

15. データベースの応用, データベースの種類

URL: <https://www.kkaneko.jp/de/ds/index.html>

金子邦彦



謝辞：この資料では「いらすとや」のイラストを使用しています

オープンデータ

オープンデータ

- 政府等が保有するデータがデジタルで公開

データカタログサイト

<https://www.data.go.jp/>

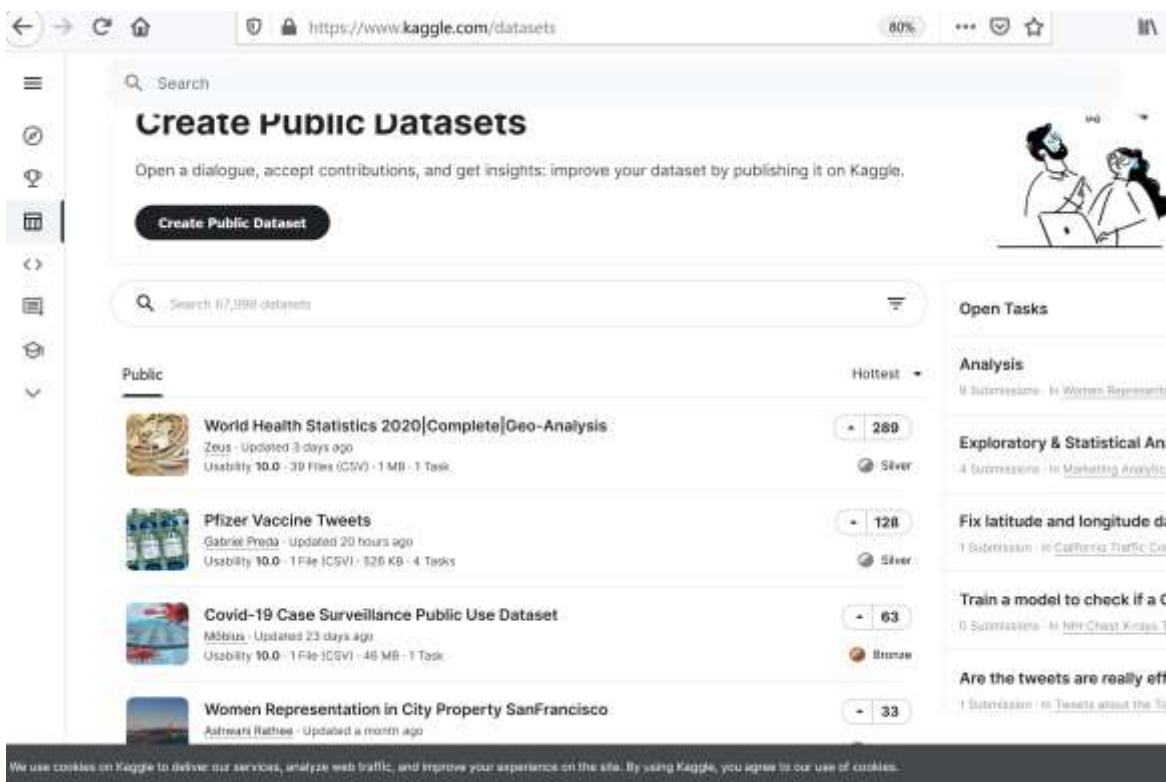
The screenshot shows the homepage of the DATA GO.JP website. At the top, there is a dark blue header with the logo "DATA GO.JP" and the text "データカタログサイト". Below the header, there is a navigation menu with links: "お知らせ", "データ", "データベースサイト一覧", "公共データ活用事例", "コミュニケーション", and "開発者向け情報". A search bar is located below the menu, containing the text "データセットを検索...". Below the search bar, there is a notice: "意見受付コーナーにて、オープンデータの掲載に関する御要望等を受け付けております (匿名可)". The main content area is divided into three columns: "データ" (Data) with icons for HTML, PDF, XLSX, CSV, and XML; "オープンデータの取組 (リンク集)" (Open Data Initiatives (Link Collection)) with a central "OPENDATA LINKS" icon; and "コミュニケーション" (Communication) with speech bubble icons. At the bottom, there are two dark blue bars: "一億総活躍社会の実現" (Realization of a Society where 100 million people thrive) on the left and "地方公共団体データベースサイト一覧" (List of Local Government Database Sites) on the right.

