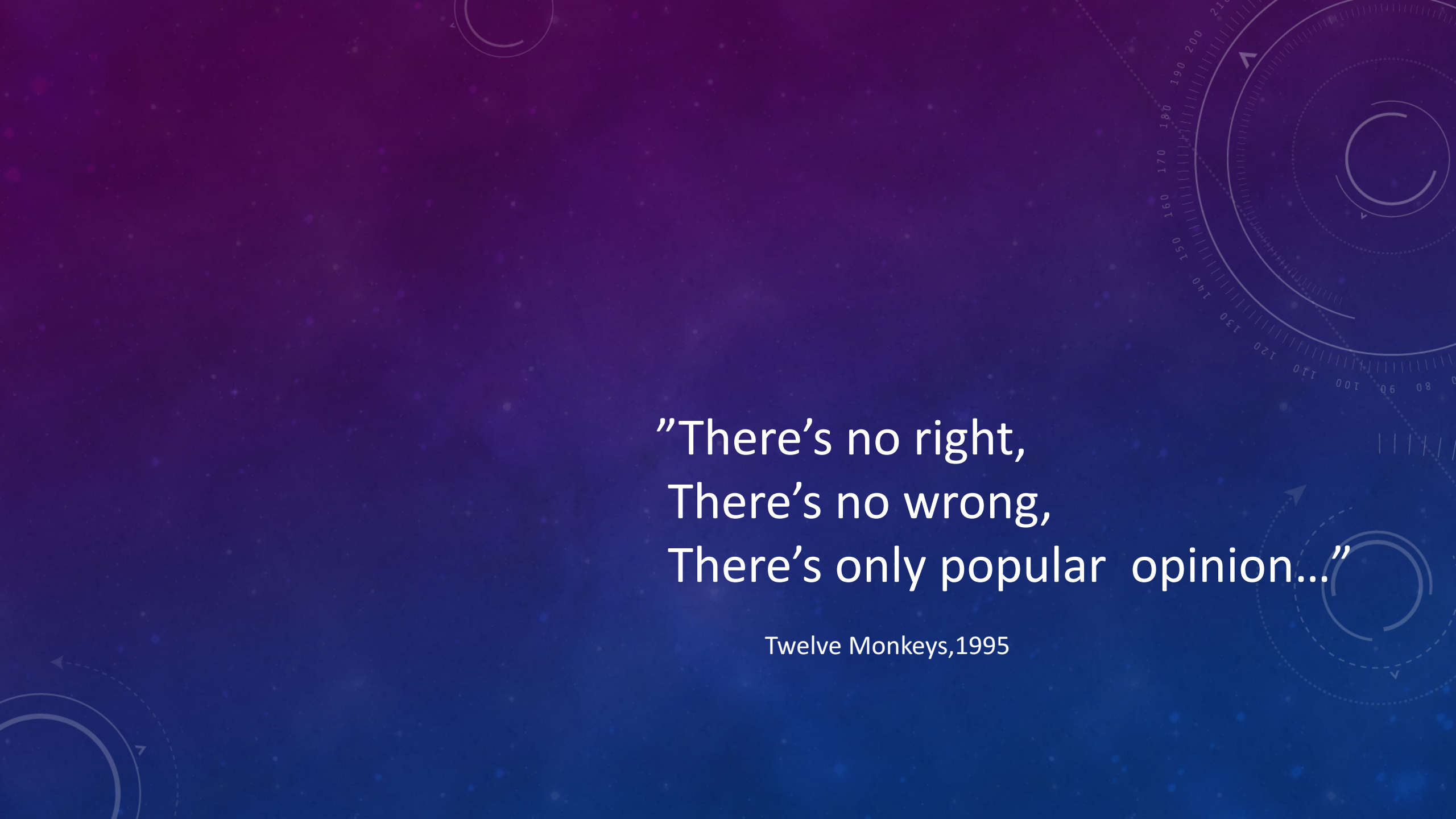




ANÁLISIS DE SENTIMIENTOS

PROCESAMIENTO DEL LENGUAJE NATURAL



“There’s no right,
There’s no wrong,
There’s only popular opinion...”

