



Gobierno de  
**México**



# Luis Guillermo Ruiz Velázquez

---

Infotec Aguascalientes

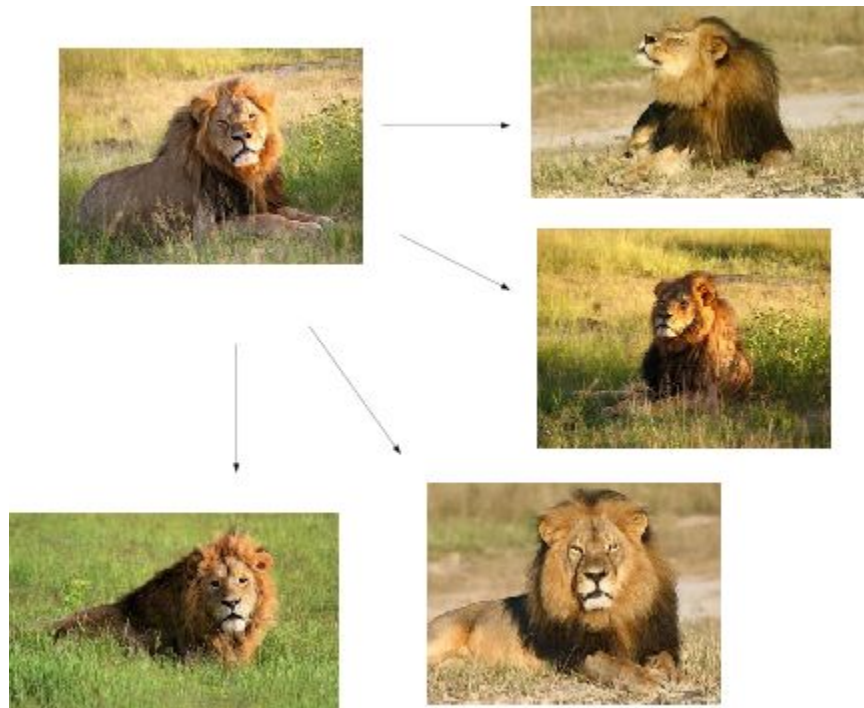


# Recuperación de Información

¿Qué es eso?

Dada una base de datos y una consulta, encontrar los elementos de la base de datos **más parecidos** a la consulta. No hay una definición universal de similaridad pues depende del problema.

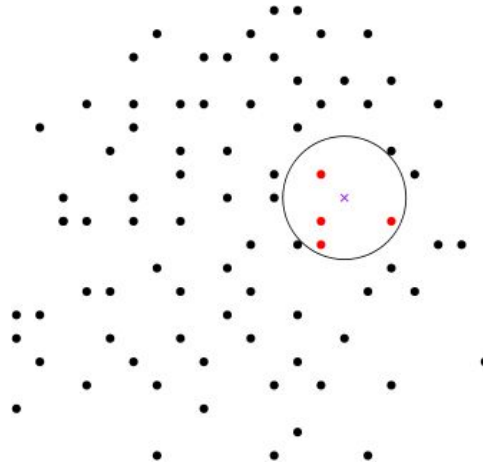
# Recuperación de Información



# Recuperación de Información

En un espacio métrico  $(U, d)$ , dado un subconjunto finito  $S \subset U$  y una consulta  $q \in U$  debemos encontrar todos los elementos de  $S$  tales que su distancia a  $q$  es menor o igual que  $r$ .

