

# Sprawozdanie - Latent Semantic Indexing

Arkadiusz Kraus

21 maja 2019

## 1 Zadanie

Zadanie polegało na zastosowaniu metody Latent Semantic Indexing (LSI), aby w zbiorze tekstów móc wyszukiwać te związane z podaną frazą.

## 2 Zbiór danych

Zbiór danych jaki zastosowałem pochodzi ze strony <https://www.kaggle.com/snapcrack/all-the-news>. Jest to zbiór artykułów z amerykańskich gazet z lat 2016-2017 o różnej tematyce. Jest złożony z ponad 140 tys. tekstów.

## 3 Zastosowane przekształcenia

Na początku przetwarzamy wstępnie przetwarzamy zbiór artykułów, aby później móc skuteczniej w nim wyszukiwać:

### 3.1 Zbiór użytych słów

Dla każdego artykułu bierzemy wszystkie użyte słowa i tworzymy wektor słów dla wszystkich tekstów. Dla kolejnych podzbiorów artykułów posiada on następujące wielkości.

liczba artykułów	rozmiar wektora
1000	44 tys.
10000	140 tys.
50000	230 tys.
wszystkie(142572)	400 tys.

### 3.2 Wstępne przetworzenie

Liczba słów uzyskana w poprzednim podpunkcie jest bardzo duża, aby uczynić algorytm wydajniejszym staramy się zredukować ten wektor. We wstępnym przetworzeniu pomijamy wielkość liter oraz znaki interpunkcyjne.

### 3.3 Stemming

Wiele słów może występować w różnych formach np. take i taking. Nas jednak interesuje sam fakt czynności, a nie koniecznie np. czas w jakim była wykonywana. Innym przypadkiem jest też liczba mnoga rzeczownika, gdzie nadal interesuje nas sam rzeczownik. Dlatego stosujemy algorytm stemmingu (dokładniej Porter Stemmer), doprowadzamy słowa do wspólnego rdzenia i usuwamy duplikaty. W połączeniu z poprzednim krokiem pozwala to znacząco zredukować rozmiar wektora:

liczba artykułów	rozmiar wektora
1000	24 tys.
10000	70 tys.
50000	127 tys.
wszystkie(142572)	234tys.

### 3.4 Stopwords

Stopwords to zbiór słów takich jak przyimki itp, które występują bardzo często we wszelkich tekstach. Nie mają one większego znaczenia dla treści dlatego również można je pominąć. Nie redukuje to wielkości wektora znacznie (o około 150 słów), jednak zwiększa jakość wyszukiwania.

### 3.5 Tworzenie macierzy rzadkiej

Chcemy teraz utworzyć macierz, która jako wiersze będzie przyjmować kolejne artykuły, a jako kolumny kolejne słowa ze zbioru. Wartością w komórce  $a_{i,j}$  macierzy będzie liczba wystąpień  $j$ -tego słowa w  $i$ -tym artykule. Zauważmy, że będzie to macierz rzadka ponieważ artykuły mają ok. 4000 tys słów, a wektor słów jest znacznie większy. Dodatkowo jeśli artykuł jest na jakiś temat to słowa często się w nim powtarzają. Aby przyspieszyć obliczenia (i w ogóle je umożliwić) stosujemy więc przechowywanie w postaci macierzy rzadkiej (dokładniej CSR - Compressed Sparse Row).

### 3.6 IDF

W celu zwiększenia jakości wyszukiwania stosujemy Inverse Document Frequency (IDF). Pozwala on nam zmniejszyć znaczenie słów, które występują w wielu artykułach i zwiększyć tych, które rzadko.

### 3.7 Normalizacja

Normalizacja pozwala uniezależnić korelację od długości tekstu. Wykonujemy ją od razu, aby potem przyspieszyć wyszukiwanie, aby nie trzeba było obliczać normy za każdym razem.

### 3.8 Odszumianie

W celu redukcji szumów stosujemy algorytm SVD i wybieramy  $k$  najciekawszych wartości własnych. Co ciekawe zwiększa to liczbę niezerowych elementów macierzy z ok 1-2 mln do 87 mln dla 10 tys. artykułów.

## 4 Obliczenia

Problem ten jest bardzo wymagający obliczeniowo oraz pamięciowo. Przechowanie macierzy wielkości *liczba artykułów* \* *liczba słów* dla większej ilości artykułów skutkuje skończeniem pamięci. W pliku *logs* znajdują się czasy wstępnego przetworzenia dla 50000 i wszystkich artykułów. Wyszukiwanie po przetworzeniu i odczytaniu wcześniejszej macierzy jest również zależne od ilości artykułów i dla większej ilości jeszcze trochę mu brakuje względem czasów osiągniętych przez Google.

liczba artykułów	rozmiar wektora	rozmiar wektora po przetworzeniu	czas przetworzenia	czas wyszukania
1000	44 tys	24 tys	natychmiast	natychmiast
10000	140 tys	70 tys	1 min	1 s
50000	230 tys	127 tys	35 min	15 sek
wszystkie(142572)	400 tys	234 tys	2,5h	1 min

Niestety dla 50000 oraz dla wszystkich artykułów stworzenie macierzy po dekompozycji SVD okazało się zbyt czasochłonne. Przy próbie utworzenia normalnej macierzy po dekompozycji dla np. 200 wartości własnych otrzymywałem brak pamięci, natomiast dla macierzy rzadkich obliczenia trwały w nieskończoność (obliczenia na nich są czasochłonnymi operacjami). Dlatego dla tych wielkości wyszukiwarka używa macierzy bez odszumienia.

## 5 Wyniki

Aplikacja jest podzielona na dwie części - jedna odpowiadająca za indeksowanie - *articles\_processor.py* i druga wyszukująca treści dla zadanej frazy - *search.py*. Obie części działają poprawnie. Poniżej znajduje się przykład:

weather rain south

Search

### Hurricane Matthew Toll Climbs to at Least 17 as North Carolina Suffers Record-Breaking Flooding - The New York Times

Correlation: 0.08182207146267585

Hurricane Matthew was downgraded to a cyclone early Sunday morning as it hit North Carolina and Virginia with a weakened but still powerful punch. Dispatches from our reporters on the ground a live storm tracker map and answers to reader questions will be updated below. ■ The storm's death toll in the United States has climbed to at least 17. Gov. Pat McCrory of North Carolina said on Sunday morning that his state's toll had risen to eight. Officials in Georgia confirmed three deaths on Saturday. At least six fatalities in Florida have been attributed to the storm. Nearly 900 people have died in Haiti, according to a Reuters report based on information from civil protection and local officials. ■ Bands of heavy rain are leading to flooding in parts of eastern North Carolina, according to the National Hurricane Center, which may result in flooding and flash flooding elsewhere in the region. Forecasters warned that areas along the Neuse River in Goldsboro, N. C. could experience flooding worse than the devastating inundation that followed Hurricane Floyd in 1999. The river was projected to hit 30.9 feet around 2 a. m. on Monday, surpassing the record of 28.9 feet caused by Floyd. ■ In Georgia, the storm created a record surge at Tybee Island, near the state's border with South Carolina. The surge reached 12.5 feet, according to the Chatham County Emergency Management Agency, which exceeded the previous high of 12.2 feet, set during Hurricane David in 1979. ■ To cover the storm and its aftermath, The New York Times has journalists deployed along the its path. Follow our correspondents on Twitter. The storm's assault on North Carolina extended into Sunday, and the governor said at least seven people had been killed in the state. "I wouldn't assume that there aren't people clinging for life right now in houses that are underwater that we have yet to reach, especially in areas," Gov. Pat McCrory told reporters in Raleigh, the state capital. "That's what my major concern is." Heavy rain was still pounding parts of the state as the governor spoke, and Mr. McCrory said floodwaters could rush through North Carolina for days. The fallout, he said, "is going to be a prolonged event." "This is still an extremely dangerous situation," Mr. McCrory said. "And I cannot stress it more especially in the areas of Rocky Mount, Kinston, Greenville, Goldsboro and other eastern towns, cities. We're going to have major issues with rivers and flooding." Mr. McCrory said 58 boat crews had rescued nearly 900 people by Sunday morning. More than 4,000 people were in shelters on Sunday, Mr. McCrory said, and about 760,000 homes and businesses were without electricity. Parts of Interstate 40 and Interstate 95 were shut down. — ALAN BLINDER The storm lashed South Carolina's Lowcountry with rain and sent the sea surging upward. It turned roads into rivers and ripped trees from the inundated ground before making landfall north of Charleston. More than 430,000 people around the state had been left without power as of late Saturday. Gov. Nikki Haley of South Carolina said on Sunday that she had lifted evacuation orders for four counties, including Charleston. But residents in four other counties — Beaufort, Georgetown, Horry and Jasper — were still urged to stay away. "We had a lot of rain, for many hours," said Capt. Bob Bromage, the public information officer for the Sheriff's Department in Beaufort County, southwest of Charleston. "We had flooding, we may have had some surge that we're not aware of yet, we have downed trees, we have reports of downed trees on houses already." — JESS BIDGOOD, in Charleston The American Red Cross has published a list of safety tips for homeowners returning to property damaged in hurricanes. Once cleanup begins, taking photographs of damaged or destroyed property before it is discarded provides a record for use in making insurance claims, says Ann Carns, who writes the Your Money Adviser consumer finance column for The Times. Gov. Rick Scott of Florida said on Saturday that the state had suffered "unbelievable amounts of beach erosion," as well as damaged roads, flooding and felled trees. But Mr. Scott, who flew along the coastline to survey storm damage, said the storm's consequences could have been far worse. "The first thing we can say is we are all blessed that Matthew stayed off our coast," he said. "I worried the whole time that even though the track was off our coast, that it would turn in and have a direct hit at some point." State and local officials said they were trying to expedite recovery. "We're going to ride each other hard," said Lenny Curry, the mayor of Jacksonville. "We're going to ride our utilities, we're going to ride all our workers to get this community back together as quickly and as safely as possible." Still, Mayor Charlie Latham of Jacksonville Beach asked for patience. *"Matthew saved lives in Florida, and he has fought with them, so you don't ever want him to hit our beautiful homes and ruin the life we live."* Mr. I. Auburn said. AT A N. H. INT'D in Orlando In Haiti, the

### Tropical Storm Hermine Leaves Trail of Power Failures - The New York Times

Correlation: 0.08009828004945116

The remnants of what had been Hurricane Hermine swept up and along the Eastern Seaboard on Saturday, disrupting the holiday weekend in much of the coastal South while preparing to bedevil the Northeast well into the week. The storm, which made its landfall early Friday near St. Marks, Fla. as a Category 1 hurricane, led to hundreds of thousands of power failures and flooded roadways. In Florida, dealt its first direct strike by a hurricane in nearly 11 years, the Tallahassee area was particularly hard hit, and officials said it could be almost a week before electricity was fully restored. "We still have a lot of work to do following the storm," Gov. Rick Scott of Florida said at a news conference on Saturday in Tallahassee, the state capital. "We'll continue to spend the coming days assessing the damage and responding to the needs of our Florida families." Mr. Scott said there was "significant damage" in the state, including "a lot of downed power lines." Residents and officials described destroyed businesses, boats set adrift, crumbled sea walls and battered homes. The National Hurricane Center said the storm, whose winds reached as high as 80 miles per hour after it spent days swirling through the Gulf of Mexico, remained powerful on Saturday, even after it crossed the Florida Panhandle, Georgia and the Carolinas. The storm became a cyclone on Saturday morning, but forecasters warned that it still "could be near hurricane intensity" for part of the week. And so the consequences of the storm's path stretched from Florida, which reported the first fatality attributed to it, to New York, where the city said beaches would be closed Sunday, to Cape Cod, Mass. which was under a tropical storm watch. The hurricane center said Saturday evening that the storm would "meander slowly offshore of the coast for the next couple of days" before churning past New Jersey and New York. Officials said storm surges would be "accompanied by large and dangerous waves," and they warned of the possibility of "inundation" from Virginia to Connecticut. Gov. Chris Christie of New Jersey declared emergencies for Atlantic, Cape May and Ocean Counties, and Gov. Andrew M. Cuomo of New York said the state's emergency operations center would be activated on Sunday. On Saturday, the National Guard was on alert in Virginia, where floodwaters prompted road closings in the Hampton Roads region, and visitors to North Carolina's Outer Banks, where strong rip currents were reported, were advised to adjust their travel schedules. "I'm relieved that everyone took this storm seriously, paid attention to the weather forecast and heeded the warnings of state and local officials," said Gov. Pat McCrory of North Carolina, whose state shut down major bridges in the Outer Banks because of winds associated with the storm. Officials said that the death of one person in a traffic accident in eastern North Carolina was probably connected to the storm. Although people in many communities said the storm would soon become only a minor memory, parts of Florida were entering what could be a protracted recovery, and there were some concerns that the absence of could prove dangerous. In Leon County, which includes Tallahassee, more than half of homes were without electricity on Saturday, and the American Red Cross opened a comfort shelter, a place "for residents to cool off, hydrate and receive information from area nonprofit agencies." The authorities planned to distribute bottled water throughout the county, where high temperatures were predicted to be in the upper 80s or lower 90s for the next several days. During his appearance on Saturday at Florida's emergency operations center, Mr. Scott made a dispiriting roll call of counties and the extent of their power failures. In Wakulla County, just south of Tallahassee, about 72 percent of customers were without electricity. "This is the No. 1 issue people are telling me about," Mr. Scott said before leaving for another afternoon of damage assessments.

### Hurricane Matthew's Toll Rises Flooding Strands 1,500 in North Carolina - The New York Times

Correlation: 0.07677973382307797

Rysunek 2: Wyniki wyszukiwania dla hasła "weather rain south" część druga

Dla większej ilości artukulów gdzie nie było możliwe przeprowadzenie SVD zdarzają się jednak wyniki takie jak np. ten:

**Want to be the next Mark Zuckerberg? His CEO sister says don't do these 3 things**

*Correlation: 0.27399432445989885*

" 'Fans of "Shark Tank" know there are some. ...'

**Sometimes They Wink At You - Breitbart**

*Correlation: 0.1842246519195048*

I can't imagine what Sony's marketing department was trying to tell us about fans of the new feminist Ghostbusters.

