

【オープンデータ情報】

提供元	データセット名	構成データ内容	国	URL
NTTコミュニケーション科学基礎研究所	INFANT	NTT 乳幼児音声データベース	日本	http://research.nii.ac.jp/src/INFANT.html
NTTコミュニケーション科学基礎研究所 国立大学法人東北大学	FW03	NTT・東北大 親密度別単語了解度試験用音声データセット	日本	http://research.nii.ac.jp/src/FW03.html
NTTコミュニケーション科学基礎研究所 国立大学法人東北大学	FW07	NTT・東北大 親密度別単語了解度試験用音声データセット2007	日本	http://research.nii.ac.jp/src/FW07.html
Twitter	Tweetデータセット	「ツイート」と称される140文字以内の短文の投稿を共有するウェブ上の情報サービスTwitter上のテキストデータ	日本	https://twitter.com/?lang=ja
Yahoo	Yahoo!ファイナンスデータ	Yahoo!ファイナンスで提供している株式に関する時系列データ	日本	http://info.finance.yahoo.co.jp/vip/purchase/
(株)ドワンゴ (有)未来検索ブラジル	ニコニコ動画データセット(メタ)	「ニコニコ動画」のメタデータ	日本	http://www.nii.ac.jp/dsc/idr/nico/nico.html
(株)ドワンゴ (有)未来検索ブラジル	ニコニコ動画データセット(コメント)	「ニコニコ動画」のに対するコメントデータ	日本	http://www.nii.ac.jp/dsc/idr/nico/nico.html
(株)ドワンゴ (有)未来検索ブラジル	ニコニコ大百科データセット	「ニコニコ大百科」で投稿された全ての記事データとそれに付随する掲示板全データ	日本	http://www.nii.ac.jp/dsc/idr/nico/nico.html
(株)ネクスト	HOME'Sデータセット(賃貸物件)	賃貸物件データ	日本	http://www.nii.ac.jp/dsc/idr/next/homes.html
(株)ネクスト	HOME'Sデータセット(賃貸物件画像)	画像データ(間取り図、室内写真等)	日本	http://www.nii.ac.jp/dsc/idr/next/homes.html

提供元	データセット名	構成データ内容	国	URL
(株)リクルートテクノロジーズ	リクルートデータセット(店舗等)	「ホットペッパービューティー」の下記データ (店舗/メニュー/セットメニュー/スタイリスト)	日本	http://www.nii.ac.jp/dsc/idr/recruit/recruit.html
(株)リクルートテクノロジーズ	リクルートデータセット(店舗ブログ)	「ホットペッパービューティー」の 店舗ブログデータ	日本	http://www.nii.ac.jp/dsc/idr/recruit/recruit.html
(株)リクルートテクノロジーズ	リクルートデータセット(口コミ)	「ホットペッパービューティー」の 口コミデータ	日本	http://www.nii.ac.jp/dsc/idr/recruit/recruit.html
気象庁	気象観測データ	気温、降水量、日照時間、湿度、気圧など気象 に関するデータ	日本	http://www.jma.go.jp/jma/menu/menureport.html
クイーンズ大学 ATR	X-Ray	音声研究用X線フィルムデータベース	日本	http://research.nii.ac.jp/src/X-Ray.html
クックパッド(株)	クックパッドデータセット	・レシピデータ ・献立データ	日本	http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html
国立研究開発法人 産業技術総合研究所 (AIST)	歩行運動データセット	日本人健常者の歩行パターンをモーションキャ プチャで計測した歩行データ	日本	https://www.dh.aist.go.jp/database/gait2015/index.html
国立研究開発法人 産業技術総合研究所 (AIST)	人体寸法・形状データセット	個人別の人体寸法及び形状のデータ	日本	https://www.dh.aist.go.jp/database/bodyDB/index.html
国立研究開発法人 産業技術総合研究所 (AIST)	頭部寸法データセット	個人別身長、体重、および頭部寸法16項目の データ	日本	https://www.dh.aist.go.jp/database/head/index.html
国立研究開発法人 産業技術総合研究所 (AIST)	ETL文字データセット	文字認識研究用に収集されたサンプルの手書 き文字(英数、かな、漢字)のデータ	日本	http://etlodb.db.aist.go.jp/
国立研究開発法人 産業技術総合研究所 (AIST)	海域地質構造データセット	音波探査により作成した海域地質画像データ	日本	https://gbank.gsj.jp/marineseisdb/index.html