Twelve Monkeys, 1995

El análisis de sentimientos (también conocido como minería de opiniones), es el proceso de determinar el tono emocional que hay detrás de una serie de palabras y se utiliza para intentar entender las actitudes, opiniones y emociones expresadas en un texto. Es extremadamente útil en la monitorización de minería de datos, opiniones del consumidor, tendencias en redes sociales ,etc.

Introducción

OBJETIVO:

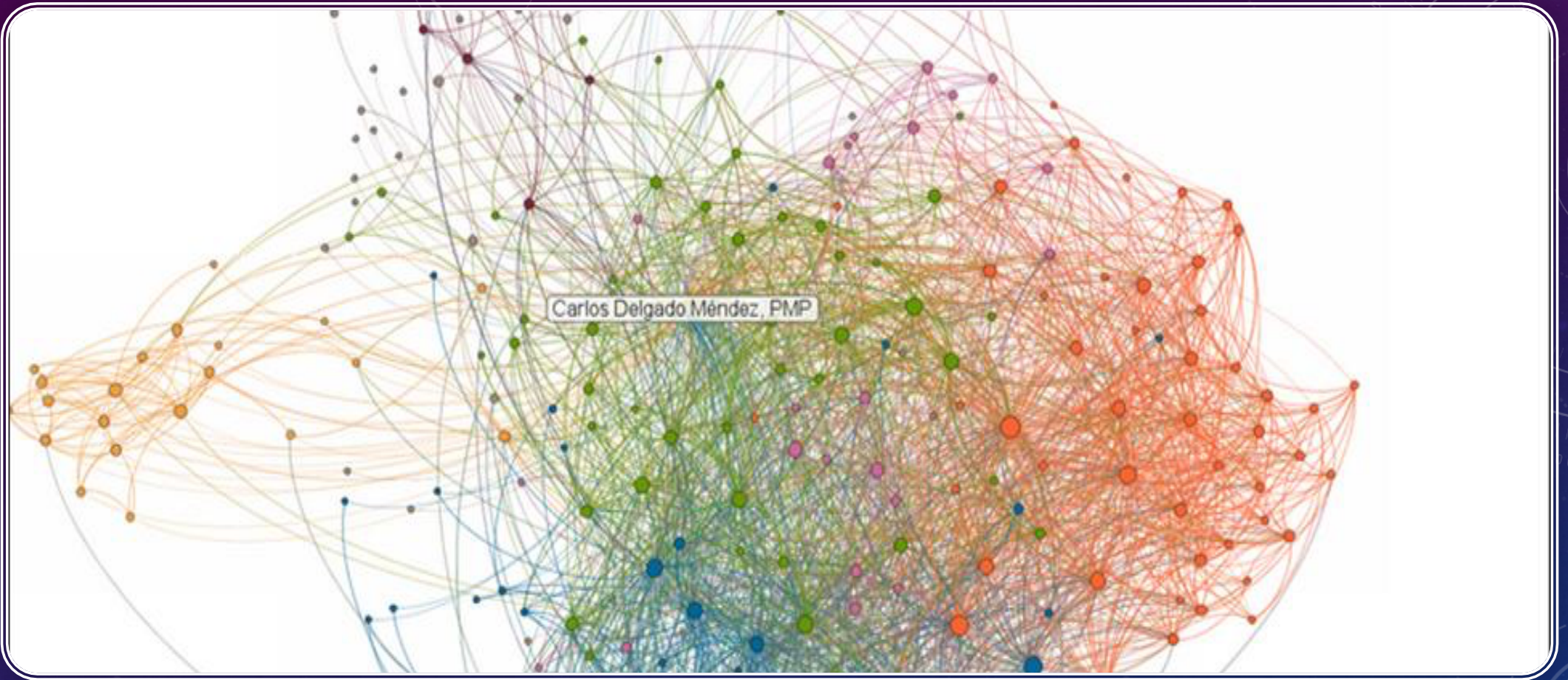
- Identificar y extraer información de textos, mediante el procesamiento computacional , (Análisis de Sentimientos), que tiene como objetivo principal determinar la “polaridad” del texto .
- Utilizando el algoritmo de Clustering Jerárquico .
- Y simulándolo a través de una pequeña aplicación.

