogu przeglądarki, informuje użytkownika o aktualnym stężeniu pyłu PM 10 i PM 2.5 oraz raportuje, ile zdań zostało zmodyfikowanych na aktualnie otwartej stronie.



Rysunek 9. Budowa utworu narzucona przez platformę

6.3. Ograniczenia

W trakcie pracy nad aplikacją uwzględniono niżej wymienione ograniczenia.

6.3.1. Ekstrakcja zdań z kodu źródłowego

Przestrzenią działania aplikacji są strony internetowe przedstawione w postaci kodu źródłowego. W kodzie, oprócz widocznych na ekranie monitora elementów graficznych i tekstowych, znajdziemy wiele bloków tekstu wykonywalnego: znaczniki HTML, skrypty, znaki specjalne, elementy stylu - odpowiedzialne za sposób wyświetlenia zawartości strony oraz jej zachowanie. Przetwarzanie i modyfikowanie takiego korpusu stwarza wiele technicznych ograniczeń. Nadpisanie nieodpowiedniego fragmentu kodu może zburzyć konstrukcję strony i spowodować, że stanie się ona niefunkcjonalna.

6.3.2. Czas ładowania się strony z uruchomioną wtyczką

Podczas projektowania, wdrażania i testowania poszczególnych wyników pracy zwrócono uwagę na przygotowanie takich reguł transformacji, które nie będą znacząco obciążały pamięci operacyjnej przeglądarki ani nie przedłużą drastycznie czasu ładowania się witryny. Internauta oczekujący na załadowanie się strony kilka sekund wyłączy wtyczkę, gdyż będzie mu przeszkadzała w codziennym korzystaniu z internetu.

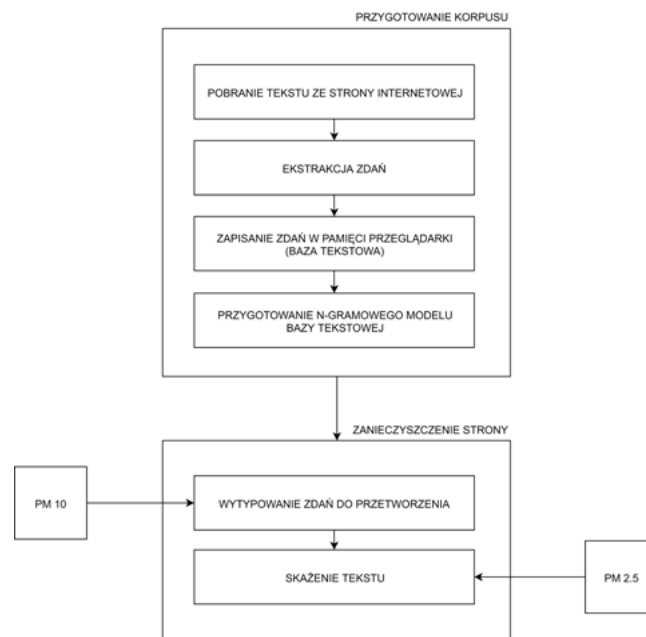
6.3.3. Waga wtyczki

Z każdym uruchomieniem strony przeglądarka wczytuje zawartość całego kodu źródłowego utworu. Im więcej linijek tekstu wykonywalnego, tym dłużej jest on ładowany do pamięci. Jednym z ważnych etapów kończących programowanie aplikacji była optymalizacja i zmniejszenie rozmiaru kodu.

6.3.4. Pamięć podręczna przeglądarki

Program zapisuje dane tekstowe pobrane z przeglądanych przez użytkownika stron w pamięci podręcznej przeglądarki, tzw. *localStorage*. W utworze skorzystano z opcji przechowywania bazy danych na dysku użytkownika, a nie synchronizowania jej z kontem użytkownika, co zwiększyło pamięć z 102 400 do 5 242 880 bajtów