**This is what happens when you slice a laptop in half with water**

*Correlation: 0.17145244151302674*

"Or, if you're looking for some fun, slicing through everyday things like shoes or a laptop." 'Business Insider talked to the creator of a new YouTube channel called "... This person didn't want to be identified since they are using the tool at work — albeit — but they did walk us through what it's like to turn water into something that can slice a baseball cleanly in half.' 'It may look like a just a stream of ...'

**Watch: Vin Scully Says Goodbye After 67-Year Career - Breitbart**

*Correlation: 0.1665685942552541*

📷 📺 pic.twitter. Legendary baseball announcer Vin Scully signed off for the final time Sunday after the Los Angeles fell to the San Francisco Giants . He said goodbye after a and reassured baseball fans everywhere that life will go on without him. Transcript as follows: "You know, friends, so many people have wished me congratulations on a career in baseball and they wished me a wonderful retirement with my family and now, all I can do is tell you what I wish for you. " 'May God give you.. For every storm, a rainbow, For every tear, a smile, For every care, a promise, And a blessing in each trial. For every problem life sends, A faithful friend to share, For every sigh, a sweet song, And an answer for each prayer.' You and I have been friends for a long time, but I know in my heart

Rysunek 3: Przykład wyniku z szumami

W tym przypadku dopiero któryś wynik jest taki jak powinien, a na początku jest szum. To samo zapytanie przy mniejszej ilości artykułów i z usuniętym szumem wygląda następująco:

**Tim Tebow Homers in First At-Bat (and Finishes 1 for 6) - The New York Times**

*Correlation: 0.08417560679864274*

Tim Tebow stepped to the plate Wednesday for his first as a professional baseball player and took a swing at the first pitch he saw. Naturally, he hit a home run. Tebow is a polarizing figure in the sports world. As a Heisman Trophy winner while at the University of Florida and, briefly, as an N. F. L. quarterback, he attracted both adulation and scorn for his overt Christian faith and his habit of Tebowing, or bowing in prayer after successes. On the field, some saw a quarterback with anemic passing statistics while others saw a player who always seemed to find a way to win. He faced considerable skepticism when, with his football days seemingly over, he announced plans to try for a baseball career at age 29. He had not played the game since high school. Still, scouts from 28 teams attended an open workout, and the Mets saw enough to sign Tebow to a minor league contract and send him to the Florida Instructional League. Game No. 1 came Wednesday against the St. Louis Cardinals' team in front of about 250 fans in Port St. Lucie, Fla. Tebow batted second and played left field. His first ended in seconds with the homer to . The pitcher was John Kilichowski, drafted out of Vanderbilt by the Cardinals in the 11th round in June. He posted a 2.70 E. R. A. in Class A with the Peoria Chiefs and the State College Spikes this past season, giving up four home runs in 11 games. Tebow's teammates rushed out to greet him at home plate after the bash. At that point, his career stats looked like this: batting average 1.000, average 1.000, slugging average 4.000. And let's make up a stat. Ratio of pitches seen to home runs delivered: 1:1. Barry Bonds? Babe Ruth? Tim Tebow was surpassing them all, and making the skeptics look pretty foolish. In his second Tebow grounded into a double play, dropping his batting average to a merely superhuman .500. Then he was 1 for 3, and 1 for 4, and he finished the day 1 for 6.

**One Season Ends and Another Begins: Baseball Playoff Matchups Are Set - The New York Times**

*Correlation: 0.07783777397365577*

The baseball gods spend six months twisting the sport into a knotty lump. No one knows quite how to untangle it, and then the final week zips by. All those possibilities, resolved just like that. The season rushes to the end. "Rushes to the end," Baltimore Orioles Manager Buck Showalter said, musing in his office at Yankee Stadium before this weekend's critical series. "That sounds like a great name for a book." Now we can close the volume on the 2016 regular season. The Orioles survived to live another day, earning a date in Toronto on Tuesday for the American League game against the Blue Jays. The Mets will host the National League game the next night, against the San Francisco Giants at Citi Field. The Orioles and the Blue Jays both won on Sunday to finish with identical records. The Blue Jays won the teams' season series to earn the right to host the game, which was also how the Mets did it. The Mets and the Giants both finished but the Mets won the season series. The A. L. winner will face the Texas Rangers, who went in a division series opener on Thursday. The Boston Red Sox will visit Cleveland that day to start their series with the Indians. The N. L. series begin Friday in Chicago and Washington, with the Cubs ( ) facing the winner and the Nationals hosting the Los Angeles Dodgers. Before we get there, though, the appetizers will be delicious. The Orioles lost the A. L. Championship Series in 2014. The Blue Jays lost it last year. Neither of the teams, rivals in the A. L. East, has won a pennant in decades, and both are trying to do it with brute force. The Orioles led the majors in homers this season, with 253, and the Blue Jays had 221 — the next highest total in the playoff field. The N. L. game features a dream pitchers' duel: the Mets' Noah Syndergaard against the Giants' Madison Bumgarner. In Syndergaard's last postseason appearance — Game 3 of last fall's World Series against Kansas City — he announced his presence with a

