

帝塚山学院大学  
TEZUKAYAMA GAKUIN UNIVERSITY

データベース概論  
GoogleのDB,ビッグデータ

中野秀男  
情報メディア学科/ICTセンター長

1 データベース概論 Google, Big Data 2017/1/30

---

---

---

---

---

---

---

---

今日の話

- ▶ 質問やコメントの回答
- ▶ ガーグルのDB
- ▶ ビッグデータ
  - ▶ ビッグデータ
  - ▶ ビッグデータの例
  - ▶ ビッグデータの利用方法

2 データベース概論 Google, Big Data 2017/1/30 帝塚山学院大学  
TEZUKAYAMA GAKUIN UNIVERSITY

---

---

---

---

---

---

---

---

Googleのデータベース

- ▶ 世界中のウェブを自動巡回プログラム(ロボット)で収集
- ▶ 安価なコンピュータや補助記憶でDBを構成
- ▶ 列(フィールド)の単位で整理せずに、行の単位で整理
- ▶ 複数の行(文書)を複数のコンピュータに保存
- ▶ 検索要求(クエリ)はありそうな複数のコンピュータに要求
- ▶ 多くの同時にくる大量のクエリに平均的に対応
- ▶ シャーディング: データの断片化
- ▶ レプリケーション: 複製
- ▶ ページランク

3 データベース概論 Google, Big Data 2017/1/30 帝塚山学院大学  
TEZUKAYAMA GAKUIN UNIVERSITY

---

---

---

---

---


---

---

---

ビッグデータ (Wikipedia)

- ▶ 通常のデータベースでは扱えない量
- ▶ 構造化データや非構造化データ
- ▶ ペタ、エクサ、ゼタ、更に上
  - ▶ K, M, G, T, P, E, Z
- ▶ 格納や活用(検索/共有/分析/可視化)
- ▶ いろいろな分野で先端技術を使う

4 データベース概論 Google, Big Data 2017/1/30  帝塚山学院大学  
TEZUKAYAMA GAKUIN UNIVERSITY

---

---

---

---

---


---

---

---

ビッグデータの例

- ▶ 政府や自治体を持つ住民や組織情報
- ▶ 家電やシステムの操作ログ
- ▶ ライフログ
- ▶ センサーネット情報
- ▶ 顧客データ
  - ▶ 政府、自治体、企業、NPO
- ▶ ソーシャルメディアデータ
- ▶ マルチメディアデータ
- ▶ ウェブサイトデータ

5 データベース概論 Google, Big Data 2017/1/30  帝塚山学院大学  
TEZUKAYAMA GAKUIN UNIVERSITY

---

---

---

---

---


---

---

---

ログ

- ▶ 操作ログ
  - ▶ 家電メーカーの思惑
  - ▶ B-CAS問題
  - ▶ フォレンジック(PC操作の記録)
- ▶ ライフログ
  - ▶ 歩数等の活動の記録
  - ▶ 読んだもの等の知識の記録
  - ▶ 写真等の記憶の記録
  - ▶ 体重や脳波等の身体の記録
  - ▶ レシート等の身体の記録
  - ▶ 位置情報等の環境の記録
  - ▶ 気分等の精神の記録

6 データベース概論 Google, Big Data 2017/1/30  帝塚山学院大学  
TEZUKAYAMA GAKUIN UNIVERSITY

---

---

---

---

---


---

---

---

ビッグデータのステップ

- ▶ 多様な複数のデータを
- ▶ 総合的に収集・分析することで
- ▶ 可視化する/新たな知見を得る
- ▶ プラットフォームの構築(高速、大容量)
- ▶ サードパーティの参加(複数で多く)

▶ 7 データベース概論 Google.Big Data 2017/1/30  帝塚山学院大学  
TEZUKAYAMA GAKUIN UNIVERSITY

---

---

---

---

---


---

---

---

時代の背景

- ▶ コンテンツ:
  - ▶ ソーシャルメディア: FBは300億/月
  - ▶ 動画配信: YouTube
- ▶ プラットフォーム
  - ▶ クラウド、Android OS
- ▶ ネットワーク
  - ▶ ワイヤレス通信、ブロードバンド
- ▶ デバイス
  - ▶ 各種センサーの普及
  - ▶ ICT端末の多様化: 携帯, PC, TV, スマホ

