

Tabla de repositorios de Datos Abiertos

Bases de datos				
Nombre de la página	Tipo	Idioma de la página	Descripción	Link de acceso
Blogger corpus	Base de datos	Inglés	"The Blog Authorship Corpus" es un repositorio de datos que contiene publicaciones recopiladas de 19,320 bloggers, pertenecientes a blogger.com, en agosto de 2004. En total incorpora 681,288 publicaciones y más de 140 millones de palabras, o aproximadamente 35 publicaciones y 7250 palabras por persona.	http://u.cs.biu.ac.il/~koppel/BlogCorpus.htm
Center for Disease Control and Prevention Data & Statistics (Centro de Control y Prevención de Enfermedades Datos y Estadísticas)	Base de datos	Inglés	Repositorio de datos abiertos del CDC que contiene datos de salud, enfermedades crónicas, cáncer, enfermedades cardiovasculares, entre otros varios de la categoría salud y medicina.	https://www.cdc.gov/datastatistics/
Computer Vision Laboratory (Laboratorio Computer Vision)	Base de datos	Inglés	"Computer Vision Laboratory" contiene 23 data sets con diferentes tipos de imágenes, dedicadas específicamente a la creación de nuevas tecnologías, como sensores de visión, modelos basados en la física para la visión y el desarrollo de algoritmos para la interpretación de escenas.	http://www.cs.columbia.edu/CAVE/databases/
EU Open Data portal (Portal de datos abiertos de la UE)	Base de datos	Inglés	El Portal de datos abiertos de la Unión Europea, proporciona acceso a una gama cada vez mayor de datos de las instituciones de la Unión Europea (UE) y otros organismos de la UE. Contiene datos de más de 70 instituciones, que abarcan temas relacionados a economía, empleo, ciencia, medio ambiente y educación.	https://data.europa.eu/euodp/en/data/
Financial times data	Base de datos	Inglés	"Financial times market data" es un repositorio que contiene una amplia base de datos de carácter financiero.	https://markets.ft.com/data/
FiveThirtyEight	Base de datos	Inglés	El repositorio de datos de FiveThirtyEight brinda el código detrás de algunos de artículos y gráficos de sus publicaciones. Es un sitio especializado para periodismo, con fuentes de datos en varias áreas como política, deporte, ciencia, economía, entre otros.	https://data.fivethirtyeight.com/
freeCodeCamp	Base de datos	Inglés	El repositorio de datos de freeCodeCamp contiene: conjuntos de datos y sus análisis, demos de proyectos basados en los datos de freeCodeCamp y enlaces a proyectos externos que involucran los datos de freeCodeCamp. Los datos pueden ser de interés para aquellos que buscan conjuntos de datos sobre: análisis de la red, análisis de redes sociales, analítica educativa, visualización de datos, desarrollo web basado en datos y bots.	https://github.com/freeCodeCamp/open-data
Global Financial data (GFD)	Base de datos	Inglés	"GFD" es una compañía que transcribió la mayor colección de archivos históricos en un formato accesible electrónicamente (datos abiertos). Brindan una cobertura completa de las acciones de EE. UU y el Reino Unido, desde la primera acción que se negoció en 1694 hasta la actualidad. Sus datos macro globales cubren 200 países a partir del año 1200.	https://www.globalfinancialdata.com/
Global Health Observatory - GHO (Observatorio global de salud)	Base de datos	Inglés	El repositorio de datos de "GHO" contiene una extensa lista de indicadores, que pueden seleccionarse por tema o mediante una funcionalidad de consulta multidimensional. Es el principal repositorio de estadísticas de salud de la Organización Mundial de la Salud.	http://apps.who.int/gho/data/?theme=main
Google open images	Base de datos	Inglés	Google Open Images, es un conjunto de datos que consta de 9 millones de enlaces a imágenes que se han anotado con etiquetas que abarcan más de 6000 categorías.	https://opensource.google/projects/open-images-dataset
HotspotQA Dataset (HotspotQA base de datos)	Base de datos	Inglés	"HotspotQA" es un conjunto de datos de respuesta a preguntas, que presenta preguntas naturales de múltiples tópicos, con una fuerte supervisión para respaldar hechos que permitan sistemas de respuesta de preguntas más explicables. Es recopilado por un equipo de investigadores de PNL en la Universidad Carnegie Mellon, la Universidad de Stanford y la Universidad de Montreal.	https://hotpotqa.github.io/
IEA Atlas of Energy (Atlas de energía)	Base de datos	Inglés	El "IEA Energy Atlas" es un repositorio de datos publicado por la Agencia Internacional de Energía (AIE), una de las referencias mundiales en términos de estadísticas de energía. El Atlas complementa los datos del país, el balance energético animado del flujo de Sankey, las bases de datos y las publicaciones disponibles en la página de estadísticas del sitio web de la AIE.	http://energyatlas.iea.org/#/hellmap/-1118783123