専門家の間でも、データの共有は当然

- Kaggle のデータセットのページ

<https://www.kaggle.com/datasets>

機械学習、データサイエンス、人工知能研究等のため



The screenshot shows the Kaggle Datasets page. At the top, there is a search bar and a button to "Create Public Datasets". Below this, a list of public datasets is displayed, sorted by "Hottest". The datasets listed are:

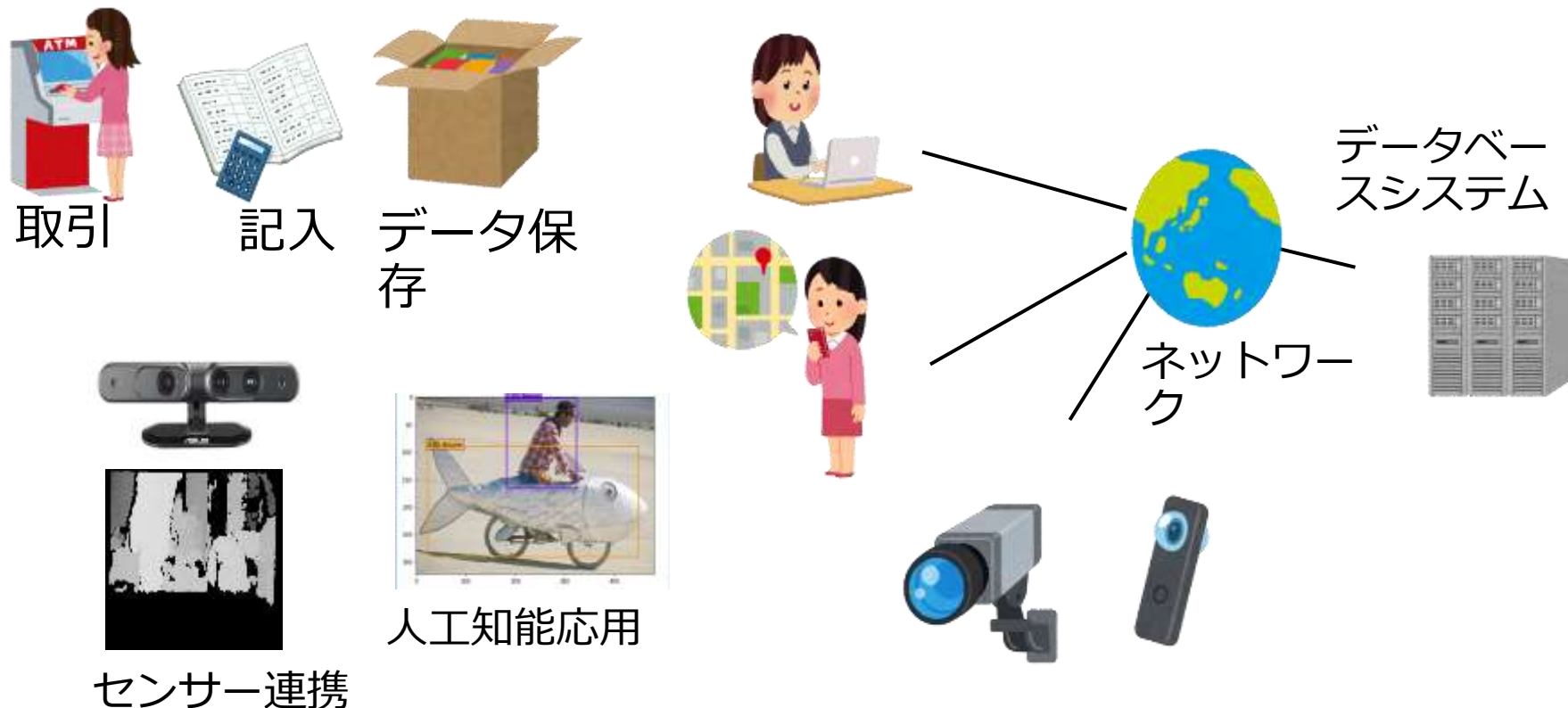
- World Health Statistics 2020[Complete][Geo-Analysis]**: Zeus - Updated 3 days ago. Usability 10.0 - 3D Files (CSV) - 1 MB - 1 Task. 289 likes, Silver badge.
- Pfizer Vaccine Tweets**: Gabriel Preda - Updated 20 hours ago. Usability 10.0 - 1 File (CSV) - 526 KB - 4 Tasks. 128 likes, Silver badge.
- Covid-19 Case Surveillance Public Use Dataset**: MSBiza - Updated 23 days ago. Usability 10.0 - 1 File (CSV) - 48 MB - 1 Task. 63 likes, Bronze badge.
- Women Representation in City Property San Francisco**: Ashwani Rathore - Updated a month ago. 33 likes.

