

Links to datasets (please share yours)

IXTIYOR AHMEDOV · SATURDAY, DECEMBER 29, 2012

Many thanks to Abdulloh Abdulkariym for starting this list!

- snap.stanford.edu/data/ - Stanford Large Network Dataset Collection
- opendata.socrata.com/ - U.S. and international data spanning a variety of topics, including U.S. Federal data and data sets in categories like Business, Fun, and Education
- data.gov — links to national and international data sources
- data.nasa.gov - open data in categories such as Aeronautics, Climate, Earth Science, Engineering, Life Science, and Space Science
- [www.kdnuggets.com/datasets/index.ht...](http://www.kdnuggets.com/datasets/index.html)
- <https://developers.google.com/place...>
- <http://archive.ics.uci.edu/ml/datas...> - datasets for machine learning
- <http://rdatamining.wordpress.com/20...>
- <http://www.quora.com/Data/Where-can...>
- <http://www.inf.ed.ac.uk/teaching/co...>
- LODD: Linking Open Drug Data Project <http://www.w3.org/wiki/HCLSIG/LODD>

For Social Data (Twitter)

<http://trec.nist.gov/data/microblog...>

For Multiple Category

<http://databib.org/>

<http://archive.ics.uci.edu/ml/datas...>

Have a look at the link given down below, it contains list of many different data set ranging from MBs to TBs and belong to many different category and which are available to use for the research purpose or for any application purpose.

<http://www.quora.com/Data/Where-can...>

Economics and FinanceQuandl Financial and Economics Search Engine:
<https://www.quandl.com/>

UMD: [http://inforumweb.umd.edu/econdata/...](http://inforumweb.umd.edu/econdata/)

World bank: <http://data.worldbank.org/indicator>

CBOE Futures Exchange: <http://cfe.cboe.com/Data/>

Google Finance: <http://finance.yahoo.com/> (R)

Google Trends: <http://www.google.com/trends?q=goog...>

St Louis Fed: <http://research.stlouisfed.org/fred...> (R)

NASDAQ: <https://data.nasdaq.com/>

OANDA: <http://www.oanda.com/> (R)

Yahoo Finance: <http://finance.yahoo.com/> (R)

Government

Archived national government statistics: <http://www.archive-it.org/>

Australia: <http://www.abs.gov.au/AUSSTATS/abs@...>

Canada: <http://www.data.gc.ca/default.asp?l...>

DataMarket: <http://datamarket.com/>

Fed Stats: <http://www.fedstats.gov/cgi-bin/A2Z...>

Guardian world governments: <http://www.guardian.co.uk/world-gov...>

London, U.K. data: <http://data.london.gov.uk/catalogue>

New Zealand: http://www.stats.govt.nz/tools_and_.....

NYC data: <http://nycplatform.socrata.com/>

OECD: <http://www.oecd.org/document/0,3746...>

RITA: http://www.transtats.bts.gov/OT_Del...

San Francisco Data sets: <http://datasf.org/>

U.K. Government Data: <http://data.gov.uk/data>

United Nations: <http://data.un.org/>

U.S. Federal Government Agencies: <http://www.data.gov/metric>

US CDC Public Health datasets: http://www.cdc.gov/nchs/data_access...

The World Bank: <http://wdronline.worldbank.org/>

UK 2011 Census Open Atlas Project: <http://www.alex-singleton.com/2011-...>

India: www.data.gov.in

Singapore: www.data.gov.sg

Machine Learning (UCI is a Good Place to Start)

Airlines Data (2009 ASA Challenge): <http://stat-computing.org/dataexpo/...>

Airports and their locations: <http://www.infochimps.com/datasets/...>

AppliedPredictiveModeling (R package): <http://bit.ly/16wyvkG>

Australian Weather: <http://www.bom.gov.au/climate/dwo/>

Causality Workbench: <http://www.causality.inf.ethz.ch/re...>

Edge data for US domestic flights 1990 to 2009: <http://www.infochimps.com/datasets/...>

GroupLens Research (movie ratings and more): <http://www.grouplens.org/node/12>

Kaggle competition data: <http://www.kaggle.com/>

KDNuggets competition site: www.kdnuggets.com/datasets/

The Koblenz Network Collection: <http://konect.uni-koblenz.de/>

Machine Learning Data Set Repository: <http://mldata.org/>

Medicare Data File: <http://go.cms.gov/19xxPN4>

Microsoft Research: <http://research.microsoft.com/apps/...>

Million songs: <http://blog.echonest.com/post/36391...>

RDataMining.com R and Data Mining ebook data: <http://www.rdatamining.com/data>

The Revolution Analytics Collection: <http://www.revolutionanalytics.com/...>

Social Networking: <http://www.cs.cmu.edu/~jelsas/data/...>

UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/>

53.5 billion clicks: <http://cnets.indiana.edu/groups/nan...>

Public Domain Collections

Academic Torrents: <http://academictorrents.com/>

Data360: <http://www.data360.org/index.aspx>

Datamob.org: <http://datamob.org/datasets>

Factual: <http://www.factual.com/topics/brows...>

Freebase: <http://www.freebase.com/>

Google: <http://www.google.com/publicdata/di...>

infochimps: <http://www.infochimps.com/>

numbray: <http://numbrary.com/>

Sample R data sets: <http://stat.ethz.ch/R-manual/R-patc...> (R)

