

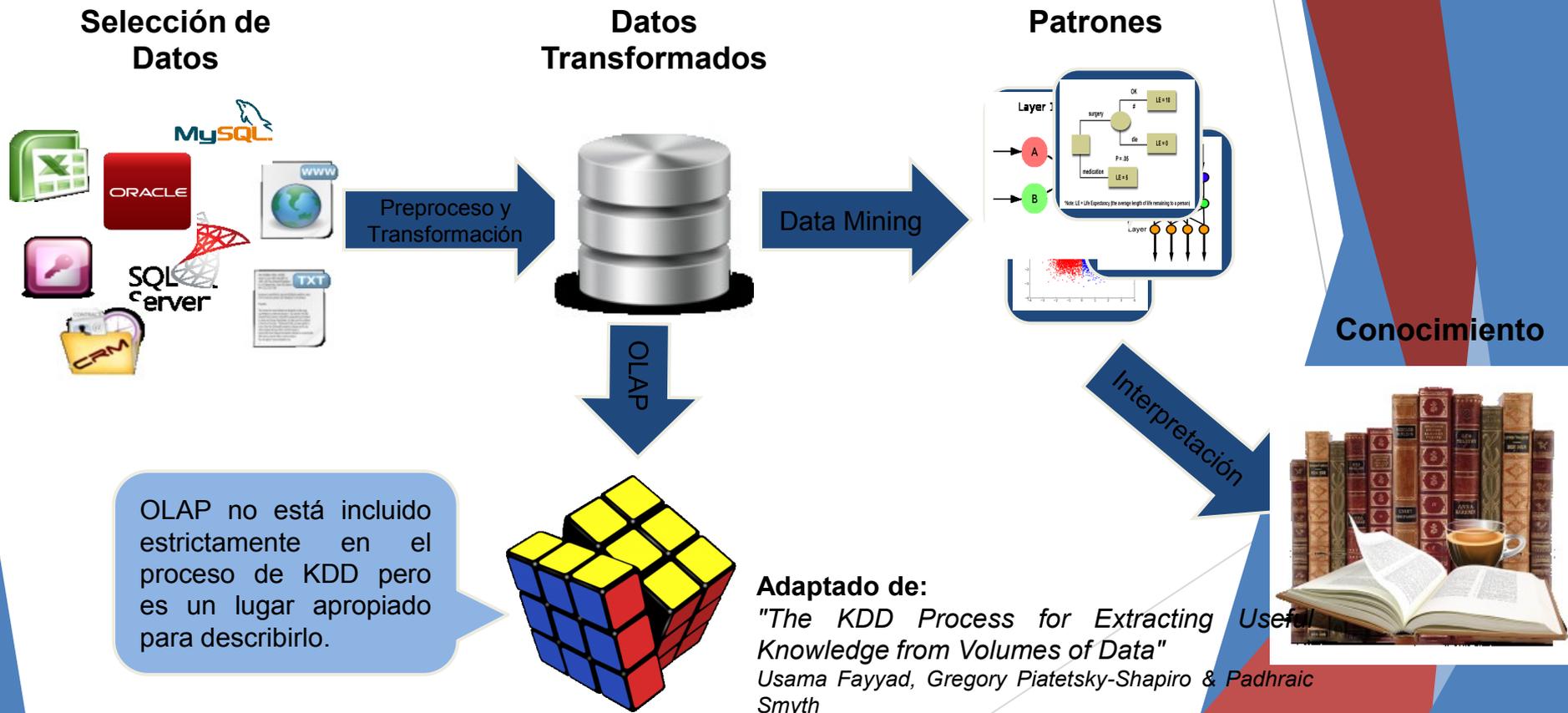


De los Datos al Conocimiento De los Minerales al Oro

Transformando los datos almacenados
en información útil para tomar
decisiones.

El Proceso de KDD

El proceso de descubrimiento de conocimiento en los datos, KDD (Knowledge Discovery in Data), se representa de la siguiente forma:



Procesamiento de los Datos

Selección de Datos

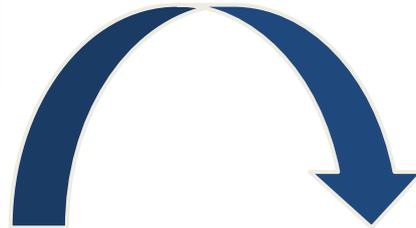
Se seleccionan datos una gran cantidad de datos de diversas fuentes.

