

まえがき

「ビッグデータ (Big Data)」という用語は、2010年頃から特にビジネスの分野を中心によく見かけるようになりました。しかし、多くの人は、ビッグデータとはいいったい何なのかを未だによくわかつていらないのが現状だと思います。実際、ビッグデータそのものには明確な定義はなく、話の文脈に応じて、大規模なデータそのものや、データが爆発している状況、大規模データを解析する技術、解析して知見を得る一連の行為など、様々な意味で用いられています。つまり、非常に便利な反面、実体をつかむのが困難な用語です。この状況をよく表しているのが、Wikipedia のビッグデータに関するエントリでしょう。頻繁に内容が更新されていますが、解説の方向性が日増しに発散し、筆者のようなデータベースの専門家が読んでも、すでに解読困難なものになっています。

そこで筆者は、本書である「ビッグデータ解析の現在と未来」を執筆することを決心しました。もちろん、ビッグデータ 자체があいまいな用語であるため、本書の視点も主観を多く含んだものになっています。しかし、データベース研究者である筆者が、ビッグデータとは何か、ビッグデータに関わる技術や課題とは何か、常日頃から考えている内容を整理して執筆した本書は、ビッグデータの初步について知りたい、学びたいと思っている読者にとっては意義のあるものだと信じています。

本書では、まず1章として、ビッグデータおよびその解析技術に

について概観します。特に、ビッグデータの特徴や注目されたようになった背景、ビッグデータの解析技術の特徴などについて解説します。2章では、ビッグデータ解析の応用事例について、代表的なものを紹介します。3章では、以降の章における技術的な解説的理解を促進するために、ビッグデータ解析の典型的な流れについて、データ収集とデータ解析に焦点を当てて解説します。4章、5章、6章、7章では、ビッグデータを支える技術として、分散処理フレームワーク、ストリーム処理エンジン、データベース、機械学習について代表的な技術をそれぞれ紹介します。次に8章では、ビッグデータ解析が今以上に広く浸透するために重要なオープンデータについて、国内外の動向を踏まえて解説します。最後に9章では、本書のまとめとして、ビッグデータに関する将来の動向や可能性について議論します。