

de los datos a la información

formatos, extraer, limpiar, exportar...

formatos

- tabular - dispuestos en columnas y filas (excel?)
- .csv - comma separated values
- .tsv - tab separated values
- .xml - extensible markup language
- .json - javascript object notation

formatos

tabular - dispuestos en columnas y filas (excel?)

nombre	profesión	sueldo mensual
juan	dentista	5.000.000
maría	profesora	800.000
jose	diseñador	6.000.000
iván	diputado	10.000.000

formatos

.csv - comma separated values

los datos pueden tener comas (números o frases complejas)

Año, País, Habitantes, Frase Típica

1997, Chile, 16.000.000, aonde la viste, pavo

1998, Argentina, 36.000.000, “somos los mejores, víste?”

Año, País, Habitantes, Frase Típica, 1997, Chile, 16.000.000, aonde la viste, pavo, 1998, Argentina, 36.000.000, “somos los mejores, víste?”

formatos

.tsv - tab separated values

los datos pueden tener espacios

Año	País	Habitantes	Frase_Típica
1997	Chile	16.000.000	aonde la viste, pavo
1998	Argentina	36.000.000	“somos los mejores, víste?”

formatos

.xml - extensible markup language

extendida desde tags html, estos pueden ser creados
puede tener una estructura de datos más compleja
puede tener elementos anidados, como un árbol
basado en nodos, sub-nodos, y así...

```
<?xml version="1.0" encoding="UTF-8"?>
<pais>
  <nombre_pais>Chile</nombre_pais>
  <habitantes>16.000.000</habitantes>
  <capital>
    <capital_nombre>Santiago</capital_nombre>
    <capital_habitantes>6.000.000</capital_habitantes>
  </capital>
</pais>
```

ejemplo: <http://xml.tutiempo.net/xml/55760.xml>

formatos

.json - javascript object notation

muy popular y eficiente, pensado para browsers
puede tener una estructura de datos más compleja
puede tener elementos anidados, como un árbol
estructura key : value

```
var pais =  
{  
  "nombre_pais" : "Chile",  
  "habitantes" : "16.000.000",  
  "capital" :  
    {  
      "capital_nombre" : "Santiago",  
      "capital_habitantes" : "6.000.000"  
      "ubicación" : { "lat" : 33, "lon": 71 }  
    }  
}
```

ejemplo: https://dev.twitter.com/rest/reference/get/statuses/user_timeline

formatos

herramientas web

<https://shancarter.github.io/mr-data-converter/>

<http://codebeautify.org/xmlviewer>

<http://jsonviewer.stack.hu/>

<http://www.jsoneditoronline.org/>

<http://www.freeformatter.com/>

extraer de .pdf

tabula - <http://tabula.technology/>

abre una dirección local en nuestro browser.

extraer de web

google docs

abrir una hoja de calculo

colocar funcion y la dirección web correspondiente

```
=ImportHTML("https://es.wikipedia.org/wiki/Anexo:Goleadores_de_la_Copa_Mundial_de_Fútbol"; "table";1)
```

limpiar datos

google google refine, ahora openrefine

herramienta avanzada para limpieza y transformación de datos.
abre una dirección local

<http://openrefine.org/>