On the right side, there is a section for "Open Tasks" with several tasks listed, such as "Analysis", "Exploratory & Statistical Analysis", "Fix latitude and longitude data", "Train a model to check if a...", and "Are the tweets are really off".

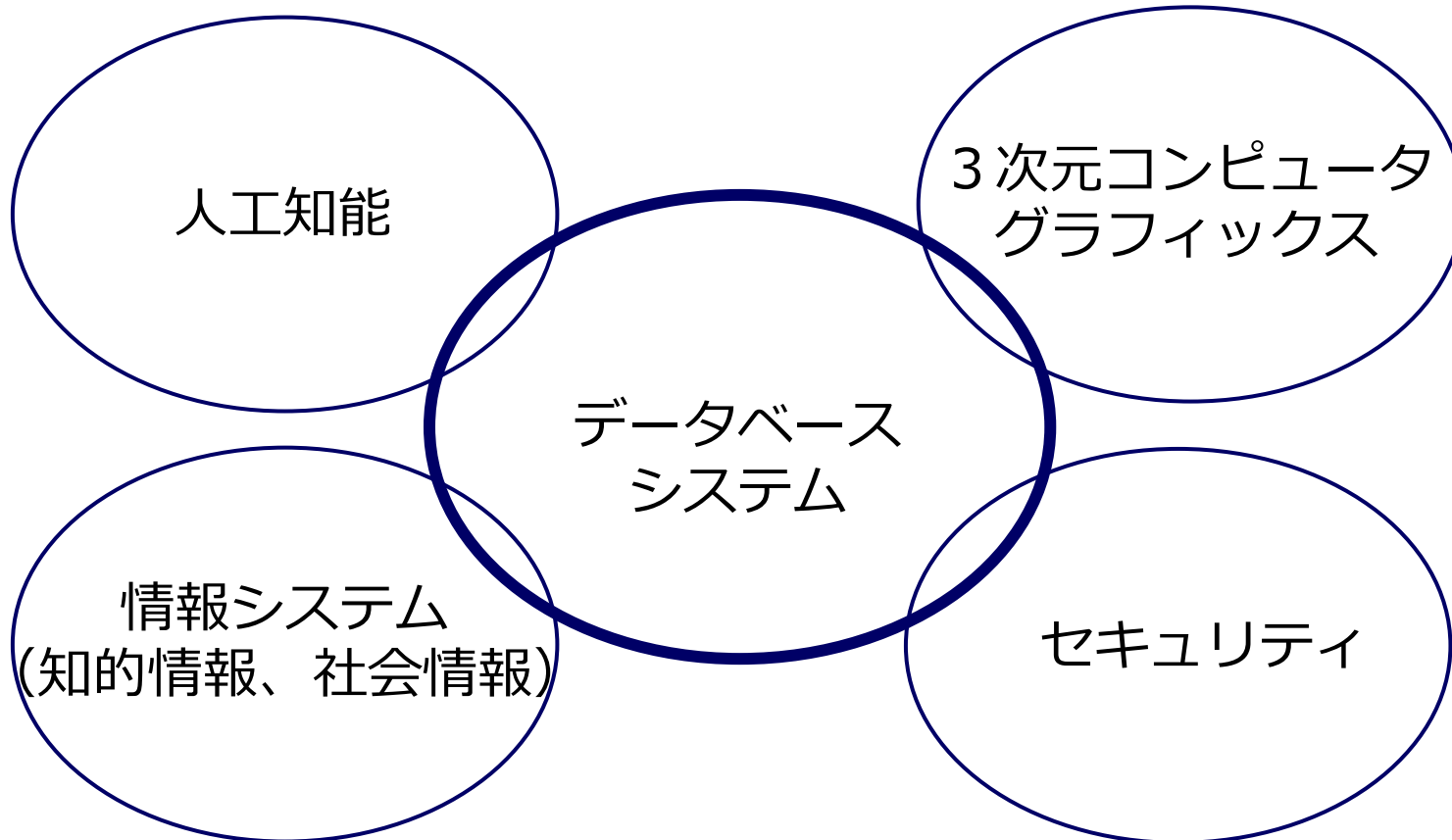
データベースシステム とその周辺分野

データベースシステム

- ・データベース管理システムは、データベースの管理等の機能を持ったソフトウェア
