



ITESO, Universidad
Jesuita de Guadalajara

ANÁLISIS Y CLASIFICACIÓN DE DISCURSOS HISTÓRICOS

Jaime Santos Orozco

JAIME A. SANTOS OROZCO

- Estudiante de la Maestría en Sistemas Computacionales
- Lic. En Ingeniería Química
- Programador Full Stack en Oracle de México.



ITESO, Universidad
Jesuita de Guadalajara

OBJETIVO DEL PROYECTO

El objetivo de la investigación es conocer si, a través de técnicas de análisis de textos y técnicas de clusterización se puede llegar a un clasificador confiable para calificar discursos de alto o bajo impacto social, basándose en análisis de discursos históricos previamente clasificados por su popularidad.

Contar con una herramienta que permita cuantificar el parecido de un discurso con aquellos clasificados como "famosos", "importantes" o de alto impacto.



ITESO, Universidad
Jesuita de Guadalajara

OBTENCIÓN DE DATOS

CONSEGUIR DATASETS

Se requiere de transcripciones confiables de discursos históricos, bien escritos y completos.

VERIFICAR FUENTE

Para asegurar los resultados confiables se necesita verificar la autenticidad de la información así como su validez

CANTIDAD DE INFORMACIÓN

Para asegurar el correcto procesamiento de los algoritmos de clasificación se requiere contar con la información suficiente para el ejercicio

Después de buscar información de diversas fuentes tales como los corpus de NLTK, Keaggle, UCI, Google, etc., se utilizaron los datos del repositorio de *American Rhetoric*, dado que incluía los audios originales a partir de los cuales se realizaron las transcripciones, lo cuál da validez a los datos.



ITESO, Universidad
Jesuita de Guadalajara

PREPARACIÓN DE DATOS

Uno de los grandes retos de la preparación de los datos fue el formato, ya que todos estaban disponibles en formato PDF que incluía membretes y publicidad, lo que de ser procesado metería ruido a los modelos. Para la limpieza y preparación de los 80 archivos se siguieron los siguientes pasos:

- Quitar los encabezados, imágenes y pie de páginas
- Pasar cada uno de los archivos a texto plano
- Remover detalles de la transcripción, tales como interrupciones al orador, signos de división de párrafos, etc.
- Nombrar cada archivo con el autor y el título de discurso para su uso futuro



ITESO, Universidad
Jesuita de Guadalajara



AmericanRhetoric.com

John F. Kennedy

Presidential Inaugural Address

delivered 20 January 1961, Washington, D.C.

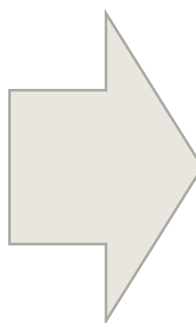


AUTHENTICITY CERTIFIED: Text version below transcribed directly from audio

Vice President Johnson, Mr. Speaker, Mr. Chief Justice, President Eisenhower, Vice President Nixon, President Truman, reverend clergy, fellow citizens:

We observe today not a victory of party, but a celebration of freedom -- symbolizing an end, as well as a beginning -- signifying renewal, as well as change. For I have sworn before you and Almighty God the same solemn oath our forebears prescribed nearly a century and three-quarters ago.

The world is very different now. For man holds in his mortal hands the power to abolish all forms of human poverty and all forms of human misery. And yet the same revolutionary beliefs for which our forebears fought are still at issue around the globe -- the belief that the rights of



[Thank you very kindly, my friends. As I listened to Ralph Abernathy and his eloquent and generous introduction and then thought about myself, I wondered who he was talking about. It's always good to have your closest friend and associate to say something good about you. And Ralph Abernathy is the best friend that I have in the world. I'm delighted to see each of you here tonight in spite of a storm warning. You reveal that you are determined to go on anyhow.

Something is happening in Memphis; something is happening in our world. And you know, if I were standing at the beginning of time, with the possibility of taking a kind of general and panoramic view of the whole of human history up to now, and the Almighty said to me, "Martin Luther King, which age would you like to live in?" I would take my mental flight by Egypt, and I would watch God's children in their magnificent trek from the dark dungeons of Egypt through, or rather across the Red Sea, through the wilderness on toward the promised land. And in spite of its magnificence, I wouldn't stop there.

I would move on by Greece and take my mind to Mount Olympus. And I would see Plato, Aristotle, Socrates, Euripides and Aristophanes assembled around the Parthenon. And I would watch them around the Parthenon as they discussed the great and eternal issues of reality. But I wouldn't stop there.

I would go on, even to the great heyday of the Roman Empire. And I would see developments around there, through various emperors and leaders. But I wouldn't stop there.

I would even come up to the day of the Renaissance, and get a quick picture of all that the Renaissance did for the cultural and aesthetic life of man. But I wouldn't stop there.

I would even go by the way that the man for whom I am named had his habitat. And I would watch Martin Luther as he tacked his ninety-five theses on the door at the church of Wittenberg. But I wouldn't stop there.

I would come on up even to 1863, and watch a vacillating President by the name of Abraham Lincoln finally come to the conclusion that he had to sign the Emancipation Proclamation. But I wouldn't stop there.

I would even come up to the early thirties, and see a man grappling with the problems of the bankruptcy of his nation. And come with an eloquent cry that we have nothing to fear but "fear itself." But I wouldn't stop there.

Strangely enough, I would turn to the Almighty, and say, "If you allow me to live just a few years in the second half of the 20th century, I will be happy."

Now that's a strange statement to make, because the world is all messed up. The nation is sick. Trouble is in the land; confusion all around. That's a strange statement. But I know, somehow, that only when it is dark enough can you see the stars. And I see God working in this period of the twentieth century in a way that men, in some strange way, are responding.

