

Localizando e obtendo os dados - Google Dataset Search

Outra fonte de dados usando ferramentas do Google é uma ferramenta chamada "Google Dataset Search". Esta ferramenta é na verdade um agregador de fontes de dados. Então, imagine usar o Google para procurar sites e páginas que publicam somente dados, mas originados de fontes oficiais. Então, essa ferramenta tenta olhar para uma infinidade de sites e, em seguida, consegue as informações, os metadados destes sites para falar sobre os dados que eles publicam.

Então, vamos digitar "educação" na barra de pesquisa do Google Search Dataset e, em seguida, teclar Enter. Agora veja, no lado esquerdo está a pesquisa de conjuntos de dados lista fontes onde encontrou dados sobre "educação". No lado direito, ela mostra a pré-visualização de toda a informação que está associada a este banco de dados específico.

Então, neste caso, a pesquisa encontrou estatísticas sobre educação nos dados abertos do Banco Mundial e diz que você pode explorar isso no Kaggle, e depois te diz quando este conjunto de dados foi atualizado, quem forneceu este conjunto de dados. Há uma licença associada normalmente, neste caso é desconhecida. Quais são os formatos disponíveis para download, a descrição, contextos e observações. E se nós continuarmos, vamos ver que a pesquisa reúne conjuntos de dados de diferentes fontes.

Assim, por exemplo, este do Marrocos, podemos clicar aqui e ver que foi atualizado pela última vez em 7 de dezembro de 2018 e podemos explorar isso em "catalogue.data.gov". E, então, clicamos e vemos o site que publicou os dados e podemos ler mais sobre ele e até mesmo ver se há links para acessar o banco de dados em si, ou até mesmo falar com as pessoas que estão de fato administrando este conjunto de dados para poder entrar em contato com a autoridade que está administrando este conjunto de dados.

O mesmo vale, por exemplo, para este conjunto de dados sobre educação na Índia. Podemos explorar isso no Kaggle, e então quando você abrir a página web ele vai mostrar a você toda a informação que está associada a isto, a licença, como você pode baixá-lo, o conteúdo e todo o resto que está associado a isto.

Então, o Google Dataset Search é uma nova ferramenta. Faz relativamente pouco tempo que lançamos, e ela está focada apenas em conjuntos de dados que são publicados por diferentes fontes por toda a web.