Rysunek 4: Przykład wyniku bez szumu

Przeanalizujemy jeszcze wyniki dla różnych współczynników k przy SVD:  
Bez SVD:

trump's wife	Search
<b>Unity: Bill and Hillary Clinton, George and Laura Bush to Attend Trump's Inauguration</b> <i>Correlation: 0.14132442211735358</i> Former President Bill Clinton and wife Hillary Clinton will attend Donald Trump's inauguration, according to several reports citing aides close to the Clintons. [New York magazine cited "sources" to the Clintons saying they "decided to do so out of a sense of duty and respect for the American democratic process." Former President George W. Bush and his wife Laura Bush will also attend the inauguration. In a statement, the Bushes said they were "pleased to be able to witness the peaceful transfer of power — a hallmark of American democracy" by attending the ceremony in Washington, D. C.	
<b>Jason Miller Backs Out of White House Communications Job Under Trump - The New York Times</b> <i>Correlation: 0.13909956022465036</i> Jason Miller, who was named as White House communications director by Donald J. Trump two days ago, has informed Mr. Trump that he will not take the job, according to a statement from Mr. Miller. In the statement, Mr. Miller said he wanted to spend more time with his family. His wife is expecting their second child in January, he said. "It's clear they need to be my top priority right now, and this is not the time to start a new job as demanding as White House communications director," Mr. Miller wrote. A veteran of several Republican campaigns, Mr. Miller was an avid supporter of Mr. Trump, appearing regularly on television on his behalf throughout the campaign. He had been a principal spokesman during the transition period. In the statement, Mr. Miller said that Sean Spicer, who was recently named press secretary, would serve as communications director, as well.	
<b>Sarah Netanyahu Tells Melania Trump of Common Bond: 'Media Hate Us but the People Love Us'</b> <i>Correlation: 0.09130866695351268</i> The Times of Israel reports: After their husbands praised the "unbreakable bond" between Israel and the US at the welcome ceremony for US President Donald Trump, Sara Netanyahu, Prime Minister Benjamin Netanyahu's wife, told US First Lady Melania Trump that the two couples had more in common than pure geopolitical interests. [Chatting candidly on the tarmac at Ben Gurion Airport on Monday, Sara Netanyahu said that, like the Trumps, Israel's "first couple" had been treated unfairly by the media but were loved by the people. "You know in Israel all the people like us. The media hate us but the people love us," Netanyahu told the first lady in a recording that picked up their private conversation. "Like you." "We have a lot in common," Trump replied. Netanyahu, with a broad smile on her face, promised to continue the discussion "over dinner." Read more here.	

## Rysunek 5: Wynik dla zapytania "Trump's wife" bez svd

Dla  $k = 50$ :

trump's wife	Search
<b>McCain: Trump Should Release His Tax Returns - Breitbart</b> <i>Correlation: 0.03697608714316153</i> Sunday on CNN's "State of the Union," Sen. John McCain () said President Donald Trump should release his tax returns. Host Jake Tapper asked, "Given that President Trump is proposing changes that could benefit him personally by tens of millions if not hundreds of millions of dollars, do you think that the American people have a right to see his tax returns so that they can know how this plan would affect his bottom line?" McCain answered, "I think I've said all along I thought every candidate for president should disclose their tax returns. I haven't changed." Follow Pam Key on Twitter @pamkeyNEN	
<b>Wallace to Priebus on Trump's Press Is the 'Enemy' Tweet: 'You Don't Get to Tell Us What to Do' - Breitbart</b> <i>Correlation: 0.03618902303723311</i> On this weekend's broadcast of "Fox New Sunday," host Chris Wallace got into a long and heated exchange with White House chief of staff Reince Priebus over President Donald Trump's tweet labeling the press an "enemy" of the American people. After several minutes of back and forth arguing over the media's coverage of the Trump administration, Wallace said, "You don't get to tell us what to do Reince. You don't get to tell us what to do any more than Barack Obama did. Barack Obama whined about Fox News all the time, but I got to say, he never said we were an enemy of the people." Later in the segment, Priebus said, "I don't know why you are going bananas." Follow Pam Key on Twitter @pamkeyNEN	
<b>WH Spox on Obama's Denial of Wiretap Claim: 'They Don't Have the Best Track Record' - Breitbart</b> <i>Correlation: 0.035386844086497504</i> Sunday on ABC's "This Week," while discussing President Donald Trump's tweets accusing the former Obama administration of wiretapping Trump Tower before the election, White House deputy press secretary Sarah Huckabee Sanders questioned President Barack Obama's denial via a spokesperson. "Well, I think, they don't have the best track record," she said. "They said they had a cardinal rule, well frankly that's just not true. The president himself got directly involved when it came to the email scandal with Hillary Clinton." Huckabee Sanders added, "All we're saying is let's take a closer look. Let's look into this. If this happened, if this is accurate, this is the biggest overreach and the biggest scandal." Follow Pam Key on Twitter @pamkeyNEN	
<b>CNN's Jim Acosta to Spicer: Is Trump 'Detached From Reality?' - Breitbart</b>	

