

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、研究の効率化を図るとともに、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。またデータや課題を共有する評価ワークショップを実施し、技術の深化とコミュニティの創生や活性化を促進します。

センターの活動内容

評価型ワークショップ (国際会議)の企画運営

詳細は
ポスター
D08へ

NTCIR

ユーザフォーラム開催

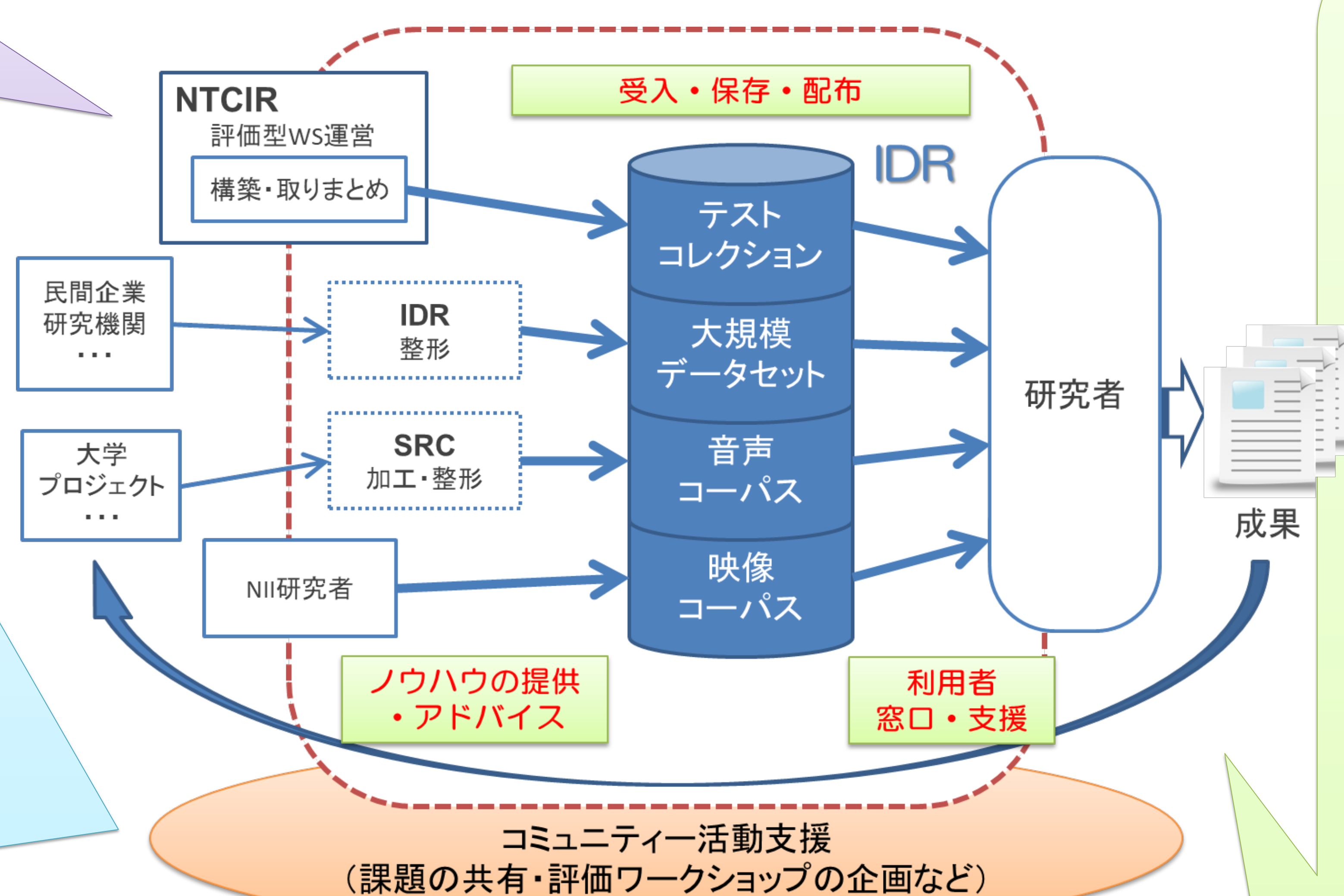
データ提供者と利用者が一同に
会し意見交換できる場の提供



↑データ提供企業登壇のパネル



↑データ利用者のポスター発表



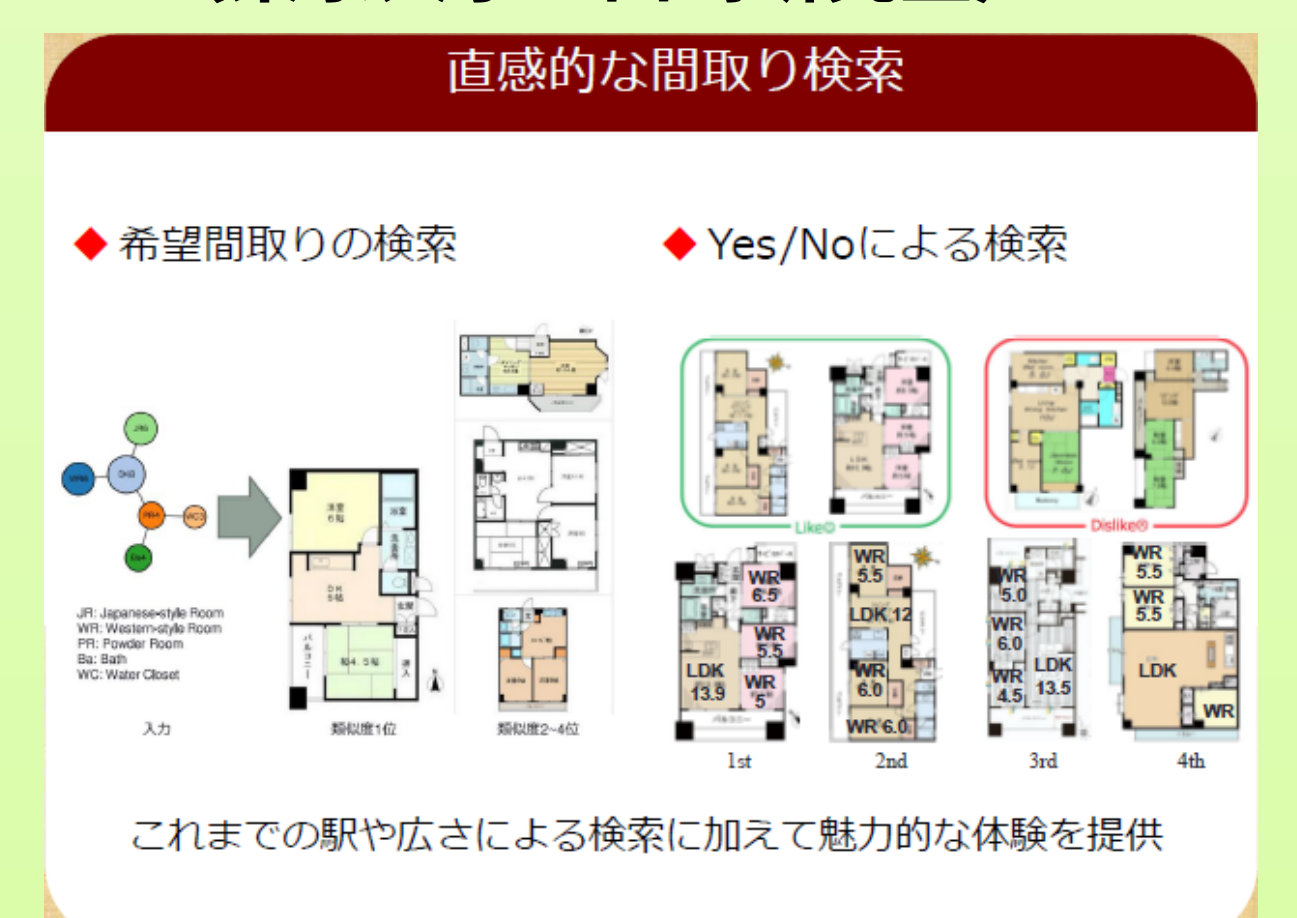
研究成果の公開