提供元	データセット名	構成データ内容	国	URL
国立研究開発法人 産業技術総合研究所 (AIST)	火山衛星画像造データセット	全世界の活火山の時系列衛星画像データ	日本	https://gbank.gsj.jp/vsldb/image/
国立研究開発法人 産業技術総合研究所 (AIST)	RWC Music Database	ポピュラー音楽やクラシック音楽の楽曲データ	日本	https://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html
国立研究開発法人 産業技術総合研究所 (AIST)	リレーショナル化学災害データベース (RISCAD)	化学物質が関連する漏洩, 火災, 爆発などの事故情報を検索しやすいキーワードで分類して収録したデータベース	日本	https://riscad.aist-riss.jp/
国立研究開発法人 産業技術総合研究所 (AIST)	H-InvDB	ヒトの遺伝子と転写産物を対象とした統合データベース	日本	http://www.h-invitational.jp/hinv/ahg-db/index_ja.jsp
国立研究開発法人 産業技術総合研究所 (AIST)	SEVENS	創薬関連の膜たんぱく質データベース, Gタンパク質共役型受容体の遺伝子領域を予測、抽出して、既知の配列と共に収めている	日本	http://sevens.cbrc.jp/
国立研究開発法人 産業技術総合研究所 (AIST)	地質文献データベース (Integrated GEOLIS)	地質調査総合センターが旧地質調査所時代から収集・所蔵している資料・地質図類から、日本地域及び日本人著者の文献、世界の地質図類の書誌情報を登録したデータベース	日本	https://gbank.gsj.jp/geolis/
国立研究開発法人 産業技術総合研究所 (AIST)	関東平野の地下地質・地盤データベース	関東平野の地下に分布する地層の、層序、物性、地質構造、埋没地形に関する研究・調査で得られたデータ、ボーリング柱状図、および各種地質モデルをご覧いただけるデータベース	日本	https://gbank.gsj.jp/kantosubsurfacegeoDB/CNV/
国立研究開発法人 産業技術総合研究所 (AIST)	海域地質構造データベース	日本周辺海域で地質調査総合センターが行ってきた海洋調査の音波探査プロファイル画像を公開。表層から海底下深部の探査まで利用可能	日本	https://gbank.gsj.jp/marineseisdb/
国立研究開発法人 産業技術総合研究所 (AIST)	活断層データベース	日本全国の活断層に関する情報を、公表文献から収集したデータベース	日本	https://gbank.gsj.jp/activefault/index_gmap.html

提供元	データセット名	構成データ内容	国	URL
国立研究開発法人 産業技術総合研究所 (AIST)	衛星データ検索システム (MADAS)	産総研地質調査総合センターから配信する衛星データ(現在はASTER)を検索し、その結果を地図上に表示するとともに、KMLや処理データのダウンロードを提供するシステム	日本	https://gbank.gsj.jp/madas/
国立研究開発法人 産業技術総合研究所 (AIST)	有機化合物のスペクトルデータベースシステム (SDBS)	化合物の質量スペクトル、赤外分光スペクトル、NMRスペクトル、レーザーラマンスペクトルと電子スピン共鳴スペクトル(ESR)を収録	日本	http://sdb.sdb.aist.go.jp/sdb/cgi-bin/cre_index.cgi
国立研究開発法人 産業技術総合研究所 (AIST)	分散型熱物性データベース	液体、固体、高温融体に関する熱伝導率、比熱容量、熱拡散率、密度、表面張力、蒸気圧などの熱物性値を収録	日本	http://tpds.db.aist.go.jp/
国立研究開発法人 産業技術総合研究所 (AIST)	標準物質総合情報システム	認証標準物質 (CRM)、標準物質 (RM)を収録	日本	https://www.nmij.jp/rminfo/
国立研究開発法人 産業技術総合研究所 (AIST)	固体NMRスペクトルデータベース (SSNMR_SD)	主に固体状態の試料について多核種のNMR (核磁気共鳴)スペクトルを収録	日本	http://ssnmr-sd.db.aist.go.jp/SSNMR/Top.php
国立研究開発法人 産業技術総合研究所 (AIST)	PoSSuM	タンパク質基質結合部位データベース、PDB中の類似したタンパク質基質結合(候補含む)部位を収録	日本	http://possum.cbrc.jp/PoSSuM/
国立情報学研究所 (NII)	情報検索用テストコレクション	NTCIR-1 のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-2 のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-3 CLIR のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-4 CLIR のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html

提供元	データセット名	構成データ内容	国	URL
国立情報学研究所(NII)	情報検索性テストコレクション	NTCIR-5 CLIR のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-6 CLIR のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-7 ACLIA のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-8 ACLIA のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
国立情報学研究所(NII)	テキスト自動要約用テストコレクション	NTCIR-2 SUMM のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-3 SUMM のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
国立情報学研究所(NII)	特許検索テストコレクション	NTCIR-3 PATENT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-4 PATENT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-5 PATENT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-6 PATENT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-7 PATENT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-8 PATENT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html

提供元	データセット名	構成データ内容	国	URL
国立情報学研究所(NII)	質問応答用テストコレクション	NTCIR-3 QA のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-4 QA のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-5 QA のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-6 QA のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-5 CLQA のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-6 CLQA のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
国立情報学研究所(NII)	Web検索評価用テストコレクション	NTCIR-3 WEB のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-4 WEB のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-5 WEB のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
国立情報学研究所(NII)	動向情報の要約と可視化テストコレクション	NTCIR-6 MuST のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-7 MuST のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html

提供元	データセット名	構成データ内容	国	URL
国立情報学研究所(NII)	意見分析テストコレクション	NTCIR-6 OPINION のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-7 MOAT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-8 MOAT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-7 PATMT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-8 PATMT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
		NTCIR-9 PatentMT のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
国立情報学研究所(NII)	音声ドキュメント検索テストコレクション	NTCIR-9 SpokenDoc のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
国立情報学研究所(NII)	コミュニティQAテストコレクション	NTCIR-8 CQA のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
国立情報学研究所(NII)	地理・時間情報の検索テストコレクション	NTCIR-8 GeoTime のタスクデータ	日本	http://www.nii.ac.jp/dsc/idr/ntcir/ntcir.html
国立大学法人 奈良先端科学技術大学院大学	NAIST Japanese Dictionary	IPA 品詞体系(THiMCO97)に基づいて作成されたIPADICの後継辞書	日本	https://osdn.jp/projects/naist-jdic/
国立大学法人 京都大学 工学研究科	PASD	模擬対話音声の音声コーパス	日本	http://research.nii.ac.jp/src/PASD.html