- ・オンラインでデータを共有するときに、特に適する



「データベースシステム」と周辺研究領域



データベースシステムとAI（人工知能）の融合 人工知能の学習にはデータが必要

モーションキャプチャ実験（マーカーレス）



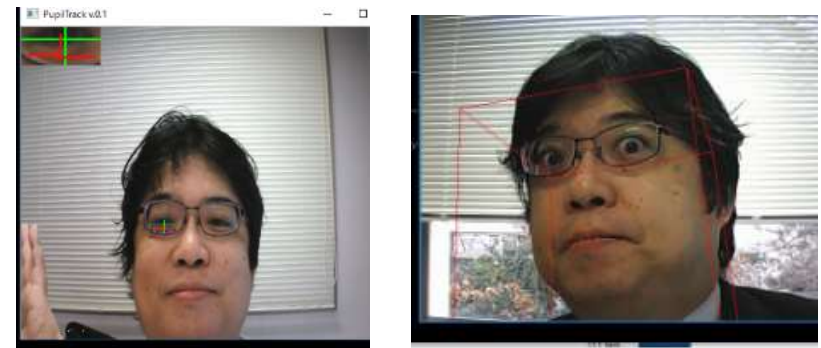
顔検出実験



ナンバープレート自動読み取り実験



眼球運動、顔の動きの自動抽出実験



データベースシステムと社会情報 データの収集, 蓄積, 活用で, データベース システムが役に立つ

情報共有型地図アプリの製作

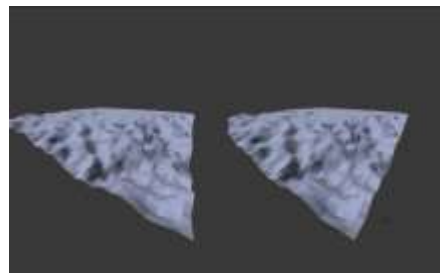
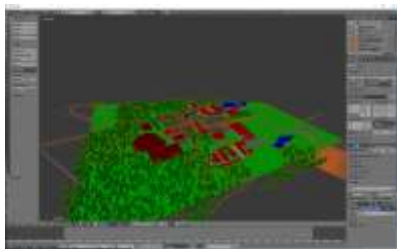


- プライベートな
グループ向けアプリ
- 地図表示
 - 位置表示
 - 写真投稿
 - メッセージやり取り

車両観測システムの製作

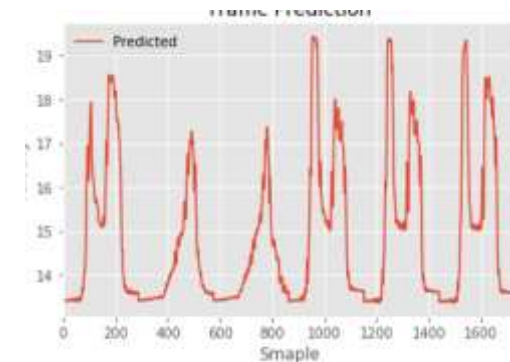


広域3次元地図の製作



ヴァーチャル・リアリ
ティ表示も可能

センサーからの交通密度プロット



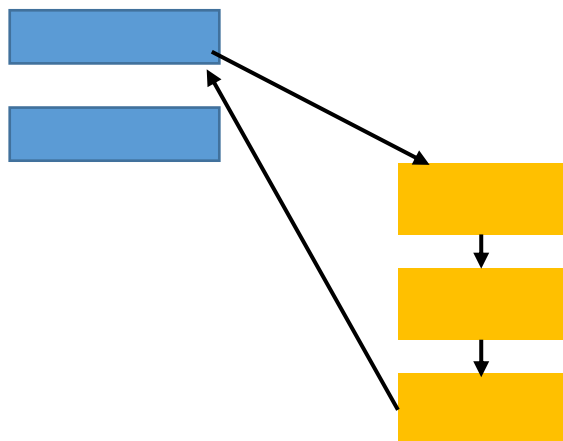
リレーショナル データベースシステムの歴史

リレーショナルデータベースシステムの歴史

- 1969年（E.F. Codd at IBM research）リレーショナルデータベースの提唱
- 1970年代前半から 実際のシステムが登場

従来のプログラミング言語や、他のデータベースシステムで出来なかったことができるようになり、大きく普及。

いまでも、**リレーショナルデータベースは、データベースシステムの主流**



関係のあるレコード同士を
ポインタでつないでおく

デューブルは**ダイナミック**に
結合する

リレーショナル
データベースシステム **以前**

リレーショナル
データベースシステム

大きな変化。データ管理が簡単に

データベースシステムの種類

データベースシステムの種類

① リレーショナルデータベースシステム

	路線コード番号	事業者コード番号	路線名称一般	路線名称一般カナ
1	1001	3	中央新幹線	チュウオウシンカンセン
2	1002	3	東海道新幹線	トウカイドウシンカンセン
3	1003	4	山陽新幹線	サンヨウシンカンセン
4	1004	2	東北新幹線	トウホクシンカンセン

テーブル

② ドキュメント指向のデータベースシステム

XMLなどのドキュメントに特化.

```
<?xml version="1.0" encoding="UTF-8"?>
<osm version="0.6" generator="CGImap
0.0.2">
  <bounds minlat="54.0889580"
minlon="12.2487570" maxlat="54.0913900
maxlon="12.2524800"/>
  <node id="298884269" lat="54.0901746"
lon="12.2482632" user="SvenHRO"
uid="46882" visible="true" version="1"
changeset="676636" timestamp="2008-09-
```

ドキュメント

③ キー・バリュー形式のデータベースシステム (キー・バリュー・ストア)

キー (鍵) バリュー (値)