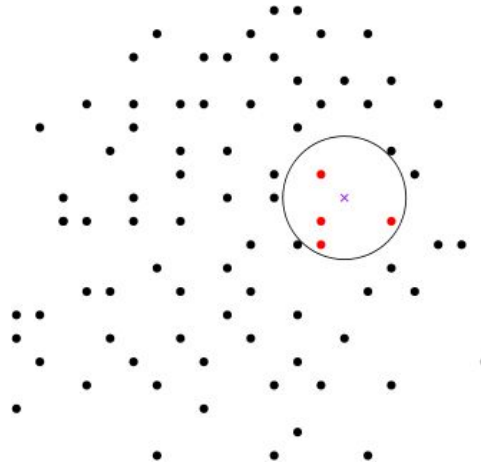
$$B_r(q) = \{x \in S \mid d(q, x) \leq r\}$$



# Recuperación de Información

En un espacio métrico  $(U, d)$ , dado un subconjunto finito  $S \subset U$  y una consulta  $q \in U$  debemos encontrar todos los elementos de  $S$  tales que su distancia a  $q$  es menor o igual que  $r$ .

$$B_r(q) = \{x \in S \mid d(q, x) \leq r\}$$



Este problema se puede resolver trivialmente usando búsqueda secuencial.

# Recuperación de Información

La forma de acelerar las consultas se logra mediante el uso de un índice de los datos.

Crear un índice puede ser muy costoso pero acelera en mucho las consultas.

# Recuperación de Información

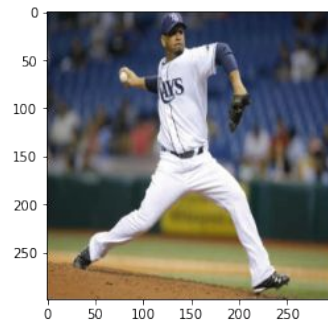
La forma de acelerar las consultas se logra mediante el uso de un índice de los datos.

Crear un índice puede ser muy costoso pero acelera en mucho las consultas.

¿Cuándo conviene construir un índice?

Cuando el número de consultas que se esperan hacer es muy grande.

# Image caption



a baseball player is about to  
throw a ball



a young girl is looking at a  
refrigerator



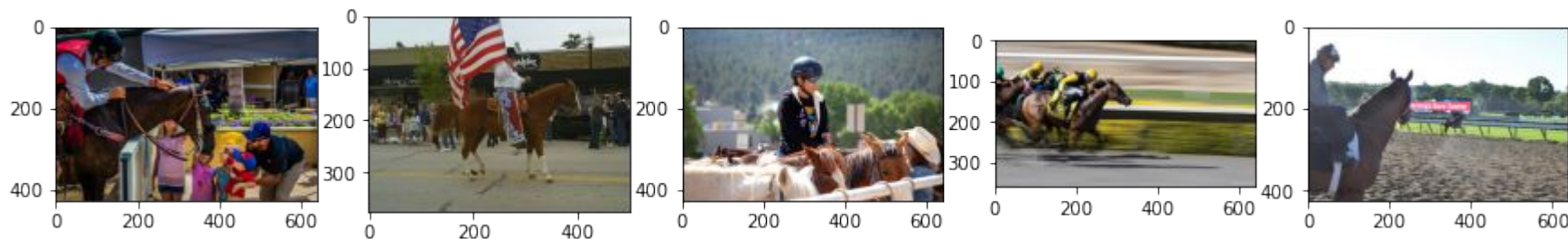
# Procesamiento de Lenguaje Natural

Usar modelos de lenguaje para clasificar texto: Análisis de Sentimiento, Detección de Discurso de Odio, medir los niveles de toxicidad de un mensaje, etc.

Usar encajes de texto para hacer búsquedas por contenido.