SourceForge Research Data: <http://www.nd.edu/~oss/Data/data.ht...>

UFO Reports: <http://www.nuforc.org/webreports.ht...>

Wikileaks 911 pager intercepts: <http://911.wikileaks.org/files/inde...>

Stats4Stem.org: R data sets: <http://www.stats4stem.org/data-sets...> (R)

The Washington Post List: <http://www.washingtonpost.com/wp-sr...>

Science

Agricultural Experiments: <http://www.inside-r.org/packages/cr...> (R)

Climate data: <http://www.cru.uea.ac.uk/cru/data/t...>

and <ftp://ftp.cmdl.noaa.gov/>

Gene Expression Omnibus: <http://www.ncbi.nlm.nih.gov/geo/>

Geo Spatial Data: <http://geodacenter.asu.edu/datalist...>

Human Microbiome Project: http://www.hmpdacc.org/reference_ge...

 Phúc Trí

Edit Profile

FAVORITES

News Fe

Message

Events

Shops

Saved

Buy and

GROUPS

đhcn1a r

CS109-D

NTU Nan

Data Min

Thống kê

Discover

Create G

APPS

Live Vide

Games

On This t

 Song Are Neenbo

Photos

Games F

FRIENDS

Close Fri

INTERESTS

Following

EVENTS

Sinh viên

Subscrib

Create E

PAGES

Pages Fe

Like Pag

Create P

PAYMENTS

Payment

MIT Cancer Genomics Data: <http://www.broadinstitute.org/cgi-b...>NASA: <http://nssdc.gsfc.nasa.gov/nssdc/ob...>NIH Microarray data: <ftp://ftp.ncbi.nih.gov/pub/geo/DATA...> (R)Protein structure: <http://www.infobiotic.net/PSPbenchm...>Public Gene Data: <http://www.pubgene.org/>Stanford Microarray Data: <http://smd.stanford.edu/>**Social Sciences**General Social Survey: <http://www3.norc.org/GSS+Website/>ICPSR: <http://www.icpsr.umich.edu/icpsrweb...>UCLA Social Sciences Archive: <http://dataarchives.ss.ucla.edu/Hom...>UPJOHN INST: <http://www.upjohn.org/erdc/erdc.htm...>**Time Series**Time Series data Library: <http://robjhyndman.com/TSDL/>**Universities**Carnegie Mellon University Enron email: <http://www.cs.cmu.edu/~enron/>Carnegie Mellon University StatLab: <http://lib.stat.cmu.edu/datasets/>Carnegie Mellon University JASA data archive: <http://lib.stat.cmu.edu/jasadata/>Ohio State University Financial data: <http://fisher.osu.edu/fin/osudata.h...>UC Berkeley: <http://ucdata.berkeley.edu/>UCLA: <http://wiki.stat.ucla.edu/socr/inde...>UC Riverside Time Series: http://www.cs.ucr.edu/~eamonn/time_...University of Toronto: <http://www.cs.toronto.edu/~delve/da...>

ImagesImageNet: <http://www.image-net.org/>

Text Mining

A number of datasets can be found at: <http://www.cs.uic.edu/~liub/FBS/sen...>

Like Comment

Scott Terry and 39 others

[View 7 more comments](#)



Hammad Alam I am unable to get closer to the data set for sentiment analysis on newspapers.. Scott Terry if you can help me out ??

May 22, 2014 at 5:05pm · [Like](#)



Agrahari Anurag hi scott, i'm searching the data set of non western musical song like hindi,sufi etc....is there any repository containing these data set

September 4, 2014 at 4:16pm · [Like](#)



Parindsheel Singh gud going... Jayant... your list is very extensive & exhaustive

September 4, 2014 at 11:31pm · [Like](#)



Jayant Kumar Parindsheel S Dhillon hehe thanks but the credit is not mine..I contributed a few links..thats all

September 4, 2014 at 11:39pm · [Like](#)



Jayant Kumar Aviska Raag what kind of corpus are you looking for? share some details please...hindi songs got me excited 😊

September 4, 2014 at 11:40pm · [Like](#)



Agrahari Anurag sir ..in musical feature extraction we have some tools and pre created data set available in the UCI repository....but there are no data set for hindi songs.....is there are any data set available for hindi song

September 7, 2014 at 1:09pm · [Like](#)



Loc Nguyen Thank you for interesting data

September 7, 2014 at 1:58pm · [Like](#)



Othman Ali <https://archive.ics.uci.edu/ml/datasets.html>

September 25, 2014 at 4:04am · [Like](#) · 1



Othman Ali www.datahub.io

September 25, 2014 at 4:04am · [Like](#) · 1



Nidhi Dadhich THANX SIR

October 28, 2015 at 4:47pm · [Like](#) · 1



Write a comment...