45343430

金子邦彦

キー・バリュー

その他さまざま

リレーショナルデータベースと オブジェクトの統合

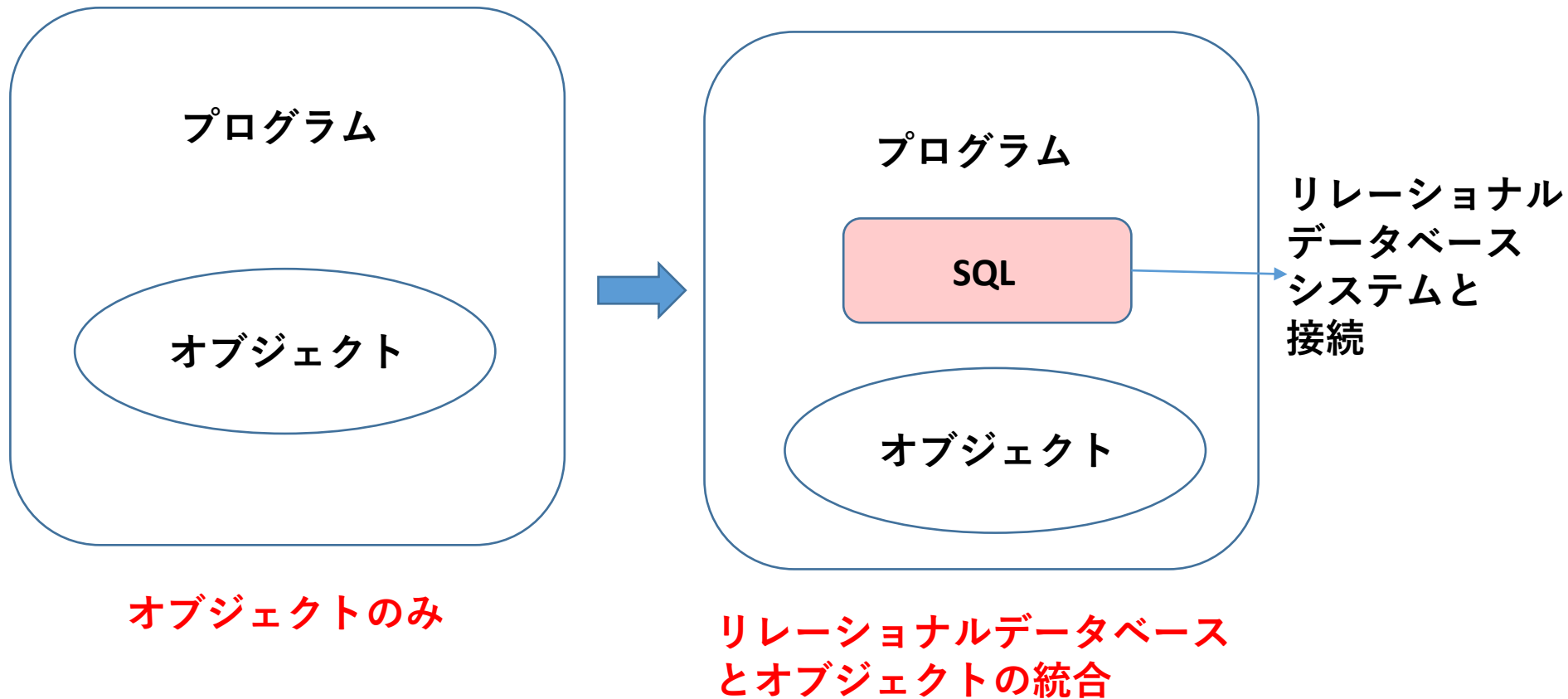
オブジェクト

- **オブジェクト**は、プログラムの中で扱うデータのかたまり

オブジェクト

- オブジェクトは、構造を持つ場合がある
構造体, 集合, リスト,
ベクトル, 行列, 配列など
- オブジェクトには種類がある = 「クラス」と
言ったりする
- オブジェクトの種類ごとに、振る舞いを関連づける
ことができる = 「メソッド」と言ったりする

リレーショナルデータベースとオブジェクト



リレーショナルデータベースとオブジェクト

d はオブジェクト

```
> library(dplyr)
> d <- data.frame(
+   date=c('4月', '5月', '6月'),
+   value=c(100, 120, 150) )
> library(sqldf)
> sqldf("select * from d") d はオブジェクト
  date value
1  4月   100   SQL
2  5月   120
3  6月   150
> sqldf("select * from d where value > 100") d はオブジェクト
  date value
1  5月   120
2  6月   150
```

R システムでの例

リレーショナルデータベースとオブジェクト

前ページのプログラムで行っていること

全体で
オブジェクト d

	data	value
1	4月	100
2	5月	120
3	6月	150

SQL も実行可能

```
select * from d;
```

```
select * from d where value > 100;
```

オブジェクト d が、テーブル d
にマッピングされている

オブジェクトとテーブル

オブジェクト

プログラムの中で扱うデータのかたまり
プログラム

C, C++, Java, Python, R, . . .