Something is happening in our world. The masses of people are rising up. And wherever they are assembled today, whether they are in Johannesburg, South Africa; Nairobi, Kenya; Accra, Ghana; New York City; Atlanta, Georgia; Jackson, Mississippi; or Memphis, Tennessee -- the cry is always the same: "We want to be free."

And another reason that I'm happy to live in this period is that we have been forced to a point where we are going to have to grapple with the problems that men have been trying to grapple with through history, but the demands didn't force them to do it. Survival demands that we grapple with them. Men, for years now, have been talking about war and peace. But now, no longer can they just talk about it. It is no longer a choice between violence and nonviolence in this world; it's nonviolence or nonexistence. That is where we are today.

And also in the human rights revolution, if something isn't done, and done in a hurry, to bring the colored peoples of the world out of their long years of poverty, their long years of hurt and neglect, the whole world is doomed. Now, I'm just happy that God has allowed me to live in this period to see what is unfolding. And I'm happy that He's allowed me to be in Memphis.

I can remember -- I can remember when Negroes were just going around as Ralph has said, so often, scratching where they didn't itch, and laughing when they were not tickled. But that day is all over. We mean business now, and we are determined to gain our rightful place in God's world.

And that's all this whole thing is about. We aren't engaged in any negative protest and in any negative arguments with anybody. We are saying that we are determined to be men. We are determined to be people. We are saying -- We are saying that we are God's children. And that we are God's children, we don't have to live like we are forced to live.

Now, what does all of this mean in this great period of history? It means that we've got to stay together. We've got to stay together and maintain unity. You know, whenever Pharaoh wanted to prolong the period of slavery in Egypt, he had a favorite formula for doing it. What was that? He kept the slaves fighting among themselves. But whenever something happens in Pharaoh's court, and he cannot hold the slaves in slavery. When the slaves get together, and they get out of slavery. Now let us maintain unity.

Secondly, let us keep the issues where they are. The issue is injustice. The issue is the refusal to recognize the rights of man in its dealings with its public servants, who happen to be sanitation workers. Now, we've got to keep the issue where it is. We always the problem with a little violence. You know what happened the other day, and the press dealt with it. I read the articles. They very seldom got around to mentioning the fact that one thousand, three hundred and thirty-five men went on strike, and that Memphis is not being fair to them, and that Mayor Loeb is in dire need of a doctor. We've got to get around to the issue.

Now we're going to march again, and we've got to march again, in order to put the issue where it is supposed to be. And I want everybody to see that there are thirteen hundred of God's children here suffering, sometimes going hungry, going to the dark and dreary nights wondering how this thing is going to come out. This is the issue. And we've got to say to the world, the issue is coming out. For when people get caught up with that which is right and they are willing to sacrifice for it, they are winning.



ITESO, Universidad
Jesuita de Guadalajara

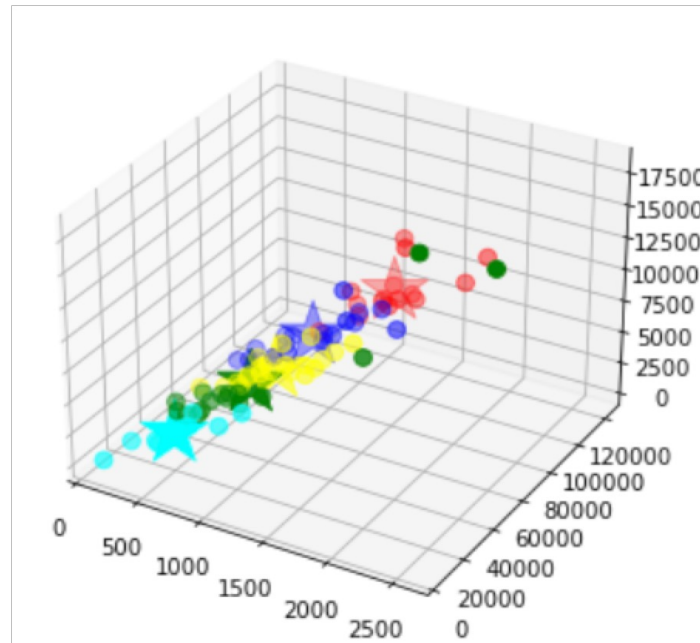
PREPARACIÓN DE DATOS

- Se crea un diccionario con palabras que no aportarán significado por sí solas, esto con ayuda de la librería de NLTK, la cuál cuenta con un dataset llamado *stopwords*
- Con ayuda de esta misma librería se genera un corpus con todos los archivos de los discursos
- Se obtienen, de cada uno, las 20 palabras más utilizadas, posteriormente se mezclan en una lista general, incluyendo el conteo de apariciones totales y la cantidad de archivos en los que aparecen
- Se genera un score con base en la totalidad de palabras, conteo de palabras excluyendo las incluidas en *stopwords* y el conteo de las palabras contenidas en la lista de las más utilizadas.
- Con los datasets generados se procede al cálculo de KNN y k-means con ayuda de la librería de *SciKit-Learn*



CONCLUSIONES Y RESULTADOS

El modelo realizado con KNN no pudo procesar de manera correcta la información, por lo que genero respuestas de 30% de confiabilidad, sin embargo, el modelo k-means fue exitoso en agrupar la información según la clasificación de los discursos, por lo que se concluye que fue posible encontrar características estadísticas que los describen.



Abstract geometric lines in the top-left corner of the slide, consisting of several thin, dark grey lines that intersect to form various polygonal shapes.

GRACIAS!