7. Mechanizm utworu

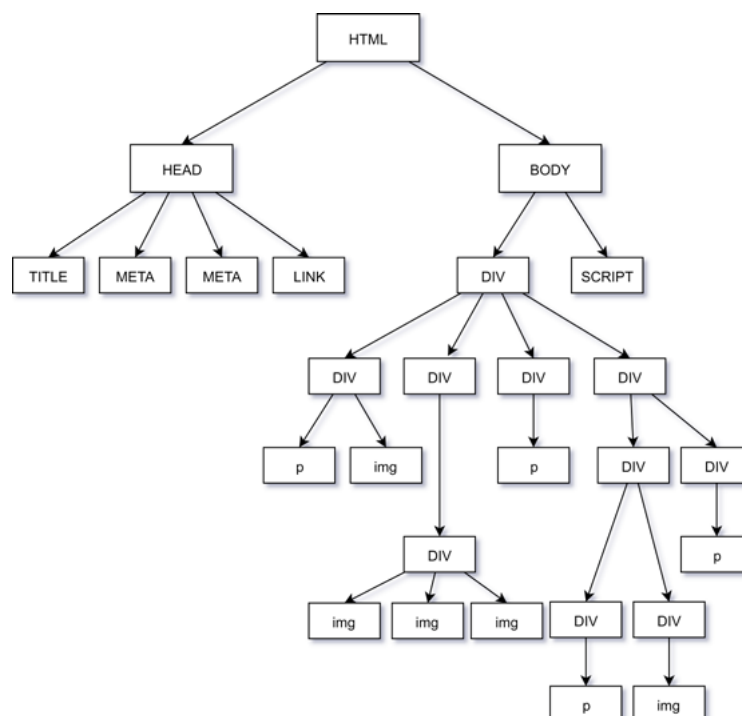


Rysunek 10. Graf przedstawiający mechanizm utworu

7.1. Budowanie korpusu

7.1.1. Pobranie tekstu ze strony internetowej

Przełęczarka po wczytaniu kodu źródłowego tworzy obiektowy model dokumentu HTML (Document Object Model, DOM), który pozwala na bezpośredni dostęp do struktury witryny oraz łatwą modyfikację jej elementów. Tym, co widzimy na ekranie komputera, jest zbiór: kontenerów (<div>), zdjęć (), odnośników (<a>) i paragrafów (<p>) umiejscowionych w strukturze drzewa.



Rysunek 11. Obiektowy model przykładowej strony internetowej

Mechanizm wtyczki przeczesuje drzewo dokumentu HTML i sprawdza, czy iterowany węzeł zawiera tekst. Jeżeli warunek zostaje spełniony, wówczas zawartość elementu kopiowana jest do nowego obiektu. Dostęp do wybranych gałęzi strony uzyskano za pomocą metody `document.getElementsByTagName(tagName)`, która pobiera kolekcję elementów po klasie.

Do rozpoznania ciągów tekstowych wykorzystano własność modelu DOM - `nodeType` (w przypadku węzła tekstowego równa się ona 3).

7.1.2. Ekstrakcja zdań

Zdrapane skrawki tekstów dzielone są na zdania przy wykorzystaniu trzech delimiterów: kropki, pytajnika oraz wykrzyknika. W przeciwieństwie do tradycyjnych korpusów literackich zawartość strony internetowej przysparza dużych trudności przy próbie ekstrakcji, gdyż mamy do czynienia z tekstem ciągłym, w którym zdania poprzedzielane są nie tylko znakami stylu (kursywa, pogrubienie, nagłówki), ale i znakami nowej linii, znakami specjalnymi czy fragmentami kodu. W wielu wypadkach zdanie zaczyna się w jednym kontenerze a kończy w kolejnym.

```
0: replace( /(^|\s)client-nojs(\s|$)/, "$1client-js$2" );
1: set({"wgCanonicalNamespace":"","wgCanonicalSpecial del": "wikitext",
"wgSeparatorTransformTable": [",\t
2: ", "\t", ], "wgDigitTransformTable":["", ""], "wgDefau...raising", "fundraising"], "wgC
coordinates":{"lat":50
3: Sport w Krakowie
4: Kraków jest drugim co do wielkości, po Warszawie, rynkiem nowoczesnej
powierzchni biurowej,
5: (ponad milion metrów kwadratowych powierzchni biurowej)
6: ze względu na swoją florę, faunę, geomorfologię i krajobraz
7: Krakowskie powietrze zanieczyszczone jest przede wszystkim pyłem
8: Pomniki w Krakowie
9: W zalanych dzielnicach komunikacja odbywała się łodziami i
10: Zespół Opery występował wcześniej w budynku Teatru im
11: Korzystasz z Wikipedii tylko na własną odpowiedzialność
```

Przykładowe ciągi tekstowe wydzielone za pomocą delimiterów (!?) ze strony Krakowa na portalu Wikipedia⁶

Celem ekstrakcji jest wyodrębnienie poprawnych gramatycznie zdań, które posłużą do zbudowania n-gramowego modelu korpusu. Po licznych testach przyjęto takie zasady, które eliminują: linijki kodu, urwane teksty, zlepki słów, ciągi tekstowe, które nie rozpoczynają się od litery lub zawierają specjalne znaki. Na potrzeby tego utworu zdefiniowano zdanie normalne spełniające warunek wyrażenia regularnego:

```
^[A-Z][a-zA-Z0-1 ]{50, 1500}([^\s]\w)$
gdzie:
^[A-Z]          zdanie rozpoczyna się dużą literą,
{50, 5000}      zdanie nie zawiera mniej niż 50 a więcej niż 1500 znaków
                 (licząc ze spacjami)
([^\s]\w)$ -    zdanie nie kończy się pojedynczą literą
```

Dodatkowo z korpusu usuwane są zdania zawierające znaki specjalne oraz kończące się na skrótowcach: „proc.”, „itd.”, „np.”