数値、文字列、日付、バイト列、
**構造体、集合、リスト、ベクトル、行列、
配列**など

保存したいときは、ファイルの読み書き
などを行うプログラムを書く

データ管理は、自分で頑張る (アルゴ
リズム、データ構造)

テーブル

リレーショナルデータベースのテーブル

SQL

数値、文字列、日付、バイト列など
(単純な構造のデータに限定)

基本は自動保存

リレーショナルデータベース管理システム
が、データ管理のためのさまざまな機能
を持つ

ここまでのまとめ

リレーショナルデータベースシステムの**テーブル**と、
オブジェクト指向の**オブジェクト**

- **テーブル**と**オブジェクト**の**マッピング**により、両方に機能をあわせもつことに。
- 状況に応じて、良い方を使用

データの種類

- テーブル
 - 構造体, 集合, リスト, ベクトル, 行列, 配列
などの構造をもったデータ
 - マルチメディア
 - テキスト／文書
 - 静止画
 - 動画
 - 音声
 - グラフィックス
- など

マルチメディア

画像、ビデオ
地図
グラフィックス
音声
音楽



マルチメディアのコンテンツ

+

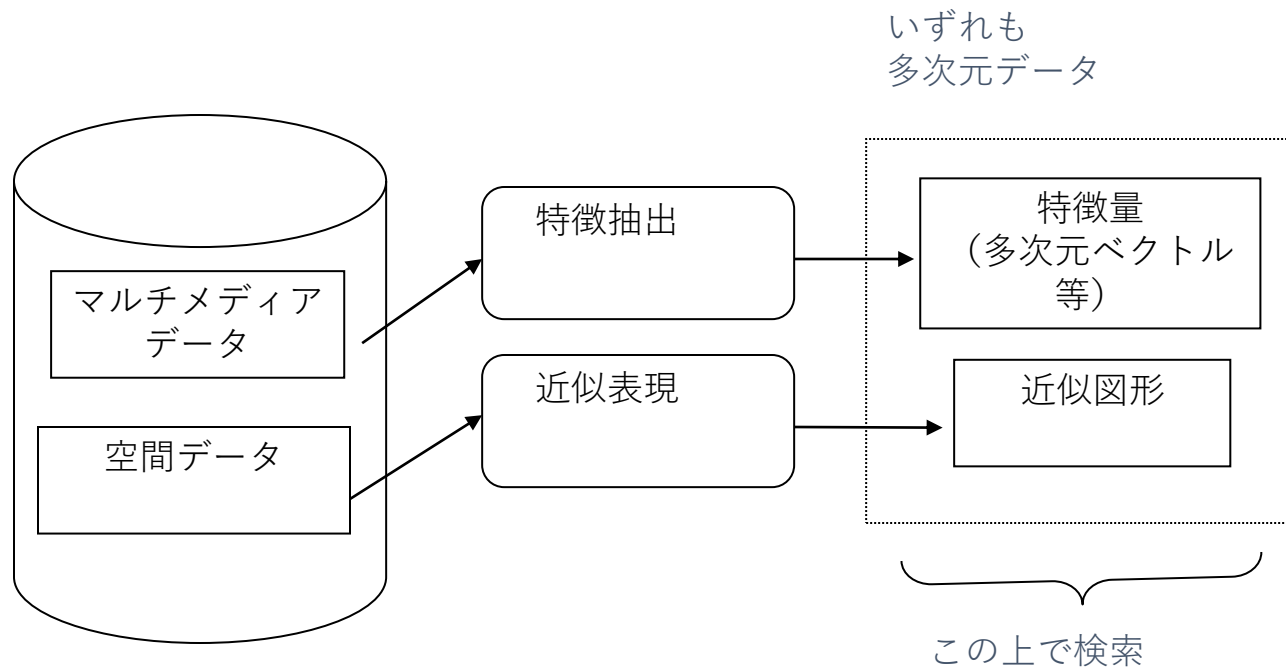
付加情報は、
検索時の有効な手がかり

タイトル
説明文
感想、批評
レーティング（点数）
価格
制作日
制作者
など

付加情報

全体で、
1つのオブジェクト
として扱うことも
多い

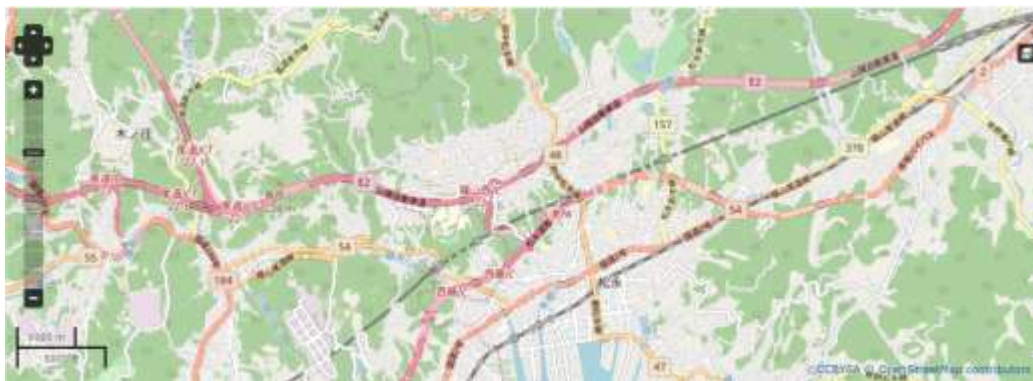
マルチメディアデータベースシステム の枠組み



ドキュメント指向のデータベース

XML形式ドキュメント

地図, グラフィックス, 書式付きの文書などは
XML で扱われることが多い



OpenStreetMap の地図表示

```
<?xml version="1.0" encoding="UTF-8"?>  
<osm version="0.6" generator="CGImap  
0.0.2">  
  <bounds minlat="54.0889580"  
minlon="12.2487570" maxlat="54.0913900  
maxlon="12.2524800"/>  
  <node id="298884269" lat="54.0901746"  
lon="12.2482632" user="SvenHRO"  
uid="46882" visible="true" version="1"  
changeset="676636" timestamp="2008-09-
```

データ (XML形式)

XML は, タグでマークアップ
された**テキストファイル**

キー・バリュー形式のデータベース

