



Escuela Profesional de
Ciencia de la Computación

Tópicos en Base de Datos

Ana Maria Cuadros Valdivia

Universidad Nacional de San Agustín de Arequipa

2020/Semestre Impar

Unidad 1: Extracción del Conocimiento en Base de Datos

- 1 Introducción
 - Motivación
 - Problemas
- 2 Propuestas
 - KDD
- 3 Referencias

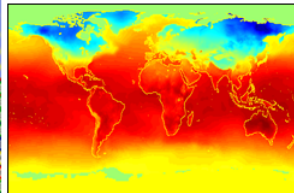
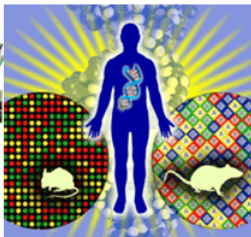
Motivación: Punto de Vista Comercial

- Grandes cantidades de datos son coleccionados y almacenados por empresas, corporaciones, etc.
 - Datos de comercio electrónico,
 - Datos de navegación en internet
 - Datos de compras de clientes en grandes tiendas por departamentos, supermercados
 - Datos de transacciones bancarias, o tarjetas de crédito
- Computadoras mas baratas y mas potentes
- Precisión de la competencia .



Motivación: Punto de Vista Científico: Med., Bio., Ing.

- Datos coleccionados y almacenados a velocidades enormes (TB/hora)
 - Sensores remotos en satélites,
 - Telescopios
 - Microarrays generando datos de expresiones de genes
 - Simulaciones científicas generando terabytes de datos



Motivación: Punto de Vista Social.

- Cambios en TICs crean nuevos hábitos de trabajo, de almacenamiento y de comunicación y difusión

Motivación: Punto de Vista Social.

● Noticias

Coronavirus: el trágico balance de muertos por la pandemia del nuevo coronavirus supera los 100.000

Redacción
BBC News Mundo



🕒 10 abril 2020

Ese era el recuento de la Universidad Johns Hopkins hasta este viernes, cuando se habían confirmado además 1.650.000 casos.

El virus, que además ha obligado a virtualmente paralizar el mundo y ha golpeado duramente la economía a nivel global, se ha cebado particularmente con España, Estados Unidos e Italia, además de China, donde fue detectado inicialmente.

- **Centroamérica, la región de récords y contrastes en sus estrategias y resultados frente al coronavirus**



El coronavirus ha matado a más de 100.000.



Motivación: Punto de Vista Social.

• Redes Sociales

The screenshot shows a Twitter search interface. At the top, there's a search bar with 'Vizcarra' entered. To the right are buttons for 'Iniciar sesión' and 'Regístrate'. Below the search bar, the results are categorized by 'Destacado', 'Más reciente', 'Personas', 'Fotos', and 'Videos'. The 'Personas' tab is selected, showing a list of users. The first user is Martin Vizcarra (@MartinVizcarraC), President of the Republic of Peru, with a 'Seguir' button. The second is José Chino Vizcarra (@ch_vizcarra), with a bio listing various football clubs and a 'Seguir' button. The third is Joan Vizcarra (@vizcarrafanclub), a fan profile, also with a 'Seguir' button. Below the list is a 'Ver todo' link. At the bottom, a tweet from 'Pasquín De Miércoles' is partially visible. On the right side of the page, there are three sections: '¿Eres nuevo en Twitter?' with a 'Regístrate' button, 'Filtros de búsqueda' with checkboxes for 'Personas' (selected), 'Ubicación' (selected), and 'Búsqueda avanzada', and 'Tendencias para ti' with a settings icon.

← Q Vizcarra

Iniciar sesión Regístrate

Vizcarra ...

Destacado Más reciente Personas Fotos Videos

Personas

 **Martin Vizcarra** ✓
@MartinVizcarraC
Presidente de la República del Perú. [Seguir](#)

 **José Chino Vizcarra**
@ch_vizcarra [Seguir](#)
R Central, Zacatepec, LDU Quito, GyE La Plata, Dvo Tachira, Rangers de Talca, San Martin SJ, Boca Unidos ctes, F.C.O, Platense, Tr. Suarez, Boca Unidos Ctes.