利用者による発表論文一覧を
JAIRO Cloud 上で試験公開中

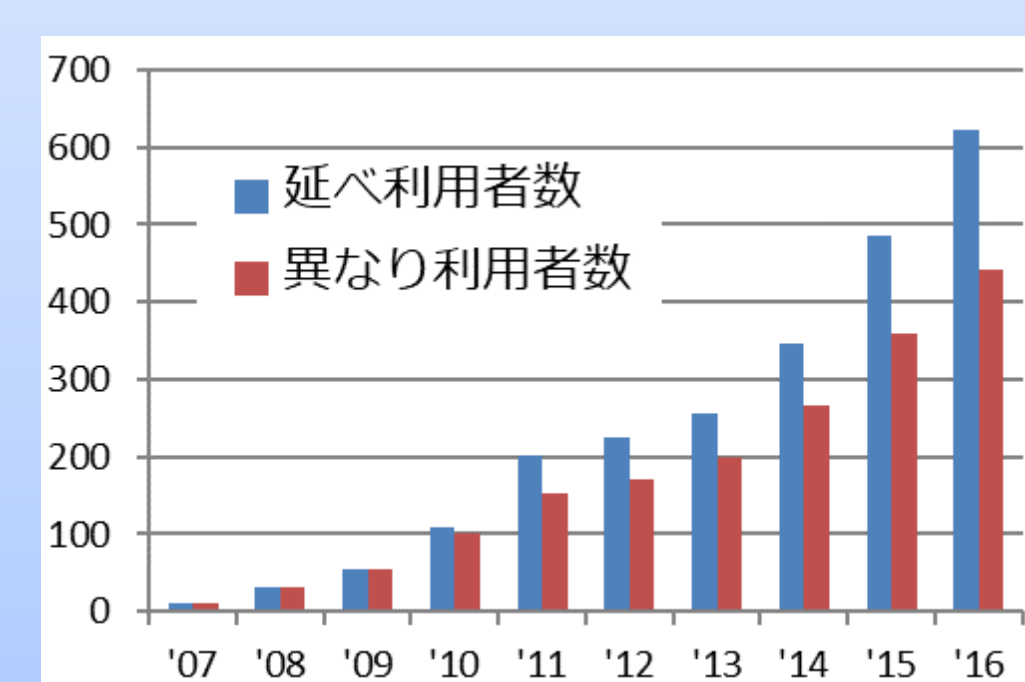
- ・楽天レシピデータを用いた研究の例
(関西学院大学・角谷研究室
& 九州大学・牛尼研究室)



- ・HOME'Sデータを用いた研究の例
(東京大学・山崎研究室)



産学界のデータの研究者への提供



利用者数（研究室単位）の推移



研究成果発表数の推移

今年度の開催イベントのご案内

音声資源活用シンポジウム

2017年9月7日(木) 開催

会場：国立国語研究所（共催）

NII-IDR ユーザフォーラム2017

2017年12月4日(月) 開催

会場：NII

NTCIR-13 カンファレンス

2017年12月5～8日 開催

会場：一橋講堂・中会議場



連絡先：国立情報学研究所 データセット共同利用研究開発センター

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、データを使いたい大学等の研究者の橋渡しをしています。

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」に投稿された質問とその回答

- 質問約1,600万件
- 回答約5,000万件 (ベストアンサー、それ以外)
- 投票数、評価数など



・ 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
- 楽天GORAのゴルフ場&レビューデータ
- 楽天レシピのレシピ情報&画像データ
- 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
- PriceMinisterのレビュー有効性情報 **NEW!!**