提供元	データセット名	構成データ内容	国	URL
国立大学法人 筑波大学 知能情報生体工学研究室	PASL-DSR	連続音声データベース（単音節/外来音節/ATR音素バランス単語/短文/文章）	日本	http://research.nii.ac.jp/src/PASL-DSR.html
特許庁	日英機械翻訳辞書データセット	日本語の公報を英語に機械翻訳するサービスで用いる日英機械翻訳辞書データ	日本	https://www.jpo.go.jp/shiryou/toushin/chousa/tokyo_dictionary.htm
文部科学省	DIAS(データ統合・解析システム)	衛星観測、大気・地上気象観測、海洋観測、気候変動予測モデルの出力データ等	日本	http://www.diasjp.net/
ヤフー(株)	Yahoo!知恵袋データセット	・「Yahoo!知恵袋」の質問データ ・「Yahoo!知恵袋」の回答データ	日本	http://www.nii.ac.jp/dsc/idr/yahoo/chiebukuro2/Y_chiebukuro.html
楽天(株)	楽天市場データセット(レーティング)	・「楽天市場」のレーティング付き全商品データ	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天市場データセット(レビュー)	・「楽天市場」のレビューデータ	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天トラベルデータセット(レーティング)	「楽天トラベル」のレーティング付き施設データ	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天トラベルデータセット(レビュー)	「楽天トラベル」のレビューデータ	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天GORAデータセット(施設)	「楽天GORA」のゴルフの施設データ	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天GORAデータセット(レビュー)	「楽天GORA」のレビューデータ	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天レシピデータセット(レシピ)	「楽天レシピ」のレシピ情報データ	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天レシピデータセット(レシピ画像)	「楽天レシピ」のレシピ画像データ	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天オークションデータセット	「楽天オークション」の取引評価情報	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html

提供元	データセット名	構成データ内容	国	URL
楽天(株)	TSUKUBAコーパスデータセット	楽天トラベルのレビューデータに対して、文単位で評価極性情報を付与したコーパスデータ	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天商品カテゴリ付きデータセット	Caltech-256にある一部のカテゴリに対応する楽天ジャンルの商品画像にカテゴリを付与したデータ	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	文字領域アノテーションデータセット	画像内の文字領域の矩形の座標を付与したもの	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天Wikiデータセット(ビデオ属性)	「楽天Wiki」のビデオ属性情報	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
楽天(株)	楽天Wikiデータセット(レーティング)	「楽天Wiki」のユーザー行動評価	日本	http://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html
一般財団法人日本情報経済社会推進協会(JIPDEC)	ASJ-JIPDEC	日本音響学会 研究用連続音声データベース	日本	http://research.nii.ac.jp/src/ASJ-JIPDEC.html
一般社団法人 情報処理学会 音声言語情報処理研究会	CENSREC	CENSREC-1 AURORA-2J 雑音環境下連続英語数字音声データ	日本	http://research.nii.ac.jp/src/CENSREC-1.html
		CENSREC-1-C 連続数字8~10発話を1ファイルとする音声データ	日本	http://research.nii.ac.jp/src/CENSREC-1-C.html
		CENSREC-1-AV音声と口唇動画像を用いたバイモーダル音声認識用データ	日本	http://research.nii.ac.jp/src/CENSREC-1-AV.html
		CENSREC-2 CENSREC-1と同じ発話内容を自動車内で収録した音声データ	日本	http://research.nii.ac.jp/src/CENSREC-2.html
		CENSREC-3 CENSREC-2と同じ実験車両で収録した単語および音素バランス文の音声データ	日本	http://research.nii.ac.jp/src/CENSREC-3.html
		CENSREC-4 残響下連続数字発声データ等	日本	http://research.nii.ac.jp/src/CENSREC-4.html

提供元	データセット名	構成データ内容	国	URL
一般社団法人 日本音響学会	JNAS	日本音響学会 新聞記事読み上げ音声コーパス	日本	http://research.nii.ac.jp/src/JNAS.html
株式会社不満買取センター	不満調査データセット	株式会社不満買取センターが運営するWebサービス「不満買取センター」に一般ユーザが投稿した様々な不満に関するデータ	日本	http://www.nii.ac.jp/dsc/idr/fuman/fuman_top.html
技術研究組合 新情報処理開発機構RWCP	RWCP	RWCP-SP96 音声対話データベース - 96年版(人間同士1対1, 対面での目的指向対話(質問応答形式の自由対話))	日本	http://research.nii.ac.jp/src/RWCP-SP96.html
		RWCP-SP97 音声対話データベース - 97年版(96年版と同様(ただし, 話題は海外旅行計画のみ))	日本	http://research.nii.ac.jp/src/RWCP-SP97.html
		RWCP-SP99 検索・要約用ニュース音声データベース(ニュース放送用の原稿をプロのアナウンサが読み上げたもの)	日本	http://research.nii.ac.jp/src/RWCP-SP99.html
		RWCP-SP01 会議音声データベース(4名以上が参加する模擬会議の音声データ)	日本	http://research.nii.ac.jp/src/RWCP-SP01.html
		RWCP-SSD 実環境音声・音響のデータベース	日本	http://research.nii.ac.jp/src/RWCP-SSD.html
国立研究開発法人 科学技術振興機構 (JST)	科学技術文献情報データベース(JDream III)	国内外の科学技術や医学・薬学関係の文献データベース	日本	http://jdream3.com/
国立研究開発法人 科学技術振興機構 (JST)	NBDCポータルサイト	公的資金で産生された生命科学分野の研究データ及びヒトに関するデータ	日本	https://biosciencedbc.jp/
国立研究開発法人 科学技術振興機構 (JST)	総合電子ジャーナルプラットフォーム(J-STAGE)	日本国内の科学技術情報関係の電子ジャーナル発行支援システム	日本	https://www.istage.jst.go.jp/browse/-char/ja/
国立研究開発法人 海洋研究開発機構 (JAMSTEC)	深海映像・画像アーカイブス(J-EDI)	地球深部探査船「ちきゅう」による掘削孔内観測システム	日本	http://www.godac.jamstec.go.jp/jedi/j/