 **Joan Vizcarra**
@vizcarrafanclub [Seguir](#)
Perfil dedicado a los fans de Joan Vizcarra, uno de los mejores caricaturistas del mundo. Dibujante entre otras de @elJueves, @interviu_ y @lesportiucat.

[Ver todo](#)

 **Pasquín De Miércoles** @PasquinDeM · 9h
¿Que nos pasa a los peruanos?
Están más preocupados defendiendo a **Vizcarra** y por eso callan las

¿Eres nuevo en Twitter?
Regístrate ahora para obtener tu propia cronología personalizada.
[Regístrate](#)

Filtros de búsqueda

Personas
De cualquiera ☒
Personas a las que sigues ☐

Ubicación
En cualquier lugar ☒
Cerca de ti ☐

[Búsqueda avanzada](#)

Tendencias para ti ⚙️
Tendencia en Perú
Vizcarra ▼

Motivación: Punto de Vista Social.

- Salud



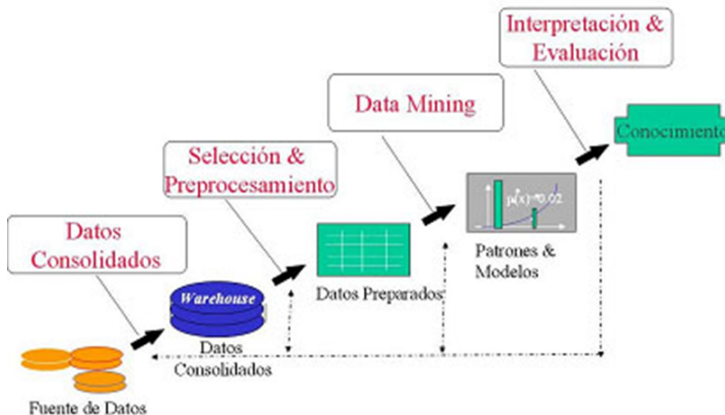
Multidimensionalidad

- Datos.
 - Son muy variados (imágenes, videos, texto, etc.)
 - Son demasiado complejos (muchas dimensiones)
 - La colección de datos es muy grande.
 - Muchas relaciones
- Estudiar sistemas naturales/artificiales implica manipular una gran cantidad de datos.
- Técnicas tradicionales no apropiadas para analizar tales datos:
 - ruidos y gran dimensionalidad .

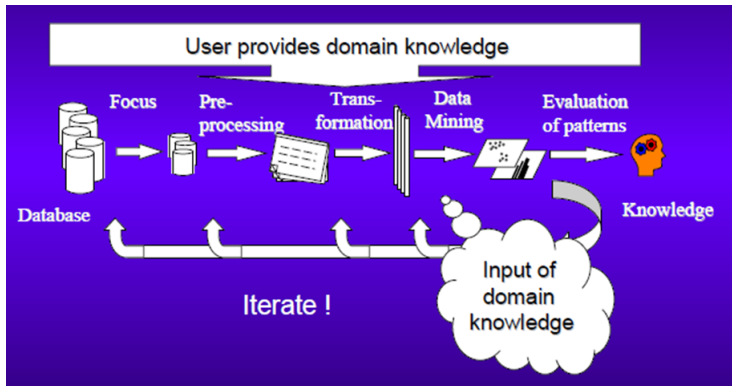
Procesamiento - Escalamiento

- Exceso de datos, de naturaleza compleja:
 - Enormes volúmenes, dimensionalidad alta, diferentes tipos de datos, organizaciones, etc.
- ¿Cómo tratar tanto volumen y variedad?
 - ultrapasan la habilidad técnica y la capacidad humana para su interpretación
- ¿Cómo saber lo que realmente es relevante?
- ¿Cómo transformar datos en información para la toma de decisiones?.

Descubierta del Conocimiento KDD



Proceso del Conocimiento Centrado en el Usuario



Bibliografia

- RHan, J., Kamber, M., Pei, J. (2011). Data mining: concepts and techniques. Elsevier.