楽天データセット

「楽天市場」の約1億5,600万の商品データと約6400万のレビューデータ、その他、楽天トラベルや楽天レシピ、アノテーションデータなど複数

商品名	レビュー数	商品名	レビュー数
商品名1	1000	商品名2	500
商品名3	2000	商品名4	3000
商品名5	1500	商品名6	4000
商品名7	1200	商品名8	2500
商品名9	1800	商品名10	3500
商品名11	1100	商品名12	2800
商品名13	1600	商品名14	3200
商品名15	1300	商品名16	2700
商品名17	1700	商品名18	3100
商品名19	1400	商品名20	2900
商品名21	1900	商品名22	3300
商品名23	1000	商品名24	2600
商品名25	1600	商品名26	3400
商品名27	1200	商品名28	2400
商品名29	1800	商品名30	3600
商品名31	1100	商品名32	2300
商品名33	1700	商品名34	3700
商品名35	1300	商品名36	2100
商品名37	1900	商品名38	3800
商品名39	1000	商品名40	2000
商品名41	1600	商品名42	3900
商品名43	1200	商品名44	2200
商品名45	1800	商品名46	4000
商品名47	1100	商品名48	2500
商品名49	1700	商品名50	3000
商品名51	1300	商品名52	2700
商品名53	1900	商品名54	3500
商品名55	1000	商品名56	2800
商品名57	1600	商品名58	3200
商品名59	1200	商品名60	2600
商品名61	1800	商品名62	3100
商品名63	1100	商品名64	2900
商品名65	1700	商品名66	3300
商品名67	1300	商品名68	2400
商品名69	1900	商品名70	3600
商品名71	1000	商品名72	2300
商品名73	1600	商品名74	3700
商品名75	1200	商品名76	2100
商品名77	1800	商品名78	3800
商品名79	1100	商品名80	2000
商品名81	1700	商品名82	3900
商品名83	1300	商品名84	2200
商品名85	1900	商品名86	4000
商品名87	1000	商品名88	2500
商品名89	1600	商品名90	3000
商品名91	1200	商品名92	2700
商品名93	1800	商品名94	3500
商品名95	1100	商品名96	2800
商品名97	1700	商品名98	3200
商品名99	1300	商品名100	2600



ニコニコデータセット

- ・「ニコニコ動画」約830万件のメタデータとコメントのデータ
- ・「ニコニコ大百科」に投稿された記事データとその掲示板データ



・ 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

クックパッドデータセット

- ・「クックパッド」に2014年9月までに掲載された約172万品のレシピ
- タイトル
- 材料
- 手順
- コツ・ポイント
- このレシピの生い立ち
- つくれぽ
- カテゴリ など
- ・上記レシピからなる献立のタイトル、主菜/副菜などのラベル



・ 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

・ LIFULL HOME'Sデータセット

- 賃貸物件データとその画像データ

・ 不満調査データセット

- 不満買取センターへの投稿データ
- 不満カテゴリ辞書データ **NEW!!**

・ Sansanデータセット **2017.5~ NEW!**

- ダミーの名刺画像データ

・ 音声コーパス **随時更新中！**

- 音声データベース

・ NTCIRテストコレクション **随時更新中！**

- 検索課題・正解/質問・解答データ
- 文書データ (Webアーカイブ)

・ 手話コーパス

- 日本手話話し言葉コーパス **(研究者向けに提供開始予定)**

リクルートデータセット

「ホットペッパービューティ」に掲載された全国的美容室、サロンのデータ

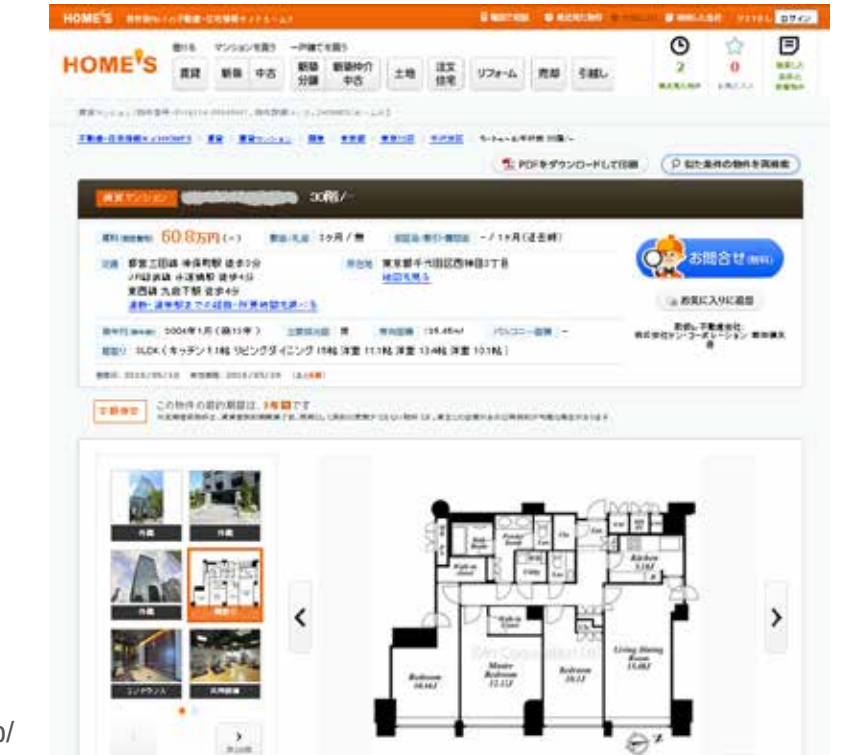
- 店舗データ (約1万件)
- クーポンデータ (約15万件)
- 店舗ブログデータ (約180万件)
- メニューデータ (約52万件)
- スタイリストデータ (約9万件)
- 口コミデータ (約36万件)



