

1-2 社会で活用されているデータ

どんなデータが集められ、どう活用されているかを知る

様々な種類のデータ

- 調査データ: 調査して収集されたデータ(アンケートなど)
- ログデータ: 自動的に収集されたデータ(位置情報, ネット履歴など)
- 実験データ: 測定されたデータ(ABテストなど)
- 観測データ: 観測して得られたデータ(気象, 天体など)

1次データ・2次データ・メタデータ

- 1次データ:自分で集めたオリジナルデータ
- 2次データ:他人が過去に集めたデータ
- メタデータ:データを説明するデータ(タイトル, 作成者, 公開日など)
- オープンデータ:公開データ

構造化データと非構造化データ

- 構造化データ: 表など
- 非構造化データ: 画像, 動画, 音声など
- アノテーション: 文章や画像などに名称を「ラベル付け」すること

アノテーションの例



参考文献

- 吉岡剛, 森倉悠介, 小林領, 照屋健作: AIデータサイエンス入門, 技術評論社(2022)
- 伊藤大河, 川村和也, 内田瑛, 河合麗奈: 大学基礎データサイエンス, 実教出版(2023)
- 東京大学 数理情報教育研究センター,
http://www.mi.u-tokyo.ac.jp/consortium2/pdf/1-2_literacy_level_note.pdf