

Introducción al Big Data

Ejercicio U11_E02

MasterD

1. EXPLORANDO FUENTES DE DATOS

Pasemos a otro ejercicio sencillo. No te descubrimos nada nuevo a estas alturas al decir que el interés y crecimiento de las tecnologías Big Data es producto del desbordante crecimiento de datos disponibles para analizar.

Hoy en día muchas empresas y organizaciones generan y manejan enormes cantidades de información en su día a día.

Pero para acceder a grandes conjuntos de datos y poder explorar y hacer nuestros pinitos con las diferentes tecnologías Big Data no necesitamos ser una de estas empresas.

Afortunadamente, existen gran cantidad de organismos y compañías que ofrecen conjuntos de datos de mucho valor y utilidad, de manera pública y gratuita a través de páginas y servicios web. Es lo que llamamos *open data*. Las principales fuentes de datos abiertos son las administraciones públicas y los centros de investigación, pero cada vez hay más empresas que publican parte de los datos que generan.

Cualquiera puede acceder y descargar estos datos, y utilizarlos simplemente para curiosear y examinarlos, o realizar estudios, o incluso desarrollar aplicaciones o servicios a partir de ellos.

El ejercicio que te proponemos aquí es que hagas un pequeño trabajo de búsqueda para encontrar algunos de estas fuentes de datos abiertos en internet.

Puedes clasificarlos en las siguientes categorías.

1.1. CATÁLOGOS O LISTADOS DE FUENTES DE DATOS ABIERTOS

1.2. CONJUNTOS DE DATOS PARA USO CIENTÍFICO, Y PARA MACHINE LEARNING Y CIENCIA DE DATOS

1.3. DATOS ABIERTOS DE ADMINISTRACIONES Y ORGANISMOS PÚBLICOS

1.4. DATOS ABIERTOS DE REDES SOCIALES

Puedes añadir más categorías si lo consideras necesario.