7.1.3. Zapisanie zdań w pamięci przeglądarki

Wyselekcjonowane ciągi tekstowe zapisywane są w „pamięci przeglądarki”. Wykorzystano do tego magazyn *Web Storage* i jego atrybut *Local Storage*, który przechowuje dane na dysku twardym użytkownika. Obiekty umieszczone w bazie pozostają w pamięci nawet po zamknięciu przeglądarki. W utworze zrezygnowano z synchronizowania danych z kontem użytkownika w chmurze - co zwiększyło obszar pamięci z 102 400 do 5 242 880 bajtów.

⁶ <https://pl.wikipedia.org/wiki/Krak%C3%B3w> [dostęp: 23.11.2018].

Zaletą *Local Storage*, w odróżnieniu od powszechnie znanych *cookies*, jest to, że treści umieszczone w ich przestrzeni nie są automatycznie pobierane przy odświeżeniu strony tylko czekają na wywołanie przez aplikację.

7.1.4. Przygotowanie N-gramowego modelu bazy tekstowej

Po zapisaniu w pamięci przeglądarki minimum 30 tys. ciągów tekstowych spełniających warunek $\{[A-Z][a-zA-Z0-1]\}^{50,1500}\{[\s\w]\}^*$ tworzona jest tablica oparta na n-gramowym modelu języka, który bada sekwencję następujących po sobie jednostek leksykalnych i pozwala obliczyć częstotliwość występowania po sobie n-gramów w korpusie. W pracy wykorzystano model 2-gramowy (bigram)⁷ oparty na ciągach tekstowych składających się z dwóch członów.

Mechanizm przetwarza zdania zgromadzone w historii przeglądania iterując po dwuelementowych sekwencjach wyrazów i odnotowując słowa, które po nich występują. Przykładowo, dla ciągu:

Jeżeli nic przez chwilę nie mówiłam, to nie dlatego, że nie miałam wiele do powiedzenia.

Stworzona została tablica, gdzie kluczem jest zbiór dwuelementowy, a wartością słowo, które po nim występuje.

BIGRAM	NASTĘPNIK
Jeżeli nic	przez
nic przez	chwilę
przez chwilę	nie
chwilę nie	mówiłam
nie mówiłam	,
mówiłam ,	to
, to	nie
to nie	dlatego
nie dlatego	,
dlatego ,	że
, że	nie
że nie	miałam
nie miałam	wiele
miałam wiele	do
wiele do	powiedzenia

Przy budowaniu n-gramowego modelu bazy tekstowej:

- zrezygnowano z oznaczania początku zdania, w zamian przygotowano oddzielną tablicę zawierającą pierwsze bigramy wszystkich ciągów tekstowych,
- nie uwzględniono najczęściej występujących w języku polskim łączników zdań współrzędnych, podrzędnych i nadrzędnych, wskaźników, zaimków oraz spójników, czyli jednostek językowych należących do grupy słów *stoplist*.
- znaki interpunkcyjne traktowane są jako oddzielne wyrazy,
- każdej wartości przypisano numer zdania w bazie tekstowej, z którego pochodzi.

7.2. Zanieczyszczenie strony

7.2.1. Wytypowanie zdań do przetworzenia

Zależnie od stężenia pyłu zawieszonego PM 10 zniekształceniu ulega proporcjonalna liczba wybranych elementów witryny. Do obliczenia ile segmentów tekstowych ma ulec deformacji

⁷ W trakcie testowania utworu sprawdzono działanie modelu 1-gramowego (unigram), badającego sekwencje jednostek jednowyrazowych, ale ze względu na liczne błędy gramatyczne, które pojawiały się przy wykorzystaniu tej metody, zrezygnowano z tego rozwiązania.

wykorzystana została funkcja mapowania⁸. Odzworowuje ona wartość aktualnego skażenia z zakresu wejściowego (którego górną granicę wyznacza stężenie substancji zagrażające życiu człowieka: 180µg/m³) na zakres wyjściowy o wartości końcowej równej liczbie wszystkich segmentów tekstowych występujących na przeglądanej przez użytkownika stronie.

$$x = \minElem + \frac{(\maxElem - \minElem)(value - lowPM10)}{(highPM10 - lowPM10)}$$

gdzie:

value - aktualny pomiar pyłu zawieszonego PM 10

lowPM10 - dolna granica stężenia pyłu zawieszonego PM 10 (wynosi zawsze zero)

highPM10 - stężenie pyłu zawieszonego PM 10 zagrażającego zdrowiu człowieka (180 µg/m³)

minElem - minimalna liczba elementów do modyfikacji na stronie (wynosi zawsze zero)

maxElem - liczba wszystkich zdań na stronie spełniających warunek

`^[A-Z][a-zA-Z0-1]{50,1500}([^\s]\w)$`

Przykładowo, pewnego dnia w trakcie sezonu grzewczego poziom pyłu zawieszonego PM10 wyniósł w Krakowie 120 µg/m³. Na uruchomionej tego dnia stronie zawierającej dwadzieścia pięć zdań spełniających warunek `^[A-Z][a-zA-Z0-1]{50,1500}([^\s]\w)$`, modyfikacji ulegnie szesnaście z nich.

