

プログラミング言語研究室

～大規模なデータを効率よく扱う～

千代 英一郎

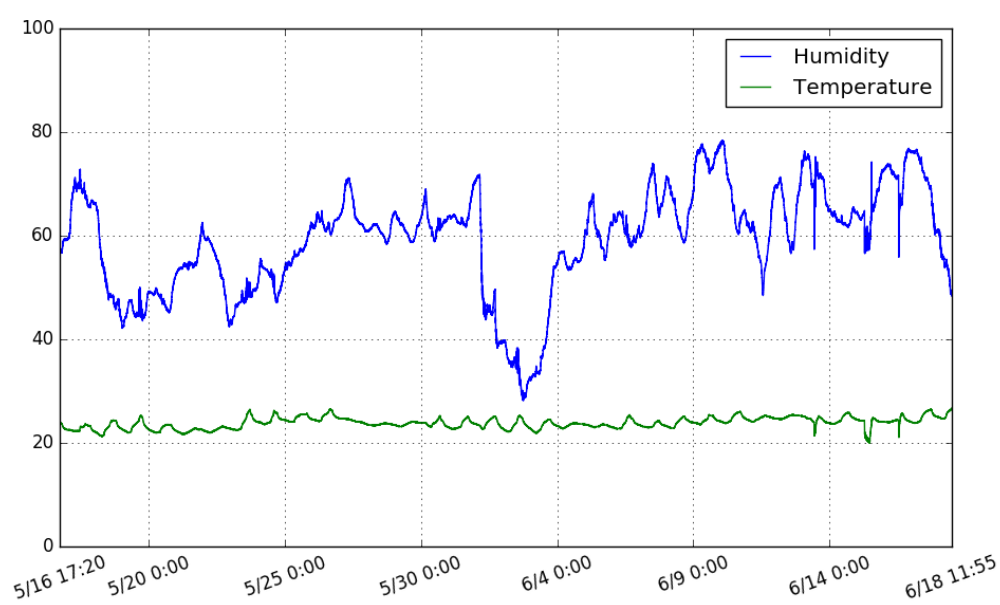
背景

世界はデータで溢れている

- ・ウェブサイト
- ・画像・動画・音声
- ・センサデータ



ウェブサイトの
テキスト・動画



温度・湿度の
センサデータ

コンピュータの 得意分野

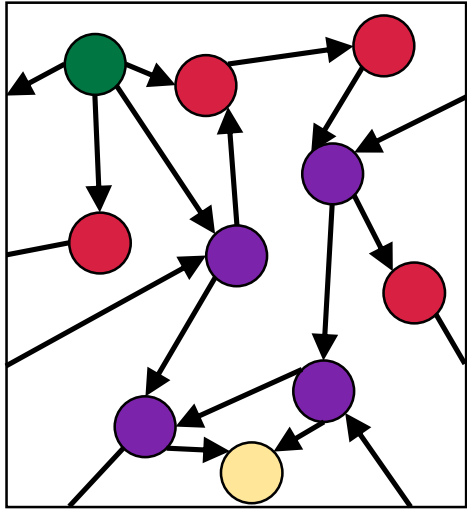
- 計算
- 記録
- 検索
- 自動化

有効に活用するためには、コンピュータの得意分野を生かして
正しく・賢く・効率よくデータを処理するしくみが必要

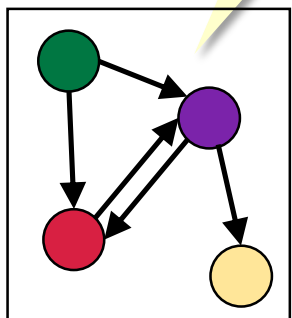
研究室の成果

大規模グラフの分析と検索

グラフデータ



要約



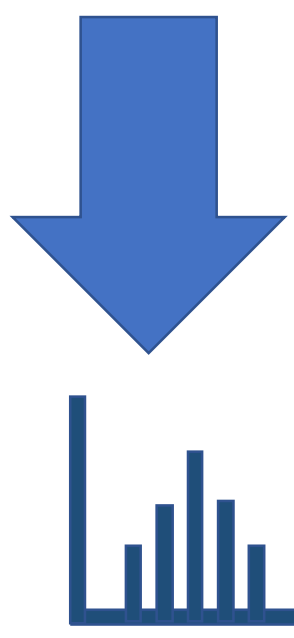
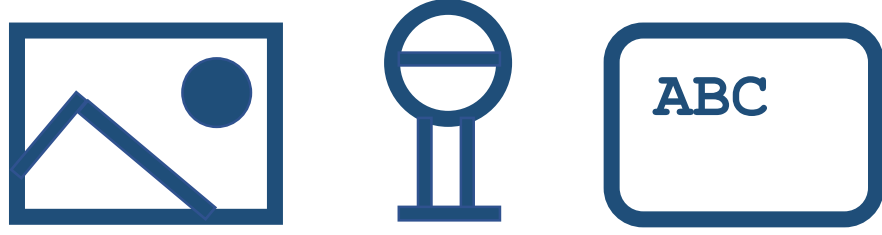
データのつながりに
注目して、データを
グループに分ける

元のデータの構造を要約した
データを利用して、効率よく
データを分散・並列処理する



データの特徴抽出・合成

画像・音声・フリーテキストなどの
非構造データは、そのままでは
検索・分析・加工しにくい



学習により
データの特徴量
を抽出

応用例

ニコニコ動画©の全動画情報を解析

一定期間に投稿された動画の再生回数上位の動画を
もとに、動画に付けるタグの推定情報を抽出

要約情報	2015年4月
あり	2.08秒
なし	148.97秒

検索時間が $\frac{1}{70}$ に短縮!

カスタムフォントの合成

既存のフォントデータから抽出した
特徴量を用いて自分好みのフォント
を合成

