



UNIVERSIDAD **CATÓLICA**
de Colombia

Uso de Herramientas Tecnológicas para Data Analytics y Big Data

Diego Alberto Rincón Yáñez MSc
darincon@ucatolica.edu.co

Afiliada a la Federación Internacional de Universidades Católicas (FIUC)
www.ucatolica.edu.co

¡Ejercicio No.3!



UNIVERSIDAD CATÓLICA
de Colombia

- Problema: Extraer la mayor cantidad de meta-datos del Libro el quijote.
- URL: <http://www.elmundo.es/quijote/index.html>

- Grupos de 4 Personas
 - Ningún compañero cercano!!



Extracción de Datos



UNIVERSIDAD CATÓLICA
de Colombia

- Ejemplo Hecho en [Java](#)
- Ejemplo Hecho en [VisualBasic 6.0](#)

Fase 1



UNIVERSIDAD CATÓLICA
de Colombia

1. Datos a Catalogar?
2. Información a extraer?
3. Como extraer los datos de la fuente?

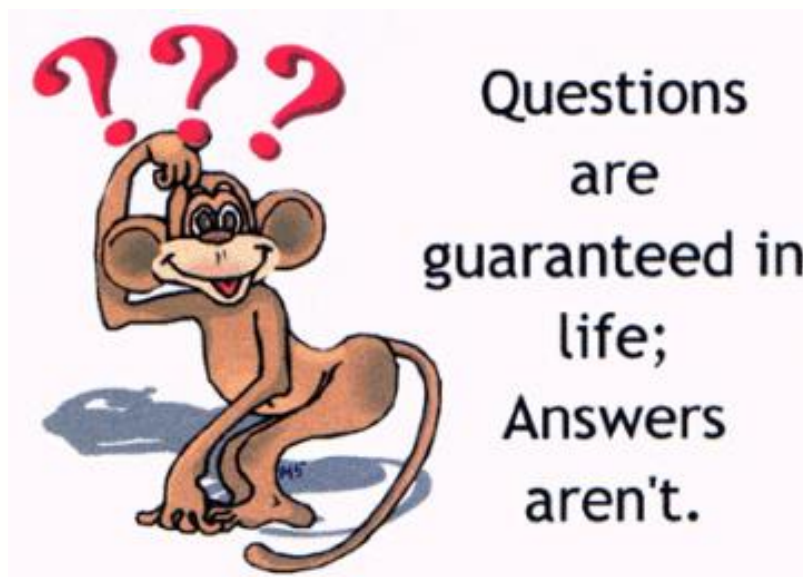
Producto: ¡Documento!



Conceptos para NoSQL

Mongo DB

- Colecciones (Collections)
 - Entidades (Entities)
 - Documentos (Documents)
 - Campos (Fields)



¿Preguntas?

Diego Alberto Rincón Yáñez MCSc.
Twitter: @d1egoprogram