提供元	データセット名	構成データ内容	国	URL
国立研究開発法人 情報通信研究機構 (NICT)	数値人体モデルデータセット	人体を細かなブロックで分割し、それぞれのブロックに筋肉や脂肪といった組織名を示す番号を付与したデータ	日本	https://www.nict.go.jp/out-promotion/data-provided/trans-edr/voxel_guide.html
国立研究開発法人 情報通信研究機構 (NICT)	日本語話し言葉音声データセット	日本語話し言葉の対話などの音声データ	日本	
国立研究開発法人 情報通信研究機構 (NICT)	静止衛星画像データセット	定期的に取得した日本から見える静止衛星軌道上の衛星画像データ	日本	https://www.nict.go.jp/out-promotion/data-provided/trans-edr/satellite_guide1.html
国立研究開発法人 情報通信研究機構 (NICT)	Wikipedia日英京都関連文書対訳文書	日英京都関連文書対訳文書データ	日本	https://alaginc.nict.go.jp/WikiCorpus/#license
国立研究開発法人 情報通信研究機構 (NICT)	Wikipedia日英京都関連文書対訳辞書	日英京都関連対訳用語集データ	日本	https://alaginc.nict.go.jp/WikiCorpus/#license
国立研究開発法人 情報通信研究機構 (NICT)	タイムビジネス用時刻配信公開データベース	GGTTSデータ、GPS-CVスケジュール、うるう秒の実施に関するデータ(無償)	日本	http://tb.nict.go.jp/database.html
国立研究開発法人 情報通信研究機構 (NICT)	宇宙天気サービス	宇宙天気情報に関する各種情報を提供(無償)	日本	http://swc.nict.go.jp/contents/index.php
国立研究開発法人 情報通信研究機構 (NICT)	国内電離層月報	各月毎の日本国内(稚内、国分寺、山川、沖縄)のイオゾンデによる電離圏観測のサマリーや電離圏パラメータの読み取り値、平磯の太陽電波観測のサマリーやイベントリストが記されたデータブック(無償)	日本	http://wdc.nict.go.jp/IONO/wdc/iono-monthly/index_j.html
国立研究開発法人 情報通信研究機構 (NICT)	南極電離層年報	各年毎の南極昭和基地のイオゾンデによる電離圏観測のサマリーや電離圏パラメータの読み取り値が記されたデータブック(無償)	日本	http://wdc.nict.go.jp/IONO/wdc/iono-antactica/index_j.html
国立研究開発法人 情報通信研究機構 (NICT)	可動基線電波干渉計による静止衛星データ	日本から可視範囲にある、静止衛星からの電波を干渉計で受信し、相関処理をすることによって、衛星の方角データを提供(有償)	日本	https://www.nict.go.jp/out-promotion/data-provided/trans-edr/satellite_guide2.html

提供元	データセット名	構成データ内容	国	URL
国立研究開発法人 情報通信研究機構 (NICT) 国立研究開発法人 科学技術振興機構 (JST)	ASPEC-JE	科学論文のAbstractから抽出した日本語と英語の対訳データ	日本	http://orchid.kuee.kyoto-u.ac.jp/ASPEC/
国立研究開発法人 情報通信研究機構 (NICT) 国立研究開発法人 科学技術振興機構 (JST)	ASPEC-JC	科学論文のAbstractから抽出した日本語と中国語の対訳データ	日本	http://orchid.kuee.kyoto-u.ac.jp/ASPEC/
国立研究開発法人 新エネルギー産業技術総合開発機構 奈良先端科学技術大学院大学	S-JNAS	新聞記事読み上げ高齢者音声コーパス	日本	http://research.nii.ac.jp/src/S-JNAS.html
国立研究開発法人 物質・材料研究機構 (NIMS)	MatNavi (NIMS物質・材料データベース)	物質・材料に関する世界最大級のデータベースシステム	日本	http://mits.nims.go.jp/
国立研究開発法人 理化学研究所	FANTOMデータベース	マウスゲノム百科事典プロジェクトで収集された完全長cDNAのアノテーションデータ	日本	http://fantom.gsc.riken.jp/jp/
国立研究開発法人 理化学研究所 免疫・アレルギー科学総合研究センター	RefDIC	免疫細胞のmRNAとタンパク質のプロファイルデータ	日本	
国立研究開発法人 理化学研究所 脳科学総合研究センター 言語知能システム研究チーム	RIKEN-DLG	理研ワープロ操作対話音声コーパスデータ	日本	http://research.nii.ac.jp/src/RIKEN-DLG.html
国立国会図書館	国会会議録データセット	国会会議録のデータは、会議の日、会議の参加者、発話した人の名前、発話自体からなる。	日本	http://kokkai.ndl.go.jp/
国立大学法人 筑波大学 知能情報生体工学研究室	UT-ML	筑波大 多言語音声コーパスデータ(単語音声/連続音声)	日本	http://research.nii.ac.jp/src/UT-ML.html
国立大学法人 筑波大学 知能情報生体工学研究室	GSR-JD	基盤研究(A)「日本語方言の地域差」方言音声コーパス (読み上げ発話/自然談話)	日本	http://research.nii.ac.jp/src/GSR-JD.html
国立大学法人 東北大学 応用情報学研究センター 松下通信工業株式会社 松下技研株式会社	TMW	東北大-松下 単語音声データベース (単語音声 音韻バランス/単語音声 鉄道駅名・線名)	日本	http://research.nii.ac.jp/src/TMW.html