# Procesamiento de Lenguaje Natural

a person riding a horse



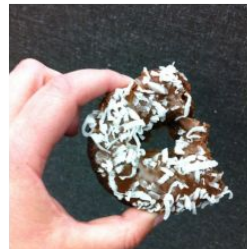
# Procesamiento de Lenguaje Natural



-



+



=

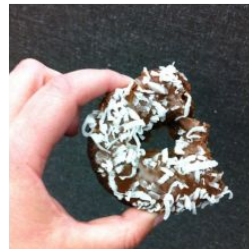
# Procesamiento de Lenguaje Natural



-



+



=

1



2



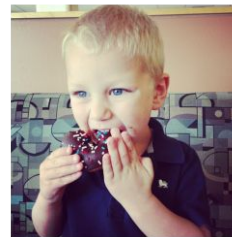
3



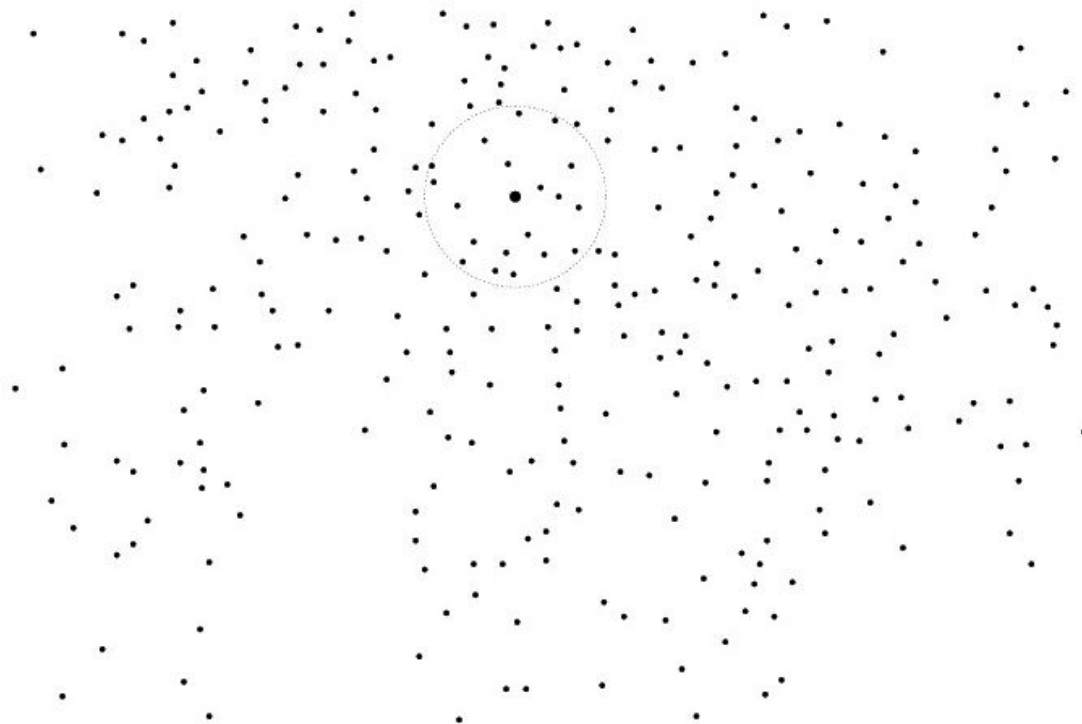
4



5

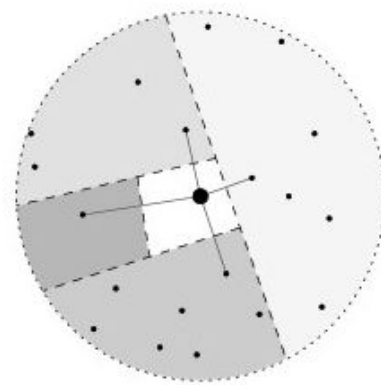
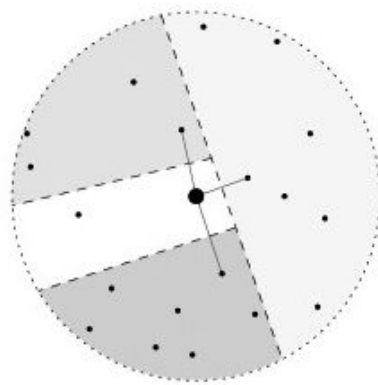
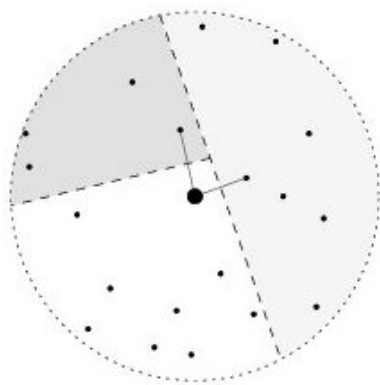
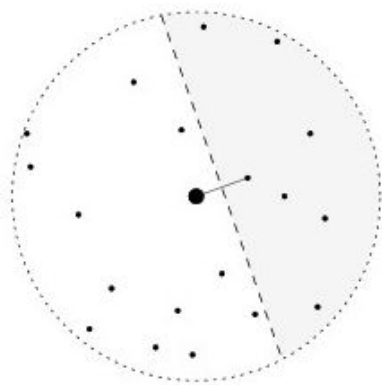