## Rysunek 6: Wynik dla zapytania "Trump's wife" i $k = 50$

### Mark Levin on Trump Wiretapping Claims: 'The Evidence Is Overwhelming' - Breitbart

Correlation: 0.03201471178590111

Sunday on Fox News Channel's "Fox Friends," while discussing President Donald Trump's tweets accusing the former Obama administration of wiretapping Trump Tower before the election, conservative radio talk show host Mark Levin broke down the wiretapping reporting. Levin said, "Well, a pleasure to be here. The evidence is overwhelming. This is not about President Trump's tweeting. This is about the Obama administration spying, and the question is not whether it spied. We know they went to the FISA court twice. The question is who they did spy on and the extent of the spying that is the Trump campaign, the Trump transition, Trump surrogates." Levin then read several news articles about the surveillance. Follow Pam Key on Twitter @pamkeyNEN

### Nicolle Wallace: 'We Have Just Elected a Man Who Bullies Female Reporters' - Breitbart

Correlation: 0.03138985523313076

Sunday on NBC's "Meet the Press," former communications chief for President George W. Bush Nicolle Wallace said Donald Trump "bullies female reporters at his rally as an applause line." Wallace said, "We have just elected a man who bullies female reporters at his rally as an applause line. We have just elected a man who started a hot war with a female anchor instead of attending a debate she moderated. We are in a new place. And I don't think it's good. And I don't think it has any parallels to the past." Follow Pam Key on Twitter @pamkeyNEN

### Donald Trump Is Sworn In as President, Capping His Swift Ascent - The New York Times

Correlation: 0.03091255892809344

WASHINGTON — Donald John Trump was inaugurated as the 45th president of the United States on Friday, ushering in a new era that he vowed would shatter the established order and reverse a national decline that he called "this American carnage." In a ceremony that capped a remarkable rise to power, Mr. Trump presented himself as the leader of a populist uprising to restore lost greatness. He outlined a dark vision of an America afflicted by "the ravages" of economic dislocation and foreign exploitation, requiring his approach to turn around. "I will fight for you with every breath in my body, and I will never, ever let you down," Mr. Trump told hundreds of thousands of admirers and onlookers in a forceful Inaugural Address from the West Front of the Capitol. "America will start winning again, winning like never before. We will bring back our jobs. We will bring back our borders. We will bring back our wealth. And we will bring back our dreams." Mr. Trump's ascension amounted to a hostile takeover of a capital facing its most significant disruption in generations. While officially a Republican, he has taken on leaders of both parties and, with no prior political career of his own, made clear that he saw himself as the ultimate outsider not beholden to the current system. "We will no longer accept politicians who are all talk and no action, constantly complaining but never doing anything about it," he said. "The time for empty talk is over. Now arrives the hour of action. Do not allow anyone to tell you that it cannot be done." Mr. Trump's view of the United States was strikingly grim for an Inaugural Address — a country where mothers and children are "trapped in poverty in our inner cities," where "factories" are "scattered like tombstones

Rysunek 7: Wynik dla zapytania "Trump's wife" i k = 50 część druga

Dla k = 100:

Trump's wife Search

### Donald Trump Is Sworn In as President, Capping His Swift Ascent - The New York Times

Correlation: 0.03828337959877941

WASHINGTON — Donald John Trump was inaugurated as the 45th president of the United States on Friday, ushering in a new era that he vowed would shatter the established order and reverse a national decline that he called "this American carnage." In a ceremony that capped a remarkable rise to power, Mr. Trump presented himself as the leader of a populist uprising to restore lost greatness. He outlined a dark vision of an America afflicted by "the ravages" of economic dislocation and foreign exploitation, requiring his approach to turn around. "I will fight for you with every breath in my body, and I will never, ever let you down," Mr. Trump told hundreds of thousands of admirers and onlookers in a forceful Inaugural Address from the West Front of the Capitol. "America will start winning again, winning like never before. We will bring back our jobs. We will bring back our borders. We will bring back our wealth. And we will bring back our dreams." Mr. Trump's ascension amounted to a hostile takeover of a capital facing its most significant disruption in generations. While officially a Republican, he has taken on leaders of both parties and, with no prior political career of his own, made clear that he saw himself as the ultimate outsider not beholden to the current system. "We will no longer accept politicians who are all talk and no action, constantly complaining but never doing anything about it," he said. "The time for empty talk is over. Now arrives the hour of action. Do not allow anyone to tell you that it cannot be done." Mr. Trump's view of the United States was strikingly grim for an Inaugural Address — a country where mothers and children are "trapped in poverty in our inner cities," where "factories" are "scattered like tombstones across the landscape" and where drugs and crime "have stolen too many lives." "This American carnage," he declared, "stops right here and stops right now." He got started right away with rolling back the policies of his predecessor, former President Barack Obama, by issuing orders freezing new regulations from recent weeks and ordering agencies to "ease the burden" of the Affordable Care Act during the transition from repealing to replacing the law. More orders are planned for next week. Wearing a dark suit and red tie and accompanied by his wife, Melania, in a suit and matching gloves, Mr. Trump took the oath administered by Chief Justice John G. Roberts Jr. precisely at noon. Michael Richard Pence, a former governor and congressman from Indiana, was sworn in minutes before as vice president by Justice Clarence Thomas. Mr. Trump assumed the presidency of a country still unsettled after a polarizing election and entered office with less support in polls than any other president in recent history. It was clear from the day that there would be no grace period either for or by the new president. The Senate confirmed two cabinet nominations — James N. Mattis as defense secretary and John F. Kelly as secretary of homeland security — but Democrats temporarily held up Mike Pompeo's confirmation as C. I. A. director. Throughout the day, there were mostly peaceful protests against the new president. Sporadic violence broke out as demonstrators smashed shop windows and burned a limousine, while police officers in riot helmets responded with tear gas. More than 200 people were arrested. Liberal groups prepared for a women's march on Saturday that they said could draw hundreds of thousands. Mr. Trump made only passing efforts to reach out to Democrats beyond thanking Mr. Obama and his wife, Michelle, for their handling of the transition. "They have been magnificent," he said in his speech. He later praised his defeated opponent, Hillary Clinton, at a luncheon with congressional leaders, asking her and former President Bill Clinton to stand for applause. "I have a lot of respect for these two people," he said. Democrats were not impressed. "I was pretty shocked by how dark it was," Senator Sherrod Brown, Democrat of Ohio, said of Mr. Trump's Inaugural Address. "I love this country, and I don't understand how a president of the United States that loves his country could paint a picture of its failures." He added, "It was interesting sitting up onstage with a bunch of billionaires hearing him say how bad the country was." The National Mall was filled with supporters, many wearing "Make America Great Again" hats and chanting "Trump! Trump! Trump!" But the lingering animosity from the presidential campaign was on display, too. When Mrs. Clinton arrived, some in the crowd chanted, "Lock her up," mimicking Mr. Trump's campaign rallies. As he took the oath, a cluster of people blew whistles and screamed, "Not my president," before being escorted out. While large, the crowds on a soggy day did not rival the energetic throngs at Mr. Obama's first inauguration eight years ago, according to aerial

photosnaps. The Washington Mall crowd seemed less than half as many as on Friday morning as in 2009, and knots of bystanders along the inaugural parade route seem not as thick. In a

Rysunek 8: Wynik dla zapytania "Trump's wife" i k = 100

Dla k = 200:

**Unity: Bill and Hillary Clinton, George and Laura Bush to Attend Trump's Inauguration**

*Correlation: 0.047420611139398675*

Former President Bill Clinton and wife Hillary Clinton will attend Donald Trump's inauguration, according to several reports citing aides close to the Clintons. [New York magazine cited "sources" to the Clintons saying they "decided to do so out of a sense of duty and respect for the American democratic process." Former President George W. Bush and his wife Laura Bush will also attend the inauguration. In a statement, the Bushes said they were "pleased to be able to witness the peaceful transfer of power — a hallmark of American democracy" by attending the ceremony in Washington, D. C.

**Donald Trump Is Sworn In as President, Capping His Swift Ascent - The New York Times**

*Correlation: 0.04355159097190291*

WASHINGTON — Donald John Trump was inaugurated as the 45th president of the United States on Friday, ushering in a new era that he vowed would shatter the established order and reverse a national decline that he called "this American carnage." In a ceremony that capped a remarkable rise to power, Mr. Trump presented himself as the leader of a populist uprising to restore lost greatness. He outlined a dark vision of an America afflicted by "the ravages" of economic dislocation and foreign exploitation, requiring his approach to turn around. "I will fight for you with every breath in my body, and I will never, ever let you down," Mr. Trump told hundreds of thousands of admirers and onlookers in a forceful inaugural Address from the West Front of the Capitol. "America will start winning again, winning like never before. We will bring back our jobs. We will bring back our borders. We will bring back our wealth. And we will bring back our dreams." Mr. Trump's ascension amounted to a hostile takeover of a capital facing its most significant disruption in generations. While officially a Republican, he has taken on leaders of both parties and, with no prior political career of his own, made clear that he saw himself as the ultimate outsider not beholden to the current system. "We will no longer accept politicians who are all talk and no action, constantly complaining but never doing anything about it," he said. "The time for empty talk is over. Now arrives the hour of action. Do not allow anyone to tell you that it cannot be done." Mr. Trump's view of the United States was strikingly grim for an Inaugural Address — a country where mothers and children are "trapped in poverty in our inner cities," where "factories" are "scattered like tombstones across the landscape" and where drugs and crime "have stolen too many lives." "This American carnage," he declared, "stops right here and stops right now." He got started right away with rolling back the policies of his predecessor, former President Barack Obama, by issuing orders freezing new regulations from recent weeks and ordering agencies to "ease the burden" of the Affordable Care Act during the transition from repealing to replacing the law. More orders are planned for next week. Wearing a dark suit and red tie and accompanied by his wife, Melania, in a suit and matching gloves, Mr. Trump took the oath administered by Chief Justice John G. Roberts Jr. precisely at noon. Michael Richard Pence, a former governor and congressman from Indiana, was sworn in minutes before as vice president by Justice Clarence Thomas. Mr. Trump assumed the presidency of a country still unsettled after a polarizing election and entered office with less support in polls than any other president in recent history. It was clear from the day that there would be no grace period either for or by the new president. The Senate confirmed two cabinet nominations — James N. Mattis as defense secretary and John F. Kelly as secretary of homeland security — but Democrats temporarily held up Mike Pompeo's confirmation as C. I. A. director. Throughout the day, there were mostly