▶ 8 データベース概論 Google.Big Data 2017/1/30  帝塚山学院大学  
TEZUKAYAMA GAKUIN UNIVERSITY

---

---

---

---

---


---

---

---

仮説を検証

- ▶ 仮説を立て
- ▶ 複数の大量のデータを集め
  - ▶ 大容量/分散ストレージ
- ▶ データマイニング等の手法で推論(?)
  - ▶ 高速/並列計算、アルゴリズム
- ▶ 例
  - ▶ 紙オムツを買った人は缶ビールも買う
  - ▶ 冬にスキー用品を買った人は、夏にテニス用品を

▶ 9 データベース概論 Google.Big Data 2017/1/30  帝塚山学院大学  
TEZUKAYAMA GAKUIN UNIVERSITY

---

---

---

---

---

---

---

---

質問やコメント(1)

- ▶ データベースの整合性を保つのは大変
- ▶ 半構造にしておく柔軟性が
- ▶ SQLの限界、なぜ曖昧な記述ができないか
- ▶ GoogleのDBの次の展開は
- ▶ Googleはアルゴリズムを時々変える
- ▶ Googleのデータの分断化のシャーディングとは
- ▶ Googleのページランクの基準はアクセス数か
- ▶ GoogleのDBで知らない単語がたくさん
  - ▶ アルゴリズムにも関係している
- ▶ Googleは自分のことを全て知っていると思うと少し恐ろしい

▶ 10

データベース概論 Google, Big Data 2017/1/30




---

---

---

---

---

---

---

---

質問やコメント(2)

- ▶ 要求するときに曖昧な要求だと間違った答えが
- ▶ Accessは使うときに覚える
- ▶ DB使っていない組織は生き残れない、信用されない
- ▶ メタデータがマルチメディアDBのキーポイント
- ▶ マルチメディアDBは曖昧でよくわからなかった
  - ▶ 要求の曖昧さが困難になっている理由
- ▶ 検索した文字と全然関係ないサイトが出る時がある
- ▶ 自分の名前で検索するのは怖くてたくない
- ▶ ウェブ上で自分の写真を勝手にあげられていたら取り消すことは可能か

▶ 11

データベース概論 Google, Big Data 2017/1/30




---

---

---

---

---

---

---

---

質問やコメント(3)

- ▶ 世界中のウェブをロボットが収集する
  - ▶ これからどんな変化が
- ▶ サイトでロボットでない証明のために文字入力や画像を選ぶ
- ▶ 顔写真で人を特定したり、写真から指紋を取るのは怖い
- ▶ なんでもコンピュータ処理でも人は必要
- ▶ 堺市の情報漏えい問題、ネットは怖い
- ▶ Twitterで炎上すると住所や学校や名前まで特定される

▶ 12

データベース概論 Google, Big Data 2017/1/30




---

---

---

---

---

---

---

---

質問やコメント(4)

- ▶ 中野ややったゲームで一番面白かったのは
- ▶ VR以上にゲームの世界に入れるゲームは出るか
- ▶ VRの次はどんな機能が。匂いとか
- ▶ 少しの誤字があっても正しい情報に誘導してくれる
- ▶ データを分類やタグ付けして整理するのは重要
- ▶ 資格は挑戦してみたい
- ▶ 自作でロボットやパソコンを作っている人
  - ▶ ロボットのプログラミングも

▶ 13

データベース概論 Google.Big Data 2017/1/30




---

---

---

---

---

---

---

---

質問やコメント(5)

- ▶ リスト、スタック、キューは覚えておきたい
- ▶ 木の構造
- ▶ アルゴリズムは何度勉強しても理解できない
- ▶ 授業アンケートもe-learningでできる
- ▶ 任天堂のスイッチが欲しい
  - ▶ 高い
- ▶ 任天堂スイッチは食堂で多くの学生が使うだろう
- ▶ 任天堂は色々なものを生み出している
- ▶ 任天堂スイッチのグラフィックがWiiUと対して変わらない
- ▶ 任天堂スイッチはスマホゲームに流れたユーザを戻すか

▶ 14

データベース概論 Google.Big Data 2017/1/30




---

---

---

---

---

---

---

---

質問やコメント(6)

- ▶ より若い人がプログラミングスキルを
- ▶ RSAの暗号は電子ショッピングで使われていて量子コンピュータができると解読される
- ▶ 服屋のディスプレイや会計でもiPadが
- ▶ SNSほど大きな窓はない
- ▶ 中野は顔が広い
- ▶ ポケモンGOプラスは買いますか？メリットが

▶ 15

データベース概論 Google.Big Data 2017/1/30




---

---

---

---

---

---

---

---