# Half Space Proximal

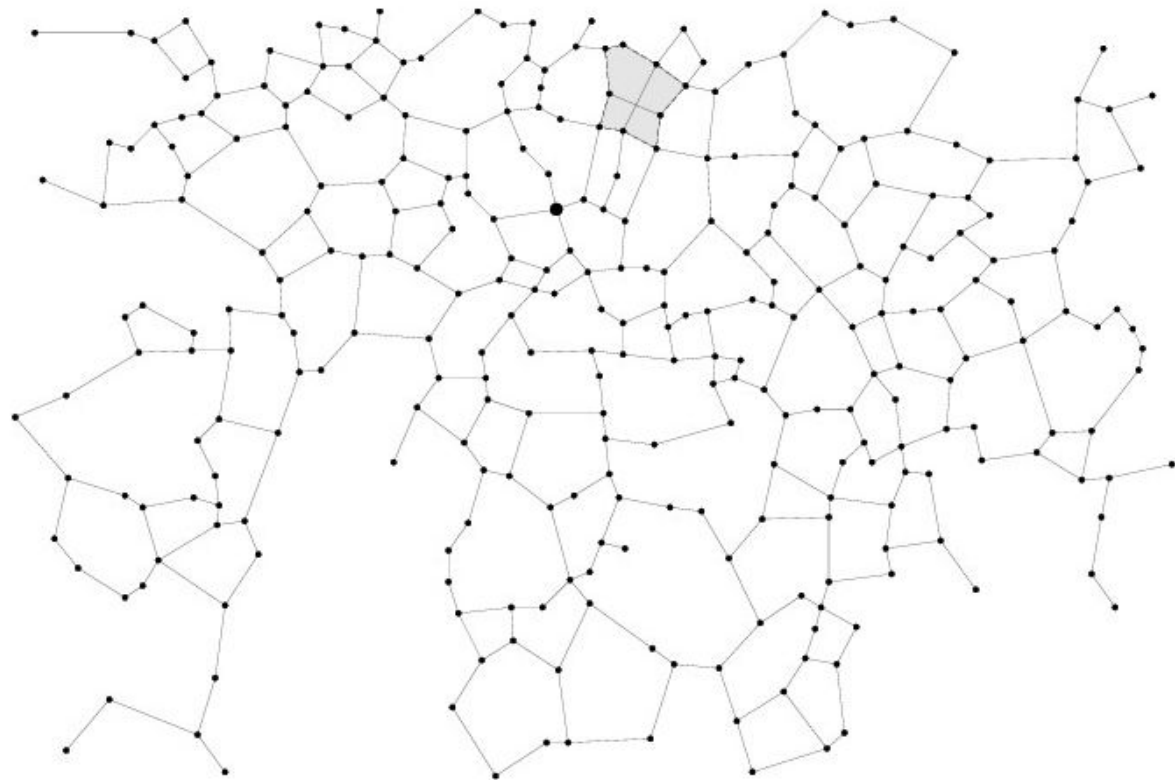


# Half Space Proximal

Figure 1: Half Space Proximal



# Half Space Proximal



# Half Space Proximal

Aplicaciones:

Recuperación de Información.

Medir la dimensión intrínseca.

Clasificación tipo k-vecinos.

Problema de empaquetado.

Clusters.

Redes de sensores.



# Gracias

[luiz.ruiz@infotec.mx](mailto:luiz.ruiz@infotec.mx)