ANÁLISIS DE TEXTO

A grandes rasgos el procesamiento de texto es:

- Tokenización.
- Stemming.
- Stopwords.
- Conversión de mayúsculas a minúsculas.





CLASIFICACIÓN DE TEXTOS

Algoritmo de Clustering

1.MATRIZ DE DATOS.

- Crear matriz de frases por palabras

¡Rosa no presentida que quitara
a la rosa la rosa que le diera
a la rosa le diera la rosa!

- Contadores: El numero de veces que la palabra aparece en la frase.

diera	la	le	no	presentida	que	quitara	rosa
0	0	0	1	1	1	1	1
1	2	1	0	0	1	0	2
1	2	1	0	0	0	0	2

2.DISTANCIAS .

- Distancia Euclidea.

$$d_{ij} = \sqrt{\sum_{p=1}^k (x_{ip} - x_{jp})^2}$$

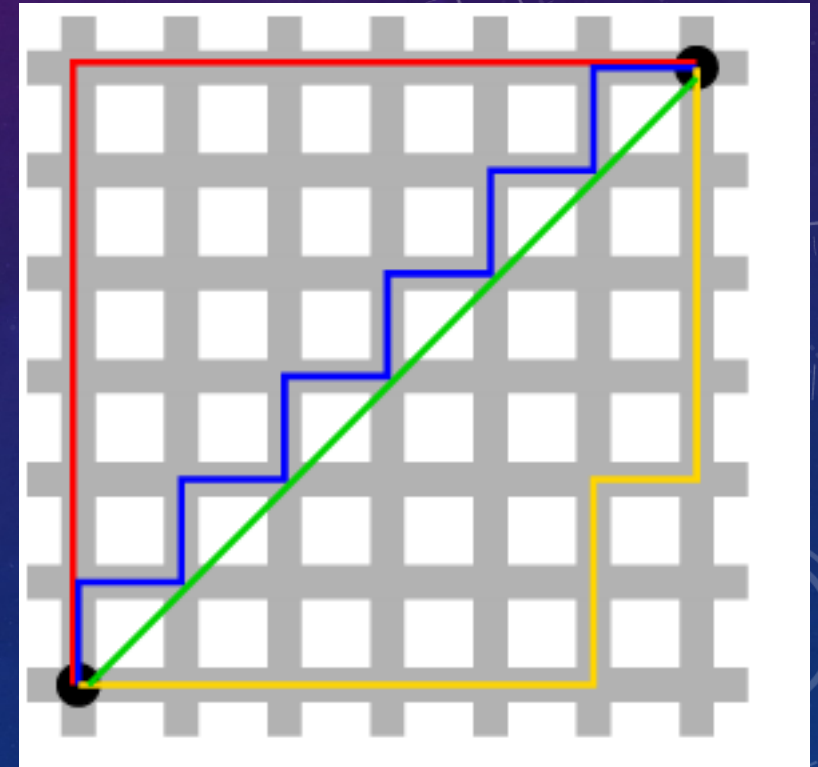
- Matriz de Similitudes / Distancias

$$d_E(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

	A	B	C	D	E
A	0				
B	1	0			
C	5	4.5	0		
D	8.5	7.8	3.6	0	
E	7.2	6.7	2.2	2	0


2.1.POR QUE NO OTRAS DISTANCIAS?