- Planillas de Cálculo
- Bases de Datos
- Archivos Planos (TXT)
- Datos de la Web
- Datos del CRM

Preprocesamiento de los Datos

Se aplican técnicas para:

- Manejar datos incompletos y faltantes
- Hacer una limpieza de los datos
- Estandarizar y normalizar los datos
- Fundir diferentes fuentes de datos

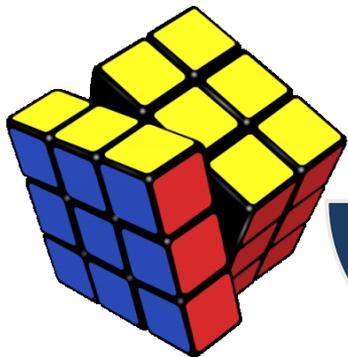


Datos Transformados

Se obtiene un repositorio de Datos (en algunos casos conocido como Data Warehouse) de donde se obtendrá la información necesaria para poder generar conocimiento.

OLAP (On-Line Analytical Processing)

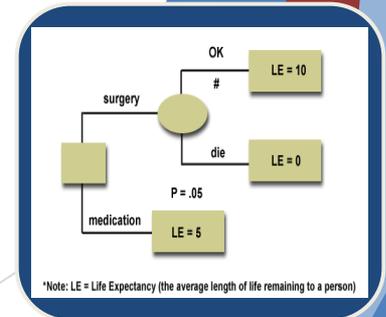
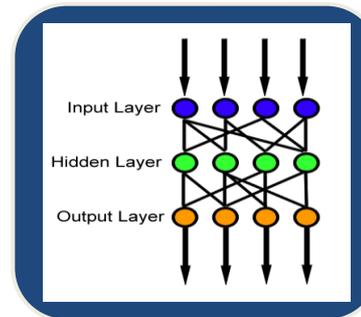
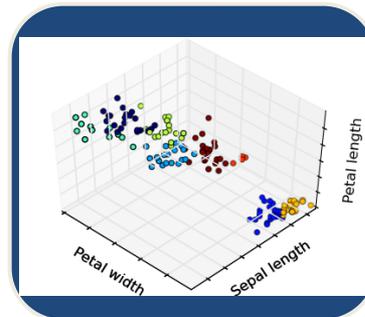
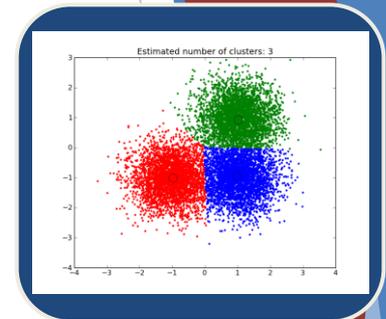
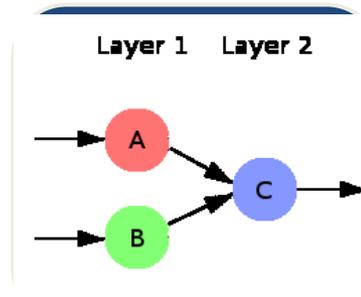
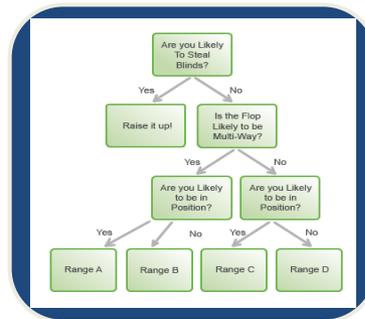
OLAP es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de Business Intelligence cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales (o Cubos OLAP) que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales (OLTP). Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares.



Data Mining

Data Mining es el proceso no trivial de identificación de patrones válidos, novedosos, potencialmente útiles y comprensibles en los datos.

Extracción de Patrones



*Note: LE = Life Expectancy (the average length of life remaining to a person)

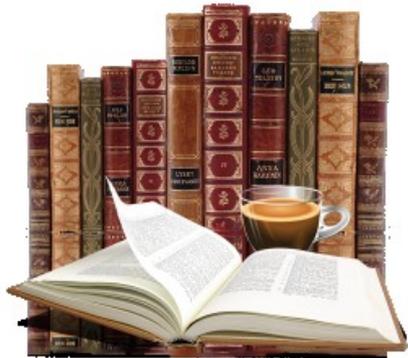


Conocimiento

El último paso del proceso es interpretar los patrones obtenidos para transformarlos en conocimiento. Generalmente para este paso se utiliza la ayuda de un experto en el dominio de los datos además del experto en Data Mining.



"El viernes por la tarde, los hombres jóvenes estadounidenses que compran pañales también tienen una predisposición a comprar cerveza"



"El Conocimiento Vale Oro"

Objetivo del Taller

- En este primer día del taller vamos a trabajar sobre distintos aspectos de este proceso