

UNIVERSIDAD PRIVADA DE TACNA
FACULTAD DE INGENIERÍA DE SISTEMAS
ESCUELA PROFESIONAL DE INGENIERÍA SISTEMAS



TRABAJO ENCARGADO
DATOS NO ESTRUCTURADOS
UNIDAD III

CURSO

Base de Datos II

DOCENTE

Mag. Patrick Cuadros Quiroga

INTEGRANTES

Tarqui Montalico, Risther Jaime - 2017057857

Limache Victorio, Víctor Piero - 2017057857

Sánchez Rodríguez, Bayron - 2017057857

Liendo Velasquez, Joaquín - 2017057857

Callata Flores, Rafael - 2017057857

TACNA - PERÚ

2020

Resumen

En este artículo se muestra la importancia de los Datos no Estructurados, dando las definiciones necesarias, también se muestra 3 formas en las cuales se pueden extraer Datos no estructurados, algunas aplicaciones conocidas que se le dan y se finaliza con un ejemplo no practico con una red social conocida.

Abstract

This article shows the importance of Unstructured Data, giving the necessary definitions, it also shows 3 ways in which unstructured Data can be extracted, some known applications that are given and ends with an impractical example with a known social network.

DATOS NO ESTRUCTURADOS

UNIDAD III

1. Introducción.

Antes de hablar sobre datos no estructurados, debe comprender qué son los datos estructurados. Cuando hablamos de datos estructurados nos referimos a la información que habitualmente se encuentra en la mayoría de bases de datos. Son archivos de tipo texto que generalmente se muestran en filas y columnas con títulos. Son datos que todas las herramientas de minería de datos pueden ordenar y procesar fácilmente. Podríamos verlo como si se tratara de un archivador perfectamente organizado donde todo está identificado, etiquetado y de fácil acceso.

2. Objetivo.

Mostrar en este artículo que existe toda una serie de datos "no tradicionales" (imágenes, audio, texto, etc.) que pueden ser analizados y de los que cuales se puede extraer valor.

3. Autores.

- Limache Victorio, Víctor Piero
- Tarqui Montalico, Risther Jaime
- Sánchez Rodríguez, Bayron
- Liendo Velasquez, Joaquín
- Callata Flores, Rafael

4. Desarrollo.

a) ¿QUÉ SON LOS DATOS NO ESTRUCTURADOS?.

En su definición más básica, simplemente significa cualquier forma de datos que no encaja fácilmente en un modelo relacional o un conjunto de tablas de base de datos desestructurado.

Los datos no estructurados, generalmente son datos binarios que no tienen estructura interna identificable. Es un conglomerado masivo y desorganizado de varios objetos que no tienen valor hasta que se identifican y almacenan de manera organizada.

Una vez que se organizan, los elementos que conforman su contenido pueden ser buscados y categorizados (al menos hasta cierto punto) para obtener información.

Aunque parezca increíble, la base de datos con información estructurada de una empresa, ni siquiera contiene la mitad de la información que hay disponible en la empresa lista para ser usada. El 80 % de la información relevante para un negocio se origina en forma no estructurada, principalmente en formato texto.

b) **TIPOS DE DATOS NO-ESTRUCTURADOS:**

Entre los distintos tipos de datos no estructurados tenemos:

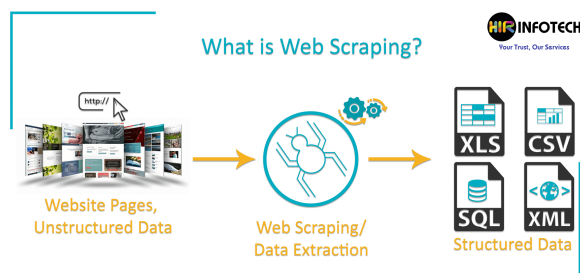
- 1) Correos electrónicos
- 2) Archivos de procesador de texto como Word
- 3) Archivos PDF
- 4) Hojas de cálculo como Excel.
- 5) Imágenes digitales como formatos bmp, tiff
- 6) Vídeo como mp4, avi
- 7) Audio como mp3
- 8) Publicaciones en redes sociales.
- 9) Presentaciones como PowerPoint

Mirando esa lista, te podrías preguntar qué tienen en común estos archivos. Se trata de archivos que pueden ser almacenados y administrados sin que el sistema tenga necesidad de entender el formato del archivo. Al no estar organizado el contenido de estos archivos, estos datos suelen ser almacenados en carpetas locales en las redes de las empresas o en la nube como Dropbox, Google drive o SharePoint.