IMDB	Base de datos	Inglés	Gran conjunto de datos de revisión de películas Se proporciona un conjunto de 25,000 críticas de películas altamente polares para capacitación y 25,000 para pruebas. También hay datos adicionales sin etiquetar para su uso. Se proporcionan formatos de texto sin procesar y bolsa de palabras ya procesada.	http://ai.stanford.edu/~amaas/data/sentiment/
Indoor scene recognition (Reconocimiento de escena interior)	Base de datos	Inglés	El repositorio de datos de "Indoor scene recognition" contiene 67 categorías de interior (casas, locales comerciales, etc) y un total de 15620 imágenes. El número de imágenes varía según las categorías, pero hay al menos 100 imágenes por categoría. Todas las imágenes están en formato jpg.	http://web.mit.edu/torralba/www/indoor.html
International Monetary Fund (IMF) open data (Fondo monetario internacional - Datos abiertos)	Base de datos	Inglés	El "IMF open data" es una base de datos que publica series temporales sobre préstamos, tipos de cambio y otros indicadores económicos y financieros. También hay disponibles manuales, guías y diferentes materiales sobre prácticas estadísticas del IMF, en países miembros y en la comunidad estadística en general.	https://www.imf.org/en/Data
Kaggle	Base de datos	Todos los idiomas	Kaggle es un repositorio con más de 19,000 conjuntos de datos públicos y 200,000 cuadernos públicos. Contiene datos de diferentes indoles, donde se destacan datos deportivos, de entretenimiento (cine, Netflix, etc), redes sociales, biología, economía e historia.	https://www.kaggle.com/datasets
Labelled faces in the wild (Rostros etiquetados en la naturaleza)	Base de datos	Inglés	"Labelled faces in the wild" es base de datos de fotografías faciales diseñadas para estudiar el problema del reconocimiento facial. El conjunto de datos contiene más de 13,000 imágenes de caras recopiladas de la web. Cada cara ha sido etiquetada con el nombre de la persona en la foto. 1680 de las personas fotografiadas tienen dos o más fotos distintas en el conjunto de datos.	http://vis-www.cs.umass.edu/lfw/
MIMIC-III	Base de datos	Inglés	"MIMIC-III" es un conjunto de datos de libre acceso desarrollado por el Laboratorio MIT para Fisiología Computacional, que comprende datos de salud desidentificados asociados con 60,000 ingresos a la unidad de cuidados intensivos. Incluye datos demográficos, signos vitales, pruebas de laboratorio, medicamentos y más.	https://mimic.physionet.org/
Pew research center data	Base de datos	Inglés	La página de datos de Pew research center data, contiene datos con tópicos: política de EE.UU, medios y noticias, tendencias sociales, religión, internet y tecnología, ciencias, hispanos, global y métodos. Se necesita usuario para acceder al repositorio.	https://www.pewresearch.org/internet/datasets/
Plenar.io	Base de datos	Inglés	"Plenar.io" es un centro para conjuntos de datos abiertos de todo el mundo, listo para buscar y descargar. Su fuerte son los datos de carácter social.	http://plenar.io/
Quandl	Base de datos	Inglés	"Queandl" es fuente de conjuntos de datos financieros, económicos y alternativos. La misma es utilizada por más de 400,000 personas, incluidos analistas de los principales fondos de cobertura del mundo, gestores de activos y bancos de inversión.	https://www.quandl.com/
Registry of open data on AWS (Registro de datos abiertos de Amazon servicios web)	Base de datos	Inglés	"Registry of open data on AWS" es un repositorio con 120 conjuntos de datos abiertos, provenientes de Amazon Web Services, de interés informático y astronómico.	https://registry.opendata.aws/
SMS Spam Collection in English (Colección de spam de SMS en inglés)	Base de datos	Inglés	"SMS Spam Collection" es un repositorio de datos de mensajes SMS etiquetados que se han recopilado para la investigación de spam en teléfonos móviles. Tiene una colección compuesta por 5,574 mensajes en inglés, reales y no codificados, etiquetados como legítimos (ham) o spam.	http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/
Stanford dogs dataset (Conjunto de datos de perros Stanford)	Base de datos	Inglés	El conjunto de datos de Stanford Dogs contiene imágenes de 120 razas de perros de todo el mundo. Este conjunto de datos se ha creado utilizando imágenes y anotaciones de ImageNet para la tarea de categorización de imagen detallada.	http://vision.stanford.edu/aditya86/ImageNetDogs/
UNICEF dataSet Repositorio UNICEF	Base de datos	Inglés	Repositorio de datos abiertos de UNICEF con conjuntos de datos relacionados a educación, trabajo infantil, discapacidad infantil, mortalidad infantil, mortalidad maternal, calidad del agua, entre otros.	https://data.unicef.org/
University of California Irvine Machine Learning Repository (Universidad de California -Repositorio de aprendizaje automático Irvine)	Base de datos	Inglés	La base de datos de "University of California Irvine Machine Learning Repository", contiene 22 conjuntos de datos de interés científico, orientados a la medicina, botánica, física y química.	http://mlr.cs.umass.edu/ml/index.html