- Geometría del taxista (Taxicab)
 - Herman Minkowski
 - También llamada distancia Manhattan
-
- Distancia Manhattan = 12 (roja, azul, amarilla)
 - Distancia Euclidea = 8.5 (verde)



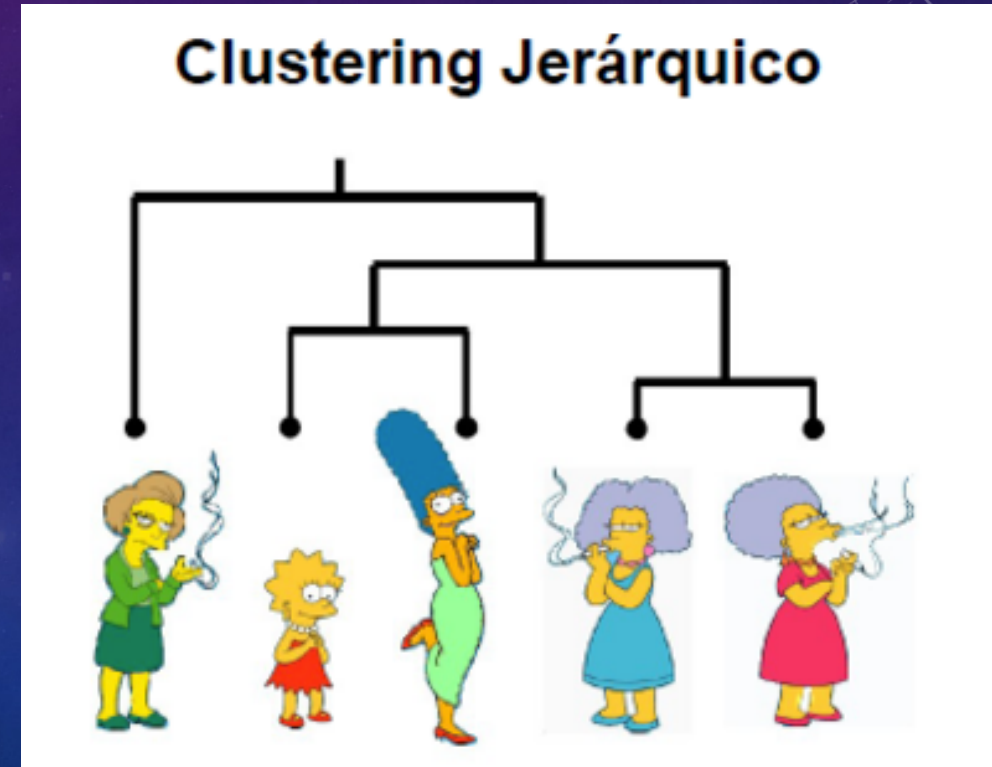
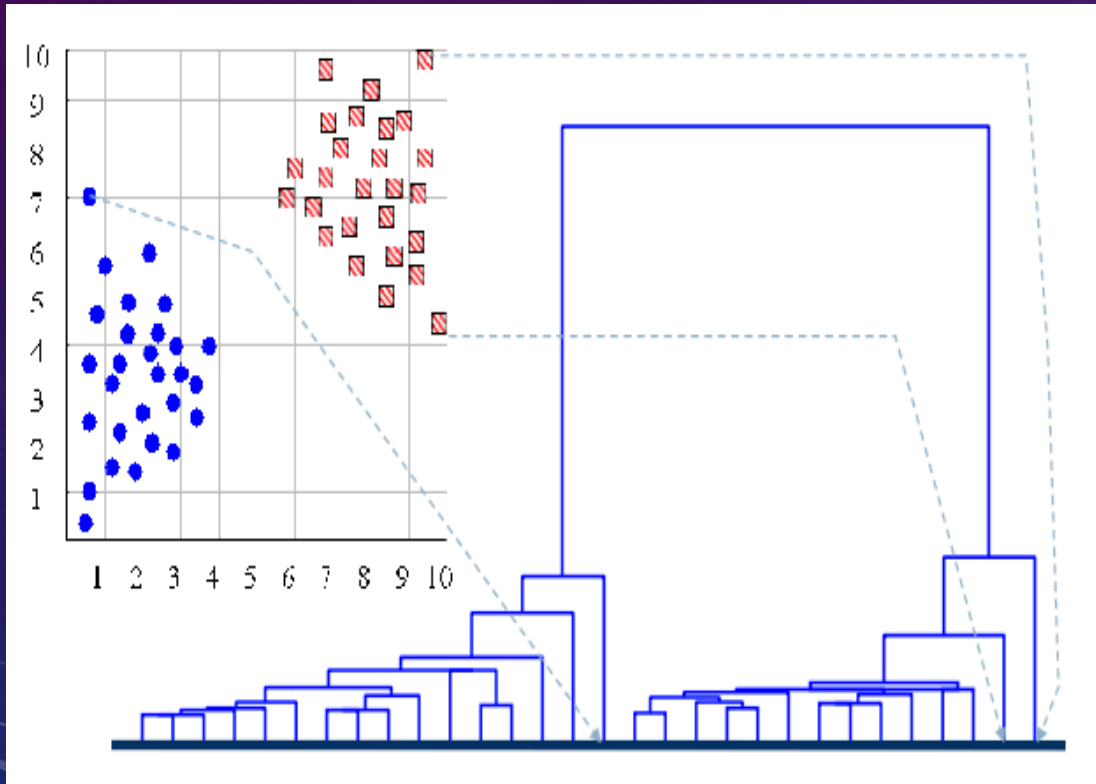
2.1.1. POR QUE NO OTRAS DISTANCIAS?