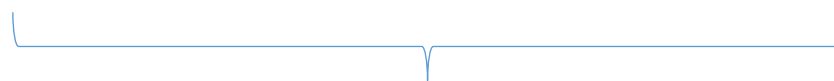
キー・バリュー形式のデータの例

ID

パスワード, 氏名, 住所



キー



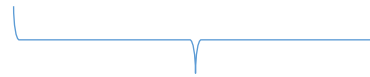
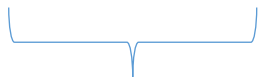
バリュー

キー・バリュー形式のデータの例

	name	price
0	apple	10
1	orange	20

0, name

apple



キー

バリュー

キー・バリュー形式のデータを扱う 2つの方法

- ファイル

JSON などの方法

- キー・バリュー形式のデータベースシステム

JSON とは

「タグ」と「値」を並べることを特徴とする
データ形式

```
{"name": "apple", "price": 10}
```

{タグ: 値, タグ: 値}

JSON の例

```
In[25]: x = {"name": "apple", "price": 10}
In[26]: print(json.dumps(x))
{"name": "apple", "price": 10}

In[27]:
```

Pythonで扱うことも簡単

JSON のデータ型

- 配列 例 [1, 2, 3]
- 数値 例 4.56
- 文字列 例 "hello" ※ 「"」で囲む
- ブール値 true false
- null

```
{"name": "apple", "price": 10}
```

- **タグ** 必ず文字列
- **値** さまざまな**データ型**でありえる。
JSON オブジェクトを値とすることも

JSON での入れ子

入れ子： **値が JSON オブジェクト** になること

id	name	price
0	apple	10
1	orange	20

リレーショナルデータベースのテーブル



同じ意味

```
{"name":{"0":"apple","1":"orange"},"price":{"0":10,"1":20}}
```

JSON での入れ子

JSON とキー・バリュー形式のデータベースシステム Cloud Firestore の違い

	JSON	Google Firebase
オブジェクトの種類	JSONオブジェクト	コレクション , ドキュメント
入れ子	OK	OK
「値」の種類	文字列 (string), 数値 (number), ブール値 (boolean), 配列 (array), null	文字列 (string), 数値 (number), ブール値 (boolean), 配列 (array), null, 日時 (timestamp), geopoint, 参照 (reference)