提供元	データセット名	構成データ内容	国	URL
国立大学法人 名古屋大学 CIAIR 統合音響情報研究拠点	CIAIR-VCV	子供の声データベース(室内(通常の生活環境下)での小学生の声を収集したもの)	日本	http://research.nii.ac.jp/src/CIAIR-VCV.html
国立大学法人千葉大学	MapTask	千葉大 地図課題対話コーパス	日本	http://research.nii.ac.jp/src/MapTask.html
国立大学法人千葉大学 国立大学法人大同大学	AWA-LTR	AWA長期間収録音声コーパス	日本	http://research.nii.ac.jp/src/AWA-LTR.html
国立大学法人千葉大学 国立大学法人東京工科大学	Chiba3Party	千葉大学 3人会話コーパス	日本	http://research.nii.ac.jp/src/Chiba3Party.html
国立大学法人東京工科大学	OGVC	感情評定値付きオンラインゲーム音声チャットコーパス	日本	http://research.nii.ac.jp/src/OGVC.html
国立大学法人東京工業大学	TITML	TITML-IDN インドネシア語	日本	http://research.nii.ac.jp/src/TITML-IDN.html
		TITML-ISL アイスランド語	日本	http://research.nii.ac.jp/src/TITML-ISL.html
国立大学法人東京工芸大学	Keio-ESD	慶應義塾大学 研究用感情音声データベース	日本	http://research.nii.ac.jp/src/Keio-ESD.html
国立大学法人東京工芸大学 国立大学法人宇都宮大学	JVPD	身体情報付き男・女・子どもの母音音声データベース	日本	http://research.nii.ac.jp/src/JVPD.html
国立大学法人北海道医療大学	MULTEXT-C	中国語MULTEXTコーパス	日本	http://research.nii.ac.jp/src/MULTEXT-C.html
情報・システム研究機構 人文学オープンデータ共同利用センター	国文研データセット	デジタル化された古典籍のデータを、日本古典籍データセットとして公開	日本	http://www.nii.ac.jp/dsc/idr/nijl/nijl.html

提供元	データセット名	構成データ内容	国	URL
大学共同利用機関法人 自然科学研究機構 国立天文台	すばる望遠鏡アーカイブシステム (SMOKA)	すばる望遠鏡等の公開データ	日本	http://smoka.nao.ac.jp/index.ja.jsp
大学共同利用機関法人 情報・システム研究機構 国立極地研究所 国立研究開発法人 海洋研究開発機構 (JAMSTEC) 国立大学法人 北海道大学	北極域研究推進プロジェクト (ArCS)	北極域観測データ、レーダー観測及び衛星観測データ、数値モデル・シミュレーション結果	日本	http://www.arcs-pro.jp/
大学共同利用機関法人 情報・システム研究機構 国立極地研究所 国立研究開発法人 海洋研究開発機構 (JAMSTEC) 国立大学法人 北海道大学	Arctic Data archive System (ADS)	北極域データアーカイブは、大気、海洋、雪氷、陸域、生態に関する観測データ(English)	日本	https://ads.nipr.ac.jp/portal/
大学共同利用機関法人 情報・システム研究機構 国立極地研究所 国立研究開発法人 海洋研究開発機構 (JAMSTEC) 国立大学法人 北海道大学	北極域データアーカイブシステム (ADS)	北極域データアーカイブは、大気、海洋、雪氷、陸域、生態に関する観測データ	日本	https://ads.nipr.ac.jp/portal/index.action?lng=ja
大学共同利用機関法人 人間文化研究機構 国立国語研究所 コーパス開発センター	現代日本語書き言葉均衡コーパス (BCCWJ)	現代日本語の書き言葉の全体像を把握するために構築したコーパス。書籍全般、雑誌全般、新聞、白書、ブログ、ネット掲示板、教科書、法律などのジャンルにまたがって1億430万語のデータを格納	日本	
電子技術総合研究所 音声研究室(現・産業技術総合研究所)	ETL-WD	電総研 単語音声データベース	日本	http://research.nii.ac.jp/src/ETL-WD.html
独立行政法人情報処理推進機構 (IPA)	共通語彙基盤	コア語彙、DMD(データモデル記述パッケージ)、IEP(情報交換パッケージ)	日本	http://goikiban.ipa.go.jp/
日本学術振興会 文部科学省	Tsuruoka91-92	鶴岡調査音声データベース91-92	日本	http://research.nii.ac.jp/src/Tsuruoka91-92.html
日本電子工業振興協会(現・電子情報技術産業協会)	JEIDA-JCSD	電子協 日本語共通音声データ	日本	http://research.nii.ac.jp/src/JEIDA-JCSD.html
日本電子工業振興協会(現・電子情報技術産業協会)	JEIDA-NOISE	電子協 騒音データベース	日本	http://research.nii.ac.jp/src/JEIDA-NOISE.html

提供元	データセット名	構成データ内容	国	URL
文部科学省	UME	UME-ERJ 日本人学生による読み上げ英語音声データベース	日本	http://research.nii.ac.jp/src/UME-ERJ.html
		UME-JRF 留学生による読み上げ日本語音声データベース	日本	http://research.nii.ac.jp/src/UME-JRF.html
文部科学省	MULTEXT-J	特定領域研究「韻律と音声処理」日本語 MULTEXT韻律コーパス	日本	http://research.nii.ac.jp/src/MULTEXT-J.html
文部科学省	Oogami	宮古大神島方言音声データベース	日本	http://research.nii.ac.jp/src/Oogami.html
國學院大學 研究開発推進機構 日本文化研究所	Aragusuku	南琉球新城方言音声データベース	日本	http://research.nii.ac.jp/src/Aragusuku.html

【オープンデータ情報】

提供元	データセット名	構成データ内容	国	URL
Adam Coates(スタンフォード大学)	STL-10 datase	飛行機、車、鳥、猫、鹿など10種類のカテゴリに分類された画像データ	アメリカ	https://cs.stanford.edu/~acoates/stl10/
Alex Krizhevsky(トロント大学)	CIFAR10	10種類のカテゴリに分類された画像データ	カナダ	http://www.cs.utoronto.ca/~kriz/cifar.html
Amazon	Landsat on AWS	ランドサット 8 号衛星により作成中の地球全土の衛星画像データ	アメリカ	http://aws.amazon.com/jp/public-data-sets/landsat/
Amazon	NEXRAD on AWS	次期世代気象レーダーのセンサーデータ	アメリカ	http://aws.amazon.com/jp/noaa-big-data/nexrad/
Amazon	Common Crawl Corpus	50 億を超えるウェブページから構成されるウェブクロールデータのコーパス	アメリカ	https://aws.amazon.com/jp/datasets/common-crawl-corpus/
Andrew Howard et.al	Radish	ロボットが行動することによって生成されるセンサーデータ及び環境地図データ	アメリカ	http://radish.sourceforge.net/index.php
AT&T ケンブリッジ研究所	Olivetti	照明、顔の表情(目の開閉、笑顔の有無)、顔の詳細(眼鏡の有無)を組み合わせて作成した顔画像データ	イギリス	http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
Bernd Krueger	Piano-midi	クラシック音楽を対象としたピアノ演奏による楽曲データ	-	http://www.piano-midi.de/
François Kawala et.al	Buzz in Social Media Dataset	テクノロジー領域に関するツイートデータ	-	https://archive.ics.uci.edu/ml/datasets/Buzz+in+social+media+

提供元	データセット名	構成データ内容	国	URL
Google	YouTube Comedy slam Preference Dataset	YouTube Comedy slamのビデオに対するレビューデータ	アメリカ	https://archive.ics.uci.edu/ml/datasets/YouTube+Comedy+Slam+Preference+Data
Google	Google Books N-grams	Google Booksを対象にしたN-gramのデータ	アメリカ	http://commondatastorage.googleapis.com/books/syntactic-ngrams/index.html
Humintell社	JACFEE	喜怒哀楽など感情が表れた表情である日本人及び白人の男女の顔画像データ	アメリカ	http://www.humintell.com/research-news/
Humintell社	JACNeuF	感情が表れていない中立な表情である日本人及び白人の男女の顔画像データ	アメリカ	http://www.humintell.com/research-news/
IntelResearch/ワシントン大学	RGB-D Object (image)	300種類の物体を撮影したRGB-D画像データ	アメリカ	http://rgbd-dataset.cs.washington.edu/index.html
IntelResearch/ワシントン大学	RGB-D Object (movie)	屋内の景色を撮影したRGB-D動画データ	アメリカ	
James Allwright	Nottingham	フォーク音楽を対象とした楽曲データ	-	http://abc.sourceforge.net/NMD/
LDC(Linguistic Data Consortium)	TIMIT Speech Corpus	アメリカ英語のスピーカーが読み上げる文章の音声データ	アメリカ	https://catalog.ldc.upenn.edu/LDC93S1
LDC(Linguistic Data Consortium)	Broadcast News	TV(ABC、CNN、CSPAN)とラジオ(NPR、PRI)の放送で話されている会話のテキストデータ	アメリカ	https://catalog.ldc.upenn.edu/LDC97S44
Lior Wolf(テルアビブ大学)	YouTube Faces DB	YouTubeから抽出した2D顔動画データ	イスラエル	http://www.cs.tau.ac.il/~wolf/vtfaces/

提供元	データセット名	構成データ内容	国	URL
Luis Seabra Lopes	Robot Execution Failures Data Set	ロボットがタスクで失敗した動作のデータ	-	https://archive.ics.uci.edu/ml/datasets/Robot+Execution+Failures
Microsoft	Microsoft Common Objects in Context (COCO)	80種類のカテゴリに分類されたキャプション付き画像データ	アメリカ	http://mscoco.org/home/
Microsoft	Microsoft Sequential Image Narrative Dataset (SIND)	ストーリーのあるアノテーション付き画像データ	アメリカ	http://www.sind.ai/index.html
Microsoft	MS MARCO (Microsoft MAchine Reading Comprehension)	匿名化された実際のデータを使った質問と回答をセットにした例となるデータセットで、AIを用いて人間のように質問を読んで回答できる認知システムの研究や開発に利用できる	アメリカ	http://www.msmarco.org/
Microsoft	MSR-VTT dataset	北京にあるMicrosoft Research Asiaによる動画キャプションデータセット	中国	http://ms-multimedia-challenge.com/dataset
MIT CSAIL	Labelme	画像内の各種物体に対して物体名やカテゴリ名など物体に関するテキスト情報が付加されたアノテーション画像データ	アメリカ	http://labelme.csail.mit.edu/Release3.0/browserTools/php/dataset.php
MPII	MPII Cooking Activities Dataset	料理の行為を撮影した2D動画データ	ドイツ	https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/human-activity-recognition/mpii-cooking-activities-dataset/
N. A. Abdulla	Twitter Dataset for Arabic Sentiment Analysis	ネガティブ・ポジティブに分類されたツイートデータ	-	https://archive.ics.uci.edu/ml/datasets/Twitter+Data+set+for+Arabic+Sentiment+Analysis
NASA	Statlog (Shuttle) Data Set	NASAスペースシャトルの動作データ	アメリカ	https://archive.ics.uci.edu/ml/datasets/Statlog+(Shuttle)
Netflix社	Netflix Prize	Netflix Prize (映画のユーザーレイティングを予測する協調フィルタリングアルゴリズムのコンペ) で用いたレーティング付きデータ	アメリカ	http://www.netflixprize.com/

提供元	データセット名	構成データ内容	国	URL
Pascal VOC	Pascal VOC	Pascal VOC(物体認識のコンペ)で用いる飛行機、自転車、車等20種類の物体カテゴリ情報、物体の位置情報などが付与された画像データ	-	http://host.robots.ox.ac.uk/pascal/VOC/
The Echo Nest社	Million Song dataset	音楽情報企業The Echo Nest社が収集したポピュラー音楽を対象とする楽曲データ	アメリカ	http://labrosa.ee.columbia.edu/million-song/
Yahoo	Yahoo Flickr Creative Commons	Yahoo! Flickrから集めた動画データ	アメリカ	http://webscope.sandbox.yahoo.com/catalog.php?datatype=i&did=67
Yahoo	Yahoo!Music User Rating of Musical artists	Yahoo!Musicから集めたアーティストに関するレーティングデータ	アメリカ	https://webscope.sandbox.yahoo.com/catalog.php?datatype=r
Yann LeCun(ニューヨーク大学)	MNIST	0から9までの手書きで書かれた数字の画像データ	アメリカ	http://yann.lecun.com/exdb/mnist/
エジンバラ大学	European Parliament Proceedings Parallel Corpus(doc)	ヨーロッパの21言語に対する対訳文書	スコットランド	http://www.statmt.org/europarl/
エジンバラ大学	European Parliament Proceedings Parallel Corpus(dic)	ヨーロッパの21言語に対する対訳辞書	スコットランド	http://www.statmt.org/europarl/
オックスフォード大学	Oxford Robotcar Dataset	ロボットカーを走行させて取得した動画データ及びセンサーデータ(LIDAR,GPS,INS)	イギリス	http://robotcar-dataset.robots.ox.ac.uk/
カーネギーメロン大学	CMU Motion Capture Database	人の基本的な動作(歩く、走る、飛ぶ等)やスポーツをしている動作(バスケット、サッカー、ボクシング等)のなど人が行う動作の動画データとその動作をモーションキャプチャーした動画データ	アメリカ	http://mocap.cs.cmu.edu/
カーネギーメロン大学	Multi-Pie	照明(19種類)、視点(15箇所)を組み合わせて作成した顔画像データ	アメリカ	http://www.multipie.org/
カリフォルニア工科大学	Caltech 101	101種類のカテゴリに分類された画像データ	アメリカ	http://www.vision.caltech.edu/Image-Datasets/Caltech101/

提供元	データセット名	構成データ内容	国	URL
カリフォルニア工科大学	Caltech 256	256種類のカテゴリに分類された画像データ	アメリカ	http://www.vision.caltech.edu/Image_Datasets/Caltech256/
カリフォルニア大学	Jester	Jester(ジョークを評価するサイト)から抽出したレーティング付きデータ	アメリカ	http://www.ieor.berkeley.edu/~goldberg/jester-data/
カリフォルニア大学	Gas sensor array exposed to turbulent gas mixtures	特定の化学的物質を検出するガスセンサーのデータ	アメリカ	Gas sensor array exposed to turbulent gas mixtures Data Set
コロンビア大学	COIL 20	マグカップ、ミニカーなど20種類の物体について、5° ずつ回転させた視点毎に作成した画像データ	アメリカ	http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
スタンフォード大学	MuseData	クラシック音楽を対象としたピアノ演奏による楽曲データ	アメリカ	http://musedata.stanford.edu/
スタンフォード大学	Visual Genome	画像内の矩形領域それぞれに総計500万超の短いキャプションが付与されていたり、170万のVQAが付与されていたりするほか、画像1枚1枚に多くの情報が付与されている。	アメリカ	https://visualgenome.org/
スタンフォード大学/Google	The Street View House Numbers (SVHN) Dataset	Google Street Viewの画像から、部屋番号が写っている画像データと映っている部屋番号の領域だけ抽出した画像データ	アメリカ	http://ufldl.stanford.edu/housenumbers/
スタンフォード大学/プリンストン大学	Imagenet	英語の辞書であるWordNetで定義された英単語及び同義語で分類された各種物体の画像データ	アメリカ	http://www.image-net.org/
ニューヨーク大学	NORB	カテゴリ(動物、人間、車等)、照明、角度、方位を変えて作成した画像データ	アメリカ	http://www.cs.nyu.edu/~vglab/data/norb-v1.0/
バージニア大/米国陸軍研究所/ ジョージア工科大学	VQA Dataset	画像に対する質問応答のためのデータセット。265,016の画像に対し、少なくとも3種類の質問と応答が付与されている。回答は10通りの正当と、妥当でない回答3通りが付与されている。	アメリカ	http://www.visualqa.org/

提供元	データセット名	構成データ内容	国	URL
バル＝イラン大学	Lexical Inference Datasets	語彙に関する命題論理式で定義された推論データ	イスラエル	http://u.cs.biu.ac.il/~nlp/resources/downloads/lexical-inference-datasets/
ビンガムトン大学	BU-3DFE	中立的な表情及び感情的な表情で構成される3Dの顔画像データ	アメリカ	http://www.cs.binghamton.edu/~liun/Research/3DFE/3DFE_Analysis.html
フライベルク大学	Book-Crossing dataset	Book-Crossing(本を評価するサイト)から抽出したレーティング付きデータ	ドイツ	http://www2.informatik.uni-freiburg.de/~ctiegl/BX/
米国商務省標準化技術研究所(NIST)	Reuters (RCV*) Corpuses	ロイター社のニュースから抽出した文書データ	アメリカ	http://trec.nist.gov/data/reuters/reuters.html
米国商務省標準化技術研究所(NIST)	NIST(fingerprint)	5つのタイプ(アーチ、左ループ、右ループ、テント状アーチ、渦巻き)に分類された指紋の画像データ	アメリカ	http://www.nist.gov/srd/nistsd4.cfm
米国商務省標準化技術研究所(NIST)	NIST(face)	男女の正面顔及び横顔を含む顔画像データ	アメリカ	http://www.nist.gov/srd/nistsd18.cfm
米国商務省標準化技術研究所(NIST)	FERET	FERET(DARPA支援による自動顔認識技術開発プログラム)で開発した顔認識アルゴリズムの評価に用いた、正面から右、左に視点を変えた頭部の画像データ	アメリカ	http://www.nist.gov/itl/iad/ig/feret.cfm
ペンシルバニア大学	Penn Treebank	ウォールストリートジャーナルから抽出した文書に品詞や文の構造情報が付与された文書データ	アメリカ	http://www.cis.upenn.edu/~treebank/
ペンシルバニア大学	Multidomain sentiment analysis dataset(text)	アマゾンの本やDVDなどカテゴリ毎にネガティブ、ポジティブに分類された商品レビューのテキストデータ	アメリカ	http://www.cs.jhu.edu/~mdredze/datasets/sentiment/
ペンシルバニア大学	Multidomain sentiment analysis dataset(rating)	アマゾンの本やDVDなどカテゴリ毎にネガティブ、ポジティブに分類された商品レビューのレーティングデータ	アメリカ	

提供元	データセット名	構成データ内容	国	URL
ボアズィチ大学	Bosphorus	中立的な表情及び感情的な表情で構成される3Dの顔画像データ	トルコ	http://bosphorus.ee.boun.edu.tr/Home.aspx
マサチューセッツ大学	Labelled Faces in the Wild	正面や帽子をかぶらないなどの制約のない顔画像データ	アメリカ	http://vis-www.cs.umass.edu/lfw/
ミネソタ大学	MovieLens	MovieLens(映画の評価、映画の推薦、映画情報を提供するサイト)から抽出したレーティング付きデータ	アメリカ	http://grouplens.org/
メリーランド大学	Dow Jones Index Data Set	Dow Jones Index	アメリカ	https://archive.ics.uci.edu/ml/datasets/Dow+Jones+Index
ヨーク大学	The 3D Face Database	中立的な表情で構成される3Dの顔画像データ	イギリス	https://www-users.cs.york.ac.uk/nep/research/3Dface/tomh/3DFaceDatabase.html
-	20 newsgroups	異なる20のニュースグループから抽出した文書データ	-	http://qwone.com/~jason/20Newsgroups/
-	Brodatz datase	壁、繊維など材料の表面の柄で構成されるテクスチャーの画像データ	-	http://www.ux.uio.no/~tranden/brodatz.html