- Distancia de Chebyshev:
 - También conocida como distancia de tablero de ajedrez.
 - Mide el numero de movimientos que el rey deben de hacer para llegar a otra casilla en el tablero

	a	b	c	d	e	f	g	h	
8	5	4	3	2	2	2	2	2	8
7	5	4	3	2	1	1	1	2	7
6	5	4	3	2	1		1	2	6
5	5	4	3	2	1	1	1	2	5
4	5	4	3	2	2	2	2	2	4
3	5	4	3	3	3	3	3	3	3
2	5	4	4	4	4	4	4	4	2
1	5	5	5	5	5	5	5	5	1
	a	b	c	d	e	f	g	h	

3. CLUSTERING JERÁRQUICO

- Dendrograma: La similitud entre dos objetos viene dada por la "altura" del nodo común mas cercano



The background is a gradient of deep blue and purple, speckled with white dots resembling a starry sky. Overlaid on this are several faint, white geometric patterns. In the top right, there is a large circular scale with degree markings from 0 to 210 and concentric circles. In the bottom right, there are concentric circles with dashed lines and arrows indicating a clockwise direction. In the bottom left, there are also concentric circles with dashed lines and arrows. A small, partial circular scale is visible in the top left corner.

Aplicación...

Tecnologías Utilizadas:

- Python (versión 2.7.6) como mínimo
- Pickle
- Nltk
- Matplotlib
- `scipy.cluster.hierarchy`
- Numpy
- tokenize

Integrantes:

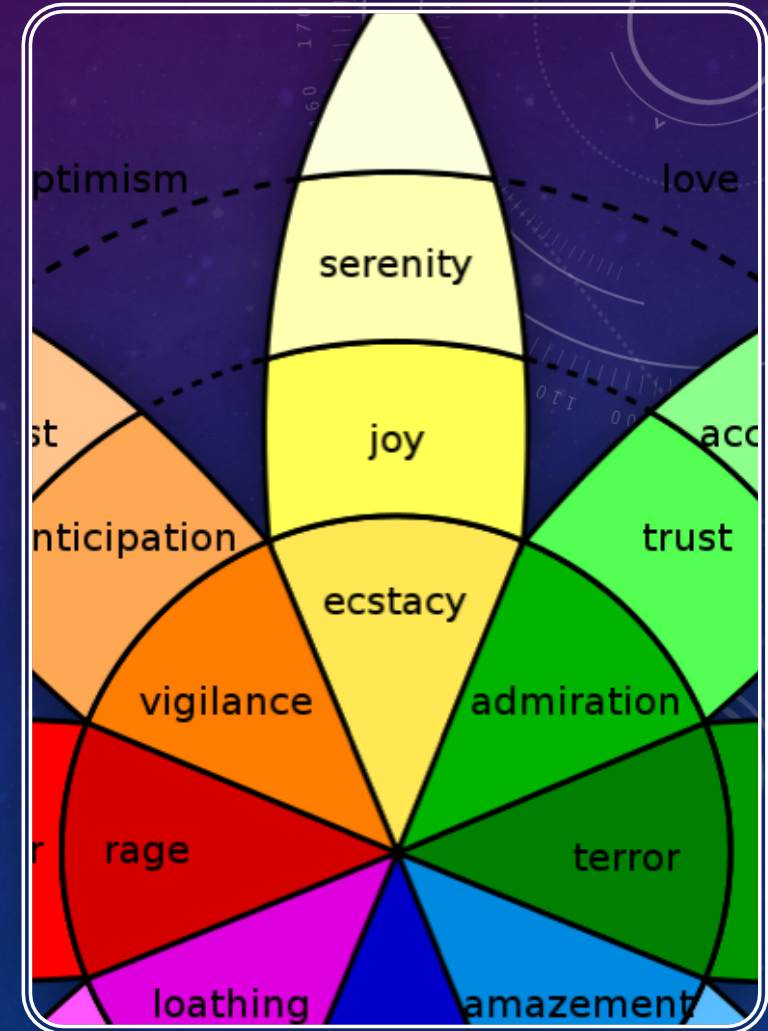
- Venegas Guerrero Fátima Alejandra.
- León Canto Ángel Efraín
- Vázquez García Palemón
- Hernández Chacón Carlos Alberto

The background is a gradient of deep blue and purple, speckled with white dots resembling a starry sky. Overlaid on this are several faint, white geometric patterns. In the top right, there is a large circular scale with degree markings from 0 to 210 and concentric circles. In the bottom right, there are concentric circles with dashed lines and arrows indicating a clockwise direction. In the bottom left, there are also concentric circles with dashed lines and arrows. A small, partial circular scale is visible in the top left corner.

Gracias !!!

BIBLIOGRAFÍA:

- *<http://pdln.blogspot.mx/2014/01/analisis-de-sentimientos-un-algoritmo.html>
- *<http://materias.fi.uba.ar/7500/Dubiau.pdf>
- *<http://isa.umh.es/asignaturas/crss/tutorialWEKA.pdf>
- *<http://materias.fi.uba.ar/7500/Dubiau.pdf>
- *<https://upcommons.upc.edu/bitstream/handle/2117/82434/113257.pdf?sequence=1&isAllowed=y>
- *<https://www.python.org>
- *<https://dlegorreta.wordpress.com/category/de-lo-concreto-a-lo-abstracto/nlp-con-python-y-r-project/>
- *<https://joernhees.de/blog/2015/08/26/scipy-hierarchical-clustering-and-dendrogram-tutorial/#Perform-the-Hierarchical-Clustering>
- *<https://pybonacci.es/2015/11/24/como-hacer-analisis-de-sentimiento-en-espanol-2/>



BIBLIOGRAFÍA

*https://es.wikipedia.org/wiki/Distancia_de_Chebyshev

*https://es.wikipedia.org/wiki/Geometría_del_taxista

*<http://wpd.ugr.es/~bioestad/guia-spss/practica-8/>

*<https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.spatial.distance.pdist.htm>

*https://www.cs.us.es/~fran/curso_unia/clustering.html