Rysunek 9: Wynik dla zapytania "Trump's wife" i k = 200

Dla k = 1000:

**Unity: Bill and Hillary Clinton, George and Laura Bush to Attend Trump's Inauguration**

*Correlation: 0.09115432146137104*

Former President Bill Clinton and wife Hillary Clinton will attend Donald Trump's inauguration, according to several reports citing aides close to the Clintons. [New York magazine cited "sources" to the Clintons saying they "decided to do so out of a sense of duty and respect for the American democratic process." Former President George W. Bush and his wife Laura Bush will also attend the inauguration. In a statement, the Bushes said they were "pleased to be able to witness the peaceful transfer of power — a hallmark of American democracy" by attending the ceremony in Washington, D. C.

**Sarah Netanyahu Tells Melania Trump of Common Bond: 'Media Hate Us but the People Love Us'**

*Correlation: 0.07895314722885481*

The Times of Israel reports: After their husbands praised the "unbreakable bond" between Israel and the US at the welcome ceremony for US President Donald Trump, Sara Netanyahu, Prime Minister Benjamin Netanyahu's wife, told US First Lady Melania Trump that the two couples had more in common than pure geopolitical interests. [Chatting candidly on the tarmac at Ben Gurion Airport on Monday, Sara Netanyahu said that, like the Trumps, Israel's "first couple" had been treated unfairly by the media but were loved by the people. "You know in Israel all the people like us. The media hate us but the people love us," Netanyahu told the first lady in a recording that picked up their private conversation. "Like you." "We have a lot in common," Trump replied. Netanyahu, with a broad smile on her face, promised to continue the discussion "over dinner." Read more here.

**Jason Miller Backs Out of White House Communications Job Under Trump - The New York Times**

*Correlation: 0.07719022526248306*

Jason Miller, who was named as White House communications director by Donald J. Trump two days ago, has informed Mr. Trump that he will not take the job, according to a statement from Mr. Miller. In the statement, Mr. Miller said he wanted to spend more time with his family. His wife is expecting their second child in January, he said. "It's clear they need to be my top priority right now, and this is not the time to start a new job as demanding as White House communications director," Mr. Miller wrote. A veteran of several Republican campaigns, Mr. Miller was an avid supporter of Mr. Trump, appearing regularly on television on his behalf throughout the campaign. He had been a principal spokesman during the transition period. In the statement, Mr. Miller said that Sean Spicer, who was recently named press secretary, would serve as communications director, as well.

Rysunek 10: Wynik dla zapytania "Trump's wife" i k = 1000

Widzimy, że najlepsze wyniki otrzymaliśmy dla k w okolicach 100-200. Bez SVD występowały szumu w postaci artykułów o dużej korelacji jednak bardzo krótkich i mało spokrewnionych z tematem. Powróciły one również dla k = 1000. Dla głównych artykułów korelacja nie zmieniała się znacząco jednak zmienia się ona dla szumu.



## 6 Wnioski

Algorytm LSI dobrze sprawdził się przy wyszukiwaniu frazy w podanym zbiorze tekstów. Zajmuje trochę czasu jego wcześniejsze przetworzenie i indeksacja, ale nie są to bardzo duże czasy. Jednak myślę, że w tej wersji nie sprawdziłby się w silniku Google.

Aby go polepszyć możnaby próbować zredukować wielkość wektora słów jeszcze bardziej ponieważ nadal po wypisaniu go widać szumy. Znaczący wpływ ma również SVD, które zostało przeze mnie pominięte dla większej ilości artykułów z powodów ograniczeń technicznych.

Ciekawą częścią algorytmu jest to, że każdy jego element można zapisać w cache'u - komentarze w pliku `articles_processor.py` - co znacząco przyspiesza pracę.