



連載

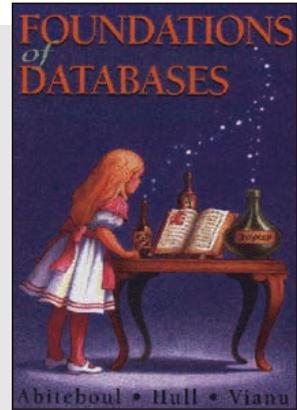
ビブリア・トーク
—私のオススメ—

… 加藤弘之 (国立情報学研究所)

Foundations of Databases

Serge Abiteboul, Richard Hull, Victor Vianu 著

Addison Wesley (1994), 685p., ISBN : 978-0201537710



いまさらデータベースの基礎?

さつと表紙の絵だけを見てみると女の子が何やら魔法の勉強をしているような姿が写っています。自然結合の関係代数式や述語論理式が書いてある魔法の壺が机の上のっています。この本で扱っている魔法は、データベース理論です。「データベースの基礎的な理論なんてどうせ古臭い関数従属性の話でしょ。今さらそんなの勉強してどうするの? 役に立たないよねー」なんて声が聞こえてきそうです^{☆1}。そんな古臭いと思われる本ですが、ACMの電子図書館によると引用数が1,607(2017年だけでも58)ある著名な本です。

まずは誰が書いたどんな本なのかを簡単に紹介します。この本の3人の著者 Abiteboul, Hull, Vianu は、南カリフォルニア大学に同じ時期に(学生または教員として)在籍していました。そのときに同大の Seymour Ginsburg 教授^{☆2} から多大な影響を受けデータベース理論の研究を行うようになったようです。それもあって、冒頭でこの本を Ginsburg に捧げると書いています。Ginsburg の影響もあってかこの本は全体として論理(一階述語論理やデータログ)に基づくデータベース理論を中心に記述されています^{☆3}。

^{☆1} 古臭いと思われる関数従属性ですが、近年のデータ交換 (data exchange) に関する研究課題では、組生成従属性 (tuple-generating dependencies, tgds) が積極的に使われていますし、これを用いた製品もあります。

^{☆2} 著名な計算機科学者で、彼が始めた研究ミーティングはその後データベース理論に関する国際会議である、ACMPODS (Symposium on Principles of Database Systems) に発展したことで知られています。また、彼は Ullman が形式言語の研究を始めるきっかけを与えたと言われています。

^{☆3} データベース理論別の側面として、プログラミング言語の観点からの見方もありますが、このあたりに興味のある方は、2014年のICDT (International Conference on Database Theory) Test-of-Time Award 受賞論文である「Naturally Embedded Query Language (ICDT1992)」がおすすです。

なお、表紙に出てくる女の子ですがアリスという名前で、各章の冒頭でアリスと3人の著者の一言ずつの掛け合いがあることから、別名アリス本 (Alice Book) と呼ばれています。たとえば、第1章の冒頭ではアリスが次のように問いかけます。「この本は理論の本だと思っていたわ。」これに対して3人の著者がそれぞれ、「そうだよ。でも良い理論には、それが示す全体図が必要なんだ。」「それに、君のおじいさんから、いったい何を勉強しているんだって聞かれたときに、君はなんて答える?」「まさか、データベース問合せにおける普遍性の基礎的含意について勉強しているって言っても、君のおじいさんには分かってもらえないだろう。」といった感じで、第1章では、データベースシステムの全体像について説明しています。このようなウィットに富む掛け合いが、各章の冒頭で繰り広げられていますので、これらだけでも面白く読むことができます。嬉しいことに出版社の好意で、Web上からこの本のPDFを無償で入手できる^{☆4}ので、気軽に読むことができます。

この本の特徴は?

この本は、学部の3年生から大学院生向けの内容となっていて、およそ700ページもある分厚い本です。この本の大きな特徴は、関係データベースの問合せ言語を論理に基づいて丁寧に解説している点にあります^{☆5}。SQLの論理的な基礎である関係論理

^{☆4} <http://webdam.inria.fr/Alice/>

^{☆5} 他の著名なデータベースの教科書は、他の要素たとえば、SQLの詳しい文法やデータベース管理システムの実現法、トランザクション処理などにもページ数を割いていますが、この本ではあくまでも論理に基づく理論的な部分に焦点を当てています。

やデータログについてページ数を割いて丁寧に記述しています。また、静的解析に基づく最適化手法についても幅広い技術 (Chase や Magic Set など) について丁寧に解説しています。

この本は、Part A から Part F に分かれていて各 Part は 3, 4 章から構成されています。ざっと内容を簡単に紹介します。

- Part A : ANTECHAMBER (はじめに)
データベースシステムについての全体像, 理論的背景, 関係データモデルを概観しています。理論的背景部分を読むにはある程度の知識が必要なので, 計算機科学の初歩的理論についてまだ学習していない人は, 適当に読み飛ばしてもよいかもしれません。
- Part B : BASICS : RELATIONAL QUERY LANGUAGES
関係データベースの問合せ言語について, 主に論理的側面から紹介しています。まずは, 否定のない簡単なクラス問合せのクラス (conjunctive queries) について紹介し, そのあとで否定が入ることでの難しさを説明しています。問合せの書き換えに基づく最適化についても紹介しています。
- Part C : CONSTRAINTS
関数従属性などの制約について, 簡単な制約から徐々に複雑な制約について紹介し, すべての従属性を記述できる一般性についても紹介しています。これらの制約を使ったデータベース設計についても簡単に紹介しています。また, 制約を使った代表的な最適化である「Chase」について紹介しています。
- Part D : DATALOG AND RECURSION
再帰を記述するためのデータログという論理的言語について紹介しています。データログに否定が入ったときの意味で知られている層状意味論 (stratied semantics) や, 商用のデータベースエンジンですでに実用化されている最適化の 1 つである Magic Set について紹介しています。

- Part E : EXPRESSIVENESS AND COMPLEXITY
(問合せ) 言語の記述能力 (expressiveness) と計算の複雑さ (complexity) は, トレードオフの関係になっていることを, チューリングマシンを使った例を用いて丁寧に説明しています。その後で, 実用的な言語である PL/SQL の基礎となっている言語の理論的な基礎について, 論理的な側面から紹介しています。
- Part F : FINALE
最後の話題としては, 不完全な内容のデータを扱う「制約データベース」, 関係データモデル以外のデータモデル, ビュー更新問題や Active データベースなどの動的側面について, 紹介しています。

読み方も人それぞれ

データベースの論理的基礎についてじっくりと勉強したい人は, 最初から丁寧に読んでいくことをおすすめします。また, 必要に応じて該当箇所を読むことでその部分の内容を理解できるので, 辞書的な読み方もできます。たとえば, 最近のデータ交換で用いられている組生成従属性を用いたスキーママッピングと, それに基づくデータ生成のための Chase を調べることもできます。また, Magic Set を深く理解するのに, Ullman の教科書¹⁾ や森下先生の本²⁾ などと比較しながら該当箇所を読んでみるのもよいかもしれません。

参考文献

- 1) Ullman, J. D. : Principles of Database and Knowledge-Base Systems, Vol.II, Computer Science Press (1989).
- 2) 森下真一: 知識と推論, 共立出版 (1994).
(2018 年 8 月 14 日受付)

加藤弘之 (正会員) kato@nii.ac.jp

国立情報学研究所アーキテクチャ科学研究系助教。データ相互運用問題, XQuery の最適化, MapReduce の最適化など, データベースプログラミング言語の分野を中心に研究に従事。博士 (工学)。