7.2.2. Skażenie tekstu

Do generowania nowych ciągów tekstowych wykorzystano metodę cięć w miejscach wspólnych, miejscach łączących zdanie źródłowe, pobrane z aktualnie otwartej strony, ze zdaniem z bazy tekstowej.

$$z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8 + [z_9, z_{10}] + b_6, b_7, b_8, b_9, b_{10}, b_{11}, b_{12}, b_{13}, b_{14}$$

gdzie:

z_x - wyrazy zdania źródłowego

b_x - wyrazy zdania z bazy danych

[] - oznaczenie miejsca wspólnego

Metoda tworzenia nowych tekstów posługuje się poniższymi regułami:

1. Lokalizacja cięcia zdania źródłowego zależy od wystąpienia w nim sekwencji dwóch wyrazów, które znajdują się w korpusie.
2. Cięcie wykonywane jest w pierwszym odnalezionym miejscu wspólnym zdania źródłowego ze zdaniem z bazy tekstowej.
3. Proces przeszukania zdania źródłowego przeprowadzany jest od trzeciego do ostatniego elementu ciągu. Rozpoczęcie kwerendy od trzeciego wyrazu eliminuje łączenia ciągów tekstowych na początku zdania.
4. W przypadku nieodnalezienia miejsca wspólnego w zdaniu źródłowym mechanizm losuje zdanie z bazy tekstowej i szuka miejsc wspólnych z innymi zdaniami zgromadzonymi w pamięci przeglądarki.
5. Zdanie źródłowe z aktualnie otwartej strony, które posłużyło jako początek nowego ciągu tekstowego, jest zapisywane w bazie danych, natomiast zdanie pobrane z pamięci przeglądarki, które wykorzystano jako drugą część ciągu tekstowego, zostaje z niej usunięte.

W podpunktach przedstawione zostaną etapy konstruowania nowego tekstu z perspektywy kodu programu.

⁸ Wzór odzwierciedlenia jednego zakresu na drugi ze strony: https://rosettacode.org/wiki/Map_range [dostęp: 23.11.2018].

1. Mechanizm pobiera zdanie źródłowe z otwartej strony internetowej.

Węgiel, ropa naftowa oraz gaz ziemny są stabilnymi źródłami energii i stanowią podstawowy surowiec dla produkcji przemysłowej w większości państw Europy.

2. Zachowujemy pierwsze dwa wyrazy ciągu.

Węgiel ,

3. Iterujemy po elementach zdania, sprawdzając, czy w zdaniach zgromadzonych w pamięci przeglądarki nie znajduje się szukany bigram. W tym kroku wykorzystano przygotowany wcześniej n-gramowy model bazy historycznej.

ropa naftowa	false
naftowa oraz	false
oraz gaz	false
gaz ziemny	false
ziemny są	false
są stabilnymi	false
stabilnymi źródłami	false
źródłami energii	false
energii i	false
i stanowią	true
	break

4. Z każdą iteracją do końca ciągu dodajemy kolejny wyraz.

Węgiel, ropa naftowa
 Węgiel, ropa naftowa oraz
 Węgiel, ropa naftowa oraz gaz
 Węgiel, ropa naftowa oraz gaz ziemny
 Węgiel, ropa naftowa oraz gaz ziemny są
 Węgiel, ropa naftowa oraz gaz ziemny są stabilnymi
 Węgiel, ropa naftowa oraz gaz ziemny są stabilnymi źródłami
 Węgiel, ropa naftowa oraz gaz ziemny są stabilnymi źródłami energii
 Węgiel, ropa naftowa oraz gaz ziemny są stabilnymi źródłami energii i
 Węgiel, ropa naftowa oraz gaz ziemny są stabilnymi źródłami energii i stanowią

5. Po znalezieniu słowa kotwicy, wyrazu, który występuje zarówno w zdaniu źródłowym, jak i w bazie tekstowej, pętla kończy działanie. Otrzymujemy listę wszystkich następników, które pojawiły się po szukanym ciągu tekstowym. W przypadku bigramu „i stanowią” otrzymujemy:

```
i stanowią: Array(10)
0: [„czynnik”]
1: [„czynnik”]
2: [„czynnik”]
3: [„jeden”]
4: [„jeden”]
5: [„kompleksowa”]
6: [„tykająca”]
7: [„jedynie”]
8: [„wskazówkę”]
9: [„właśnie”]
```

Obliczamy rozkład prawdopodobieństwa zgromadzonych słów, dzieląc ich wystąpienia przez liczbę wszystkich elementów.