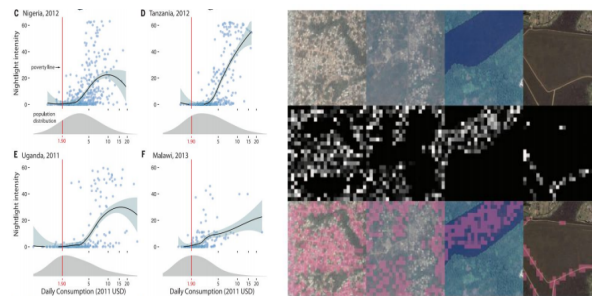
c) **FORMAS DE EXTRACCIÓN DE DATOS NO ESTRUCTURADOS:**

Entre los métodos de extracción de datos no estructurados tenemos:

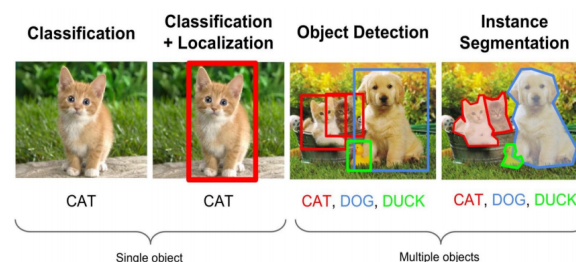
- 1) **Web Scraping (Rascado de datos)** Se podría definir como la técnica por la que un equipo de desarrolladores es capaz de rascar, escapear o liberar datos de páginas web de gobiernos, instituciones públicas u organizaciones para acceder a datos privados o públicos que puedan ser publicados o distribuidos en formato abierto. El problema es que la mayoría de los datos de interés están en formatos no reutilizables y poco transparentes como un PDF, por ejemplo.



- **Predicción de pobreza usando imágenes satelitales**



- **Clasificación de imágenes**



- **Ejemplo sencillo tratamiento datos no estructurados (redes sociales)**

Dada la variada naturaleza de los datos no estructurados, hay infinidad de posibles procesos relacionados con ellos.

El objetivo de este análisis de datos es conocer la percepción que existe sobre el precio de determinado producto en twitter.

- **Extracción:** Utilizando una clase de java (ejemplo twitter4j) leemos el feed de Twitter disponible en <https://twitter.com/search/realtime>. Añadimos a los campos disponibles calificaciones del tipo: idioma, localización geográfica.
- **Transformación:** Filtramos todos aquellos tuits que contengan el nombre del producto. Refinamos el filtro introduciendo campos del tipo (“precio”) + (“barato”, “caro”, “económico”, etc..) , teniendo en cuenta el idioma en el que se generan los tuits. Valorar la opción en base al volumen de obtener una muestra representativa de los datos extraídos y filtrados.
- **Volcado a BBDD :** Insertamos en una tabla el registro del tuit con la calificación identificada (idioma, localización geográfica).
- **Informes:** Creamos informe que permita realizar análisis por tiempo y campos de calificación. Hay que considerar que este informe puede ser actualizado en tiempo real.

5. Conclusiones

- Concluimos que no sólo los datos estructurados pueden ser analizados y que existen muchos más tipos de datos.
- Hay una gran disponibilidad de datos no estructurados.
- Actualmente existen múltiples técnicas para analizar distintos tipos de datos no estructurados.
- En la actualidad, todo lo que tenga que ver con los Datos no estructurados está creciendo exponencialmente.

6. Bibliografía

- Datos no Estructurados. Recuperado de: <https://sometet.mx/blog/que-son-los-datos-no-estructurados>
- Diferencia entre datos estructurados y no estructurados. Recueperado de: <https://www.kyocereworkspaces/insights-hub/articles/diferencia-entre-datos-estructurados-y-no-estructurados.ht>
- Herramientas de extracción de datos. Recuperado de: <https://bbvaopen4u.com/es/actualidad/de-extraccion-de-datos-para-principiantes-y-profesionales>
- Extraccion y analisis de datos no estructurados. Recuperado de: <http://www.eco.unc.edu.ar/f>