



Title	検索処理の高速化をめざしたデータベーススキーマ設計に関する研究
Author(s)	中西, 通雄
Citation	大阪大学, 1995, 博士論文
Version Type	VoR
URL	https://doi.org/10.11501/3100702
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

氏 名	中 西 通 雄
博士の専攻分野の名称	博 士 (工 学)
学 位 記 番 号	第 1 1 6 8 1 号
学位授与年月日	平成 7 年 2 月 16 日
学位授与の要件	学位規則第4条第2項該当
学 位 論 文 名	検索処理の高速化をめざしたデータベーススキーマ設計に関する研究
論文審査委員	(主査) 教授 橋本 昭洋 (副査) 教授 都倉 信樹 教授 西尾章治郎 教授 萩原 兼一 助教授 松田 秀雄

論 文 内 容 の 要 旨

データベースのスキーマは、概念レベル、論理レベル、物理レベルの順に設計される。特に、論理レベルのスキーマの設計手法については、数多くの研究が行われてきたが、データベースに関する検索質問を効率良く処理できるようにするためのスキーマ設計手法については、まだ研究の余地がある。本論文では、検索質問を効率良く処理するためのスキーマ設計手法に関して、まず最初に、関係データベースでの設計手法を理論的に考察し、次に、オブジェクト指向データベースについて、実際の構築経験を基にした設計上の知見を述べる。

関係データベースにおいて、結合を含む検索質問の処理効率を向上させることは重要な問題である。本論文の前半では、定型業務における検索質問の処理効率を向上させるために、データベーススキーマを変形して、結合処理の対象になることが多い基底関係を結合した形で蓄積せることにより、結合処理の数を減らすことを考える。まず、データベーススキーマを変形する2つの操作「複写」と「併合」を定義する。「複写」はある関係スキーマに属する関数従属を別の関係スキーマにも重複して持たせる操作であり、「併合」は2つの関係スキーマを1つにまとめる操作である。データベーススキーマに対してこれらの操作を適用しても、質問を実行したときに新たな不整合を生じないために、質問およびデータベーススキーマが満たすべき必要十分条件を示す。なお、条件の必要性から、その条件を満たす限り併合と複写を適用してえられるデータベーