情報学データ資源の共同利用

Shared Use of Informatics Data Resources

データセット共同利用研究開発センター

Center for Dataset Sharing and Collaborative Research

大山敬三，神門典子，佐藤真一，宮尾祐介，小野順貴，山岸順一，大須賀智子

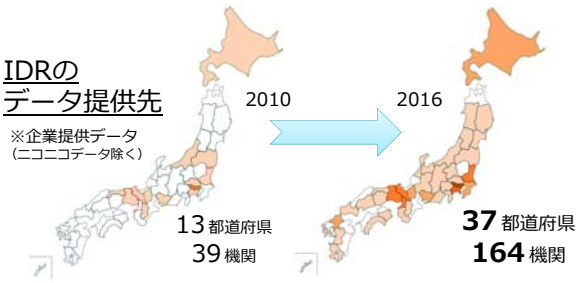
どんな活動？

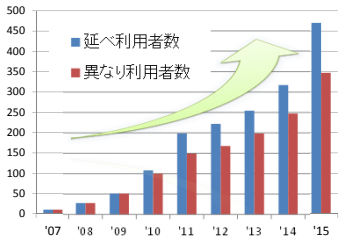
情報学の研究では大量の実データが欠かせませんが、本当に欲しいデータはなかなか手に入りません。そこで、データセット共用プラットフォームを構築して、企業などが様々なデータを提供しやすく、研究者にも使いやすい、循環型の環境づくりを目指します。

何ができる？

データセットの共同利用を通じて研究者と企業などを結びつけ、オープンサイエンス、オープンイノベーションの推進に貢献します。また、共通の技術評価プラットフォームを確立し、研究の透明性を高めるとともに、研究者間の競争と協調を促進します。

センターの活動内容





NTCIR-12来月開催！
詳細はポスターG18へ

研究者へのデータ提供窓口

- IDR（情報学研究データリポジトリ）
  - データセットの獲得・提供
- SRC（音声資源コンソーシアム）
  - 音声コーパスの収集・提供

NII-SRC Speech Resources Consortium　NTCIR　IDR

評価型ワークショップの運営

- NTCIR（情報アクセス研究のためのテストベッドとコミュニティ）
  - テストコレクションの構築
  - 研究コミュニティの形成

データセット共同利用研究開発センター

データ共有基盤

クラウドベースデータセット共用プラットフォームの研究開発

共同研究の推進

データセットの共同利用による新しいワークショップ型共同研究

シンポジウムを開催します

初開催！NII-IDR ユーザフォーラム

2016年11月30日開催　於：NII （予定）

データ提供企業見学会も企画中！

- 提供データを用いた研究発表
- データ提供者も交えたラウンドテーブル　など

データをお持ちの方へ

提供データを募集します!!

様々な実データをお持ちで、研究者への提供をお考えの方はご相談下さい。

実情に応じた提供方法をご提案します。

NII

連絡先：国立情報学研究所　DSC事務局

URL：http://www.nii.ac.jp/dsc/　　Email：dsc@nii.ac.jp

IDR：情報学研究データリポジトリ

Informatics Research Data Repository

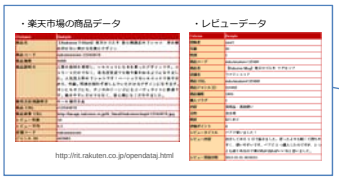
どんな活動？

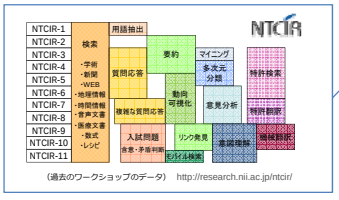
情報学の最新の研究分野では、音声や映像、Web上にあるテキストなど、大量のデータを必要としています。IDRでは、これらのデータを持っている産学界と、データを使いたい研究者の橋渡しをしています。

何ができる？

大学共同利用機関として、研究の効率化と研究者の裾野の拡大に寄与しています。研究者にとっては、共通のデータセットを用いて評価を行うことで、研究の客観性や再現性も担保できるようになります。

現在提供中のデータ





- **Yahoo!データセット**
  - Q&Aサイト「Yahoo!知恵袋」の投稿データ
- **楽天データセット**
  - 楽天市場の全商品＆レビューデータ
  - 楽天トラベルの施設＆レビューデータ UPDATE!!
  - 楽天ゴルフの施設＆レビューデータ
  - 楽天レシピのレシピ情報＆画像データ UPDATE!!
  - 楽天オークションの取引＆評価データ
  - アノテーション付きデータ
  - 楽天Vikiのビデオ情報＆ユーザ情報 NEW!!
- **ニコニコデータセット**
  - ニコニコ動画のメタデータ＆コメントデータ
  - ニコニコ大百科データ
- **リクルートデータセット**
  - ホットペッパービューティデータ
- **クックパッドデータセット**
  - レシピデータ
  - 献立データ
- **HOME'Sデータセット** NEW!!
  - 賃貸物件データ
  - 画像データ
- **不満調査データセット** NEW!!
  - 不満買取センターへの投稿データ
- **国文研データセット** NEW!!
  - 古典籍データ（書誌，画像，タグ，本文テキスト）
- **音声コーパス** 随時更新中！
- **NTCIRテストコレクション** 随時更新中！
  - 27種類の検索課題･正解／質問･回答データ
  - 2種類の文書データ（Webアーカイブ）
- **その他映像データ等**
  - 会話データ，手話データなどの公開を予定

・HOME'Sデータセット

http://www.homes.co.jp/

・不動産・住宅情報サイト「HOME'S」に2015年9月時点で掲載されていた全国約533万件の賃貸物件データ
- 賃料，管理費，敷金，礼金，更新料
- 面積，間取り
- 立地（市区町村，最寄り駅，徒歩分）
- 築年数
- 建物構造
- 諸設備　など

・上記の全物件データに対する間取り図や室内写真など約8,300万枚の画像データ

・間取り図に関しては高精細度の画像データも追加提供開始

・不満調査データセット

・2015年9月までに「不満買取センター」に一般ユーザが投稿した，さまざまな不満に関するデータのうち，オペレータがタグ付けをした約25万件のデータ

（実際に投稿された不満データの例 ↓）

http://fumankaitori.com/

・音声コーパス

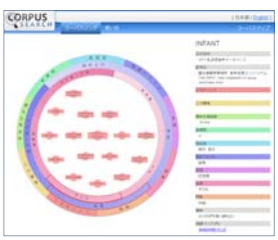
※音声資源コンソーシアムにて受け入れ・加工・提供中

http://research.nii.ac.jp/src/

－44種類の音声データベース

読み上げ／講演／対話，成人／乳幼児／高齢者，方言，多言語，非母語話者，雑音下，残響下，などなど…

－音声コーパス検索システム



http://corpus-search.nii.ac.jp/

発話音声データ

環境音・騒音

連絡先：国立情報学研究所　IDR事務局

URL：http://www.nii.ac.jp/dsc/idr/　　Email：idr@nii.ac.jp