LIFULL HOME'Sデータセット

不動産・住宅情報サイト「LIFULL 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とその画像データ約8,300万枚

- 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
- ※間取り図に関しては高精細の画像データも追加提供開始



Sansanデータセット

データ分析コンテスト「人工知能は名刺をどこまで読解できるのか」で使用されたダミーの名刺データ3,481枚 (※氏名等は実在しないサンプル)

- 各領域に以下のラベルが付与
- 会社名
- 氏名
- 役職
- 郵便番号、住所
- 電話・FAX・携帯番号
- E-mailアドレス
- HPのURL

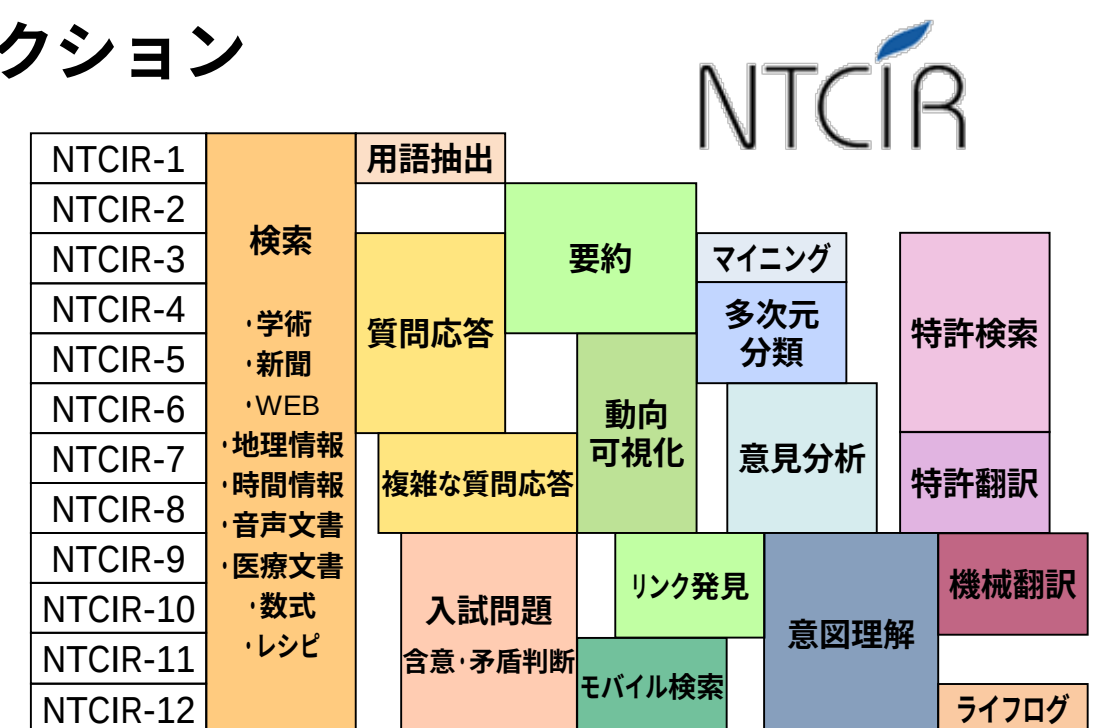


NTCIRテストコレクション

情報アクセス技術の評価を行うワークショップ
※現在NTCIR-13が進行中！

→過去のデータを配布中

- ・タスクデータ32種類
- 検索課題と正解判定
- 質問と解答
- ・WEB文書データ2種類



NTCIRプロジェクト: <http://research.nii.ac.jp/ntcir/>

データの入手をご希望の方は…

それぞれ利用申請 → (審査) → 契約手続きが必要です。まずはIDRサイトの案内に従い利用申請をしてください。

NII IDR

検索

実情に応じた提供方法をご提案致します。まずはご相談下さい。

メール (IDR事務局) : idr@nii.ac.jp

NII

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

