

Búsqueda y obtención de datos –Google Dataset Search

Otra fuente de datos utilizando herramientas de Google es una herramienta llamada “Google Dataset Search”. Esta herramienta es en realidad un agregador de fuentes de datos. Así que imaginen usar Google para buscar sitios y páginas que solo publiquen datos, pero que provengan de fuentes oficiales. Entonces esta herramienta trata de ver una multitud de sitios y luego obtener la información, los metadatos de estos sitios para hablar sobre los datos que publican.

Entonces, vamos a digitar “educación” en la barra de búsqueda de Google Dataset Search y en seguida damos Enter. Ahora vean, en el lado izquierdo se encuentran las fuentes de búsqueda de conjunto de datos donde encontró datos sobre “educación”. En el lado derecho, muestra la vista previa de toda la información asociada con esta base de datos en particular.

Entonces, en este caso, la búsqueda encontró estadísticas sobre educación en los datos abiertos del Banco Mundial y dice que ustedes pueden explorarlos en Kaggle, y luego les dice cuándo se actualizó este conjunto de datos, quién proporcionó este conjunto de datos. Existe una licencia asociada generalmente; en este caso es desconocida. Qué formatos están disponibles para descargar, la descripción, contextos y observaciones.

Y si continuamos, veremos que la búsqueda reúne conjuntos de datos de diferentes fuentes. Entonces, por ejemplo, este de Marruecos, podemos hacer clic aquí y ver que se actualizó por última vez el 7 de diciembre de 2018 y podemos explorar esto en “catalog.data.gov”. Y luego hacemos clic y vemos el sitio que publicó los datos y podemos leer más al respecto e incluso ver si hay enlaces para acceder a la base de datos como tal, o incluso hablar con las personas que de hecho administran este conjunto de datos para poder contactar a la autoridad que administra este conjunto de datos.

Lo mismo ocurre, por ejemplo, para este conjunto de datos de India en Kaggle. Podemos explorar esto en Kaggle, y luego, cuando abran la página web, les mostrará toda la información asociada con esto, la licencia, cómo pueden descargarlo, el contenido y todo lo demás asociado con esto.

Entonces Google Dataset Search es una nueva herramienta. Ha pasado un tiempo relativamente corto desde su lanzamiento, por lo que se centra solo en los conjuntos de datos publicados por diferentes fuentes en la web.