

Cápsula 1: Semántica de datos y sus tipos

Hola, bienvenidxs a una cápsula del curso Visualización de Información. En esta hablaré sobre semántica de datos y sus tipos.

Partiremos con un ejemplo. En pantalla podrán ver una secuencia de valores: la palabra "Santiago", el número 3, la letra "E", y las palabras "Nacimiento" y "Sí". Si esto es todo lo que tengo y si tuviera que adivinar que describe esta secuencia de valores, no podría contestarles, porque podría ser cualquier cosa.

Podría ser que describe el evento de un nacimiento, de una persona llamada "Santiago" el 3 de enero, y que al ser por cesárea se marca con un "Sí". Pero tal vez "Santiago" se refiere al lugar de nacimiento, pero eso solo lo pienso porque al ser de nacionalidad chilena, y saber que Santiago es el nombre de una ciudad. En realidad, Nacimiento también es el nombre de una ciudad chilena, en la región del Biobío. Solo al contar con el resto de los datos que le acompañan a esta secuencia, podemos darnos cuenta que en realidad se tratan de asientos en viajes en bus entre ciudades de Chile.

¿Qué pasó? Lo que hacía falta en el ejemplo era el contexto, o más bien, la **semántica de los datos**. Con semántica aquí nos referimos a su significado en el mundo real, a que hacen alusión o que describen. ¿Es la palabra "Santiago" el nombre de una persona, de una ciudad, de una marca, o de un meme? ¿Es un número en realidad una fila de asientos en un bus, la altura de una persona, representa un mes o un año? La semántica de datos entonces nos permite analizar y entender mejor también su tipo.

El **tipo**, en cambio, hace referencia a su interpretación estructural o matemática. Si lo que vemos es un valor numérico o no, si es una entidad, o incluso si es una relación entre entidades.

Es importante notar que el tipo de un dato se ve afectado por su semántica. Por ejemplo, un número de código postal, si bien es un número funciona como un valor identificador, más que un valor numérico. Ya que sumar o restar códigos postales no entrega información con sentido, ya que su interpretación no es numérica, si no que de identificación.

En las cápsulas de este tema revisaremos cómo catalogar no solo datos singulares, sino también conjuntos de datos completos, para así contar con un repertorio de conceptos de análisis en datos.

Primero, veremos **tipos de datos singulares**. Un dato singular puede ser uno de cinco tipos: un atributo, un ítem, un enlace, una posición o una grilla.

El más simple y reconocible es el **atributo**. Un atributo es cualquier propiedad específica que se puede medir, observar o registrar. Atributos son por ejemplo: monto salarial de una

persona, la altura de personas, precio de artículos, nivel de azúcar en la sangre, temperatura, etc...

Por otro lado, un **ítem** es una entidad individual y discreta que puede ser simple o compleja. Pueden ser personas, acciones de la bolsa, cafeterías, ciudades, genes, etc... Por lo general, un conjunto de datos contiene varios ítems descritos según atributos. Como una persona que posee varios atributos: su altura, peso, identidad de género, edad. Otro ejemplo fueron los asientos de viajes en bus del ejemplo inicial.

Luego está el dato de tipo **enlace** es una relación entre ítems. Representa una conexión entre dos entidades, como lo puede ser la amistad entre personas, viajes entre ciudades, relaciones laborales y jerarquías, etc...

Un dato de tipo **posición** se refiere a datos específicamente espaciales. Provee información de ubicaciones en espacio unidimensional, bidimensional o tridimensional. Pueden ser, longitud y latitud en un mapa, ubicación en un laberinto, o la posición 3D de alguna medición en un escáner médico.

Finalmente, también existe el tipo de dato de **grilla**. Este representa una estrategia para obtener una muestra de datos continua, lo que determina relaciones geométricas y topológicas entre celdas. Puede ser un poco difícil de entender en un principio.

Se utiliza generalmente para representar mediciones sobre objetos complejos, como en las mediciones de escáneres médicos que actúan sobre cuerpos tridimensionales, o mediciones meteorológicas a lo largo de la tierra.

Con eso termina el contenido de esta cápsula, en las próximas extenderemos nuestras definiciones sobre tipos de conjuntos de datos. Recuerda que si tienes preguntas, puedes dejarlas en los comentarios del video para responderlas en la sesión en vivo de esta temática. ¡Chao!