„czynnik”	(0.3)
„jeden”	(0.2)

```

„kompleksową”      (0.1)
„tykająca”         (0.1)
„jedynie”          (0.1)
„wskazówkę”       (0.1)
„właśnie”          (0.1)

```

Poszczególne rozkłady rozpisujemy jako osobne zbiory, każdemu elementowi zbioru przypisujemy kolejną liczbę.

```

[
  [1, 2, 3]
  [4, 5]
  [6]
  [7]
  [8]
  [9]
  [10]
]

```

Zbiory, które mają taką samą liczbę elementów, umieszczamy w jednej tablicy:

```

[
  [1, 2, 3]
  [4, 5]
  [6, 7, 8, 9, 10]
]

```

W kolejnym kroku korzystamy z dolnych i górnych granic wspólnego indeksu jakości powietrza dla pyłu zawieszonego PM 2.5. Zgodnie z indeksem: dolny zakres to 15 uq/m³, górny: 110 uq/m³. Iterujemy po wszystkich zbiorach i obliczamy zakres zanieczyszczenia dla każdego zbioru, pobierając wartość ostatniego elementu tablicy.

```

function map_range (value, low1, high1, low2, high2) {
  var wynik = low2 + (high2 - low2) * (value - low1) / (high1 - low1);
  return Math.floor(wynik);
}

```

gdzie:

```

value - wartość ostatniej liczby z iterowanego zbioru
low1 - minimalna liczba elementów we wszystkich zbiorach
high1 - liczba wszystkich elementów we wszystkich zbiorach
low2 - dolna granica zanieczyszczenia powietrza, w przypadku PM 2.5 będzie to
zawsze 15
high2 - górna granica zanieczyszczenia powietrza, w przypadku PM2.5. będzie to
zawsze 110

```

Wyniki działania funkcji dla kolejnych zbiorów przedstawiają się następująco:

```

map_range(3, 1, 10, 15, 110) = 36 uq/m3
map_range(5, 1, 10, 15, 110) = 62 uq/m3
map_range(10, 1, 10, 15, 110) = 110 uq/m3

```

Otrzymujemy zakres zanieczyszczenia dla każdego zbioru:

```

15-36 uq/m3  [„czynnik”, „czynnik”, „czynnik”]
37-62 uq/m3  [„jeden”, „jeden”]
63-110 uq/m3 [„kompleksową”, „tykająca”, „jedynie”, „wskazówkę”, „właśnie”]

```

Następnie sprawdzamy aktualne zanieczyszczenie i losujemy jeden wyraz ze zbioru, który odpowiada stężeniu pyłu PM 2.5 w powietrzu. W momencie pisania tego tekstu pomiar wskazywał 76 uq/m³, a mechanizm wybrał wyraz „tykającą”.

6. Po wylosowaniu następnika pobieramy z bazy tekstowej zawartość całego ciągu tekstowego, w którym wystąpił.

Gazy cieplarniane są główną przyczyną globalnych zagrożeń [i stanowią] {tykającą} bombę, która zagraża nie tylko środowisku naturalnemu, ale i nam samym.

gdzie:

[] - bigram (n) - wyraz-kotwica, miejsce cięcia

{ } - następnik (n+1)

7. Ze zdania z bazy tekstowej odcinamy fragment: od początku ciągu do wyrazu kotwicy włącznie.

tykającą bombę, która zagraża nie tylko środowisku naturalnemu, ale i nam samym.

8. Następnie łączona jest pierwsza część zdania źródłowego z drugą częścią zdania z bazy tekstowej.

(Węgiel, ropa naftowa oraz gaz ziemny są stabilnymi źródłami energii) [i stanowią {tykającą bombę, która zagraża nie tylko środowisku naturalnemu, ale i nam samym}

gdzie:

() - fragment zdania źródłowego

[] - bigram (n) - wyraz kotwica, miejsce cięcia

{ } - fragment zdania z bazy tekstowej

8. Konteksty kulturowe

Smog to nie tylko jedno z najważniejszych zagadnień dla ekologów, polityków czy mieszkańców miast i wsi, ale również inspiracja dla artystów.

8.1. Smog w sztuce i smog jako problem społeczny

Zanieczyszczenie powietrza odgrywało istotną rolę w sztukach plastycznych już w XIX wieku. Na wielu obrazach malarzy tworzących w Londynie można odnaleźć słynną londyńską mgłę, która - jak już dziś wiadomo - była przekleństwem dla zdrowia, zwanym dziś smogiem (od ang. *smoke* - dym i *fog* - mgła). Był to czas zaawansowanego rozwoju przemysłowego opartego na spalaniu węgla. Słynna londyńska „mgła” jest nieodzowna we wszystkich ekranizacjach przygód Sherlocka Holmesa i powieści Dickensa. Niektórzy malarze wychodzący w plener narzekali, jak Luke Fildes odnotowujący w październiku 1880 roku: „Znosiliśmy, i wciąż musimy znosić najobrzydliwiej ciemną i beznadziejną zimę (...) mieliśmy nieprzerwanie mgłę przez pięć kolejnych dni w zeszłym tygodniu. Jest zbyt ciemno, by malować”⁹. Z kolei francuski malarz, twórca impresjonizmu Claude Monet był tą tajemniczą mgłą zafascynowany i w swoich wspomnieniach odnotowywał: „Kiedy wstałem, byłem przerażony, ujrzawszy, że nie ma mgły, ani śladu. Byłem w rozpacz (...). Lecz wtem, stopniowo, zaczęto rozpalać paleniska, i dym i mgła wróciły”. Ta dziwna mgła ustępowała najczęściej w niedziele, gdy kominy kopalni i fabryk przestawały dymić: „Jakim ponurym dniem jest przeklęta angielska niedziela. (...) Wszystko jest martwe, żadnych pociągów, żadnego dymu, nic, co by mnie inspirowało”¹⁰ - odnotowywał wówczas w pamiętniku. Brakuje jednak jakiegokolwiek zdania o szkodliwości tej „mgły”, tak twórczo inspirującej francuskiego impresjonistę.