World bank open data (Banco mundial - Datos abiertos)	Base de datos	Inglés	A"World bank open data" es un repositorio que da acceso libre y gratuito a datos sobre desarrollo global.	https://data.worldbank.org/
xview	Base de datos	Inglés	"xview" es uno de los conjuntos de datos de imágenes generales más grandes disponibles públicamente. Contiene imágenes de escenas complejas de todo el mundo, anotadas mediante cuadros delimitadores. Es necesario registrarse para poder hacer uso de los datos.	http://xviewdataset.org/#dataset
YelpData	Base de datos	Inglés	"YelpData" es un repositorio de datos, que contiene 1,223,094 consejos de 1,637,138 usuarios, más de 1.2 millones de atributos comerciales (horarios, estacionamiento, disponibilidad y ambiente) de cada una de las 192,609 empresas registradas en su página.	https://www.yelp.com/dataset

Bases de datos y herramientas de búsqueda				
Labelme	Base de datos y herramienta	Inglés	LabelMe es una herramienta de anotación en línea para crear bases de datos de imágenes para la investigación de visión por computadora. Posee también una extensa base de datos con imágenes etiquetadas por su cualidad.	http://labelme.csail.mit.edu/Release3.0/browserTools/php/dataset.php
MSCoco	Base de datos y herramienta	Inglés	MSCoco es una herramienta de detección de objetos en imágenes. También cuenta con una larga base de datos en imágenes.	http://cocodataset.org/#overview
Visual genome	Base de datos y herramienta	Inglés	"Visual Genome" es un conjunto de datos que contiene 108,077 Imágenes, 5.4 millones de descripciones de región, 1.7 millones de respuestas visuales a preguntas, 3.8 millones de instancias de objetos, 2.8 millones de atributos y 2.3 millones de relaciones; mapeados en WordNet Synset.	https://www.visualdata.io/
Herramientas de búsqueda				
American economic association (Asociación económica estadounidense)	Herramienta	Inglés	"American Economic Association" es una página que contiene diferentes links, de diversas bases de datos, sobre empleo, producción económica y otras variables macroeconómicas de importante relevancia.	https://www.aeaweb.org/resources/data/us-macro-regional
Ckan	Herramienta	Inglés	"CKAN" es un poderoso sistema de administración de datos que hace que los datos sean accesibles, al proporcionar herramientas para optimizar la publicación, el intercambio, la búsqueda y el uso de datos	https://ckan.org/
Data Verse	Herramienta	Inglés	"Dataverse" es una aplicación web de código abierto que permite compartir, preservar, citar, explorar y analizar datos de investigación. Facilita que los datos estén disponibles para otros y le permite replicar el trabajo de otros más fácilmente.	https://dataverse.org/
Datasets subreddit	Herramienta	Inglés	DataSets subreddit es un blog dentro de la página Reddit, el cual está destinado a compartir, encontrar y discutir distintos datasets, de interés informático.	https://www.reddit.com/r/datasets/
Google data search (Búsqueda de datos de Google)	Herramienta	Inglés	"Google data search" es un motor de búsqueda que ha indexado 25 millones de conjuntos de datos o "datasets", especialmente datos oficiales de los gobiernos y entidades públicas, estudios de mercado y datos científicos. Las áreas temáticas son muy diversas; desde economía hasta literatura y la información procede de varios países.	https://datasetsearch.research.google.com/
Google trends	Herramienta	Inglés	"Google Trends" es una herramienta de acceso libre y gratuito brindada por Google, que permite comparar la popularidad de búsqueda de varias palabras o frases; de esta manera es posible conocer el nivel de búsqueda de un determinado término (keyword) durante un periodo de tiempo determinado, permitiéndose identificar las variaciones en las búsquedas en valores relativos basados en una escala de 0 a 100, donde 100 representa el punto más alto en niveles de búsquedas realizadas respecto a un término o palabra clave.	https://trends.google.com/trends/
Open data kit	Herramienta	Inglés	"Open Data Kit" produce software gratuito y de código abierto para recopilar, administrar y utilizar datos en entornos con recursos limitados.	https://opendatakit.org/
Opendata network	Herramienta	Página en Inglés, documentos en varios idiomas	Opendata network es una herramienta que permite acceder y consultar datos sobre : finanzas, infraestructura, transporte, ambiente, economía, seguridad pública, demografía, educación, salud, desarrollo de vivienda, servicios sociales, recreación y política.	https://www.opendatanetwork.com/
Visual data	Herramienta	Inglés	Visual data es una página destinada a albergar links, que comparten diferentes usuarios, que llevan a diferentes páginas contenedoras de datos y proyectos, en donde se utilizan imágenes de visión por computadora.	https://www.visualdata.io/