Smog inspirował także pisarzy od czasów rewolucji przemysłowej, co udowodnił na początku XX wieku T.S. Eliot w *Pieśni miłosnej J. Alfreda Prufrocka*:

Żółta mgła, co ociera się grzbietem o szyby,
 Żółty dym, co ociera się mordą o szyby,
 Wsuwa język w zakątki wieczoru,
 Pełza po mokrym bruku i bajorach,

9 J. Jędrak, *Smog w sztuce*, <https://smoglab.pl/smog-w-sztuce> [dostęp: 17.10.2018].

10 Tamże.

Sadzę kominów zbiera na odwłoku,
Wciska się pod balkon, podrywa do skoku,

A widząc, że cicha noc październikowa,
Zwija się wokół domu i zasypia.

I to zaiste będzie czas
Dla żółtych dymów rozsnutych wzdłuż ulic.

Nie zadawano sobie jednak pytań o wpływ tej „żółtej mgły” na zdrowie. Ten trend w sztuce rozpoczął się dopiero z końcem XX wieku i zaczął mocniej wybrzmiewać z początkiem wieku XXI. Aktywizm uprawiany przez wielu artystów i zwrócenie uwagi na problem środowiska, w którym żyjemy, i powietrza, którym oddychamy, sprawia, że sztuka zaczyna przybierać także aspekt edukacyjny. Można przytoczyć kilka przykładów z ostatnich lat.

W 2011 roku artyści Zack Danfeld i Cat Kramer zrealizowali projekt Smog Tasting (w ramach artystycznego think-tanku Center for Genomic Gastronomy) polegający na ubijaniu piany z białek w różnych dzielnicach Bangalora (aglomeracja w południowych Indiach), gdzie odnotowano szczególnie duże zanieczyszczenie powietrza. Piana z białek, stanowiąca w 90 proc. powietrze, stanowiła doskonały materiał do przygotowania bezy, co umożliwiło zilustrowanie, w jakim stopniu jakość żywności zależy od otoczenia, w którym jest produkowana, w tym stężenia smogu. Bezy te były następnie wysyłane do polityków, urzędników i innych ważnych dla sprawy smogu decydentów.

Joanna Jeśmian w swoim eksperckim komentarzu zwraca uwagę także na inny artystyczny projekt. W 2015 roku Nicola Twilley - autorka bloga o jedzeniu Edible Geography - we współpracy ze wspomnianymi Danfeldem i Kramerem stworzyli projekt „Aeroir”. Jego tytuł nawiązuje do terminu „terroir”, stosowanego na przykład przy produkcji win czy serów, który określa splot warunków klimatycznych i geologicznych nadających produktom spożywczym specyficzny smak. Jak pisze Jeśmian:

„Aeroir” to przypominająca wózek z hot dogami przewoźna piekarnia, która oferuje możliwość skosztowania bez w czterech charakterystycznych smakach. W Stanach Zjednoczonych występują bowiem cztery rodzaje smogu: tzw. smog londyński wywołany paleniem węgłem i emisją dwutlenku siarki; smog z Atlanty, gdzie do zanieczyszczeń tworzonych przez człowieka dochodzą opary organiczne, terpenoidy (pochodzące z drzew iglastych i innych odpadów organicznych); smog z doliny kalifornijskiej nasycony amoniakiem emitowanym przez farmy zwierzęce oraz smog typu Los Angeles, fotochemiczny, który powstaje w upalne dni. Aby bezy smakowały, pachniały i składały się z odpowiednich składników, artyści we współpracy z naukowcami stworzyli niewielką komorę smogową, w której tworzyli swoje „przepisy” z odpowiednich związków chemicznych¹¹.

Smog to bardzo poważny problem także w Chinach. Performer Wang Renzheng postanowił przemierzyć Pekin, jak pisze Jeśmian:

(...) przez 100 dni w towarzystwie odkurzacza przemysłowego. W ten sposób zebrał około 100 gramów pyłu, który następnie dodał do odpowiedniej mikstury i wypalił cegłę. Projekt miał uświadomić ludziom czym oddychają, a cegła posłużyła do budowy fabryki w mieście Tangshan. Rozenheng chciał by w ten sposób cegła symbolicznie zniknęła, jak kropla wody w oceanie¹².

Niemiecka artystka Beatriz Da Costa wraz z grupą przyjaciół stworzyła z kolei „The Pigeon Blog”, dzięki któremu mieszkańcy Kalifornii mogli na żywo śledzić dane dotyczące zanieczyszczenia powietrza. Było to możliwe dzięki miniaturowemu i lekkiemu czujnikowi zanie-

11 J. Jeśmian, *Państwo i artyści kontra smog*, <https://www.swps.pl/centrum-prasowe/archiwum-centrum-prasowego/279-komentarz-ekspercki/3887-panstwo-i-artysci-kontra-smog> [dostęp: 12.10.2018].

12 Cyt. za: tamże.

czyszczeń przyczepionemu do szyi gołębia. Następnie we współpracy z hodowcą gołębi urzędnika przymocowano do szyi stada ptaków, które najczęściej przebywają dokładnie w tych samych miejscach, co ludzie.

Jeśmian wspomina o jeszcze jednym projekcie - holenderskiego artysty Daan Roosegaarde „Smog Free Project”. W praktyce jest to:

(...) ogromna wieża oczyszczająca powietrze, którą wybudował wraz z zespołem w Rotterdamie za 113 tysięcy funtów zebranych na Kickstarterze. Projektem natychmiast zainteresowali się Chińczycy, którzy zlecili stworzenie podobnej wieży w Pekinie. Siedmiometrowa budowla, działająca podobnie do domowych urządzeń tego typu, oczyszcza około 30 000 metrów sześciennych powietrza z pomocą wolnej od ozonu technologii jonowej, a w dodatku pełni rolę sztuki w przestrzeni miejskiej ze względu na swój wyjątkowy design. Wiedząc, że większość wylapywanych przez filtry materiału to węgiel, Roosegaarde postanowił przetworzyć go w biżuterię. Kupując pierścionek lub spinki do mankietów możemy przekazać 1000 metrów sześciennych czystego powietrza dla mieszkańców miast, w których zlokalizowane są wieże. Artysta tak mocno zaangażował się w działania na rzecz ochrony środowiska, że biorąc udział w spotkaniu organizowanym przez Organizację Narodów Zjednoczonych zaproponował wprowadzenie nowego prawa człowieka - prawa do *schoonheid*, co oznacza piękno i czystość¹³.

8.2. Smog w kulturze polskiej

Smog w sztuce i literaturze w Polsce jest zjawiskiem nowym, chociaż w środowisku obecny jest od dziesięcioleci. Szczególnie mocno dokuczliwy jest dla mieszkańców Krakowa po wybudowaniu kombinatu hutniczego Nowa Huta, a także przez tradycyjny sposób opalania większości kamienic i domów jednorodzinnych za pomocą węgla - najczęstsze źródło niskiej emisji. Zwiększony ruch samochodowy w mieście stanowi tylko niewielką część tego źródła, co udowodniły liczne badania i pomiary wykazujące problem natężenia smogu głównie w okresie grzewczym.

Polskich kontekstów dla *Wiersza smogowego* nie ma w Polsce zbyt wiele, choć warto przypomnieć refleksję Tadeusza Peipera sprzed blisko stulecia zawartą w artykule *Miasto. Masa. Maszyna*. Dziś niektóre wypowiedzi Peipera należałoby czytać w kontekście nie tylko modnych „izmów”, ale i ochrony środowiska, szczególnie gdy proponował na wzór wodociągów wybudować „tlenociągi, które będą sprowadzały tlen z warstw atmosferycznych unoszących się wysoko nad domami albo też z odległych lasów zamiejskich”¹⁴. Osobnym zagadnieniem w sztuce i literaturze jest gloryfikacja dymiących kominów Nowej Huty w poezji i prozie okresu socrealizmu, a także rozwoju przemysłu ciężkiego w okresie PRL.

Obecnie, gdy skład smogu został już poznany, a jego wpływ na stan zdrowia dobrze zdefiniowany, podejmowane są działania artystyczne, które mają na celu zwrócenie uwagi jak największej grupy osób na ten społeczny problem. 19 lutego 2018 r. krakowskie kino KIKA zorganizowało Smog Film Fest - festiwal, którego smog stał się głównym bohaterem. Według organizatorów to właśnie smog sprawia, że:

Kraków przoduje w światowych rankingach. Dzięki niemu zima w Polsce na zdjęciach wygląda naprawdę magicznie. Dzięki niemu nie musisz kupować zagryzki, wystarczy, że przeżujesz powietrze. A tak na poważnie: smog truje, smog nie maleje, smog zdaje się nie zaprzętać włodarzy miast i wsi. Ale smog też inspiruje (już od jakiegoś czasu!) autorów filmów i książek.

Jednodniowy festiwal był przeglądem twórczości smogowej skupionym na projekcji filmu animowanego *Smog* (1985) Leszka Galewicza, filmu dokumentalnego *Oczyścić niebo* (2017) Pawła

13 Tamże.

14 T. Peiper, *Miasto. Masa. Maszyna*, „Zwrotnica” 1922, nr 2 (lipiec), s. 25.

Augustynka Halnego i Piotra Augustynka oraz filmu fabularnego *PM 2.5* (2017) Piotra Biedronia. Festiwalowi towarzyszyło spotkanie literackie dotyczące książki *Sezon grzewczy*, której autorami są Piotr Marecki i miasto Kraków (2017), zwaną eksperymentalną, mozaikową „kroniką zapyłonego czasu, która pokazuje, że nawet coś tak ulotnego jak powietrze może być ważnym czynnikiem życia społecznego”¹⁵. W dniach 15-27 października 2018 roku w Galerii Instytutu Cervantesa przy ul. Kanoniczej 12 w Krakowie odbył się wernisaż wystawy „Pejzaż smogowy” w ramach wydarzeń New Earth International Film Festival. Do zaprezentowania prac zostali zaproszeni artyści z Polski i zagranicą: Ronda Bautista López, Krzysztof Kiwerski, Andrzej Lichota, José Vicente Losada, Jacek Sroka oraz Mercedes de la Zarza.

Niewiele wspólnego ze smogiem, ale wiele z samym procesem czytania w specjalnych warunkach atmosferycznych ma książka *Niewidzialni* stworzona przez Fundację Kapucyńską, którą można czytać tylko na mrozie. Było to możliwe dzięki użyciu specjalnej termoaktywnej farby, reagującej na mróz.

Kiedy książka znajduje się w ciepłym pomieszczeniu, widzimy tylko białe kartki, a kiedy się ją otworzy na mrozie, pojawia się tekst. Tak naprawdę możemy zobaczyć, co gra w duszy osobom bezdomnym, ale też poczuć, jak to naprawdę jest, gdy jest się bezdomnym. Książka zawiera teksty tworzone przez osoby bezdomne. Składają się na nią krótkie formy literackie, wspomnienia i wiersze, w których opowiadają o swoim życiu, doświadczeniach i tym, jak próbują wyjść na prostą, mimo piętrzących się trudności¹⁶.

Warto dodać, że również okładki wykonane zostały z kawałków odpadowej tektury, a tytuł książki wydrukowany na taśmie pakowej, którą bezdomni własnoręcznie naklejali na książki.

Ten polski eksperyment, który miał zwracać uwagę na bezdomnych zagrożonych przez wysokie mrozy, najbardziej bliski jest idei *Wiersza smogowego* - utworu również istniejącego tylko w określonych warunkach atmosferycznych, tj. szczególnie mocnego nasilenia stężenia smogu w powietrzu, który wystawia na co dzień nasze zdrowie na szwank i zagraża przedwczesną śmiercią.

Bibliografia

- Jeśman J., *Państwo i artyści kontra smog*, <https://www.swps.pl/centrum-prasowe/archiwum-centrum-prasowego/279-komentarz-ekspercki/3887-panstwo-i-artysci-kontra-smog> [dostęp: 12.10.2018].
- Jędrak J., *Smog w sztuce*, <https://smoglab.pl/smog-w-sztuce> [dostęp: 17.10.2018].
- Jouet J., *Z przymusami (ale i bez)*, tłum. J. Franczak, „Nowy Wiek” 2006, nr 10.
- Korporacja Ha!art, <http://www.ha.art.pl/wydawnictwo/nowe-ksiazki/5261-sezon-grzewczy.html> [dostęp: 12.10.2018].
- Map range*, https://rosettacode.org/wiki/Map_range [dostęp: 23.11.2018].
- Parikka J., *The Sensed Smog: Smart Ubiquitous Cities and the Sensorial Body*, „The Fibreculture Journal” 2017, No. 29, <http://twenty-nine.fibreculturejournal.org/fcj-219-the-sensed-smog-smart-ubiquitous-cities-and-the-sensorial-body> [dostęp: 9.10.2019].
- Peiper T., *Miasto. Masa. Maszyna*, „Zwrotnica” 1922, nr 2 (lipiec).
- Różański M., „*Niewidzialni*”. *Książka do czytania... na mrozie*, <http://www.niepelnosprawni.pl/ledge/x/451615> [dostęp: 12.10.2018].
- Shannon C.E., *A Mathematical Theory of Communication*, „The Bell System Technical Journal” 1948, No. 27.

¹⁵ Korporacja Ha!art, <http://www.ha.art.pl/wydawnictwo/nowe-ksiazki/5261-sezon-grzewczy.html> [dostęp: 12.10.2018].

¹⁶ M. Różański, „*Niewidzialni*”. *Książka do czytania... na mrozie*, <http://www.niepelnosprawni.pl/ledge/x/451615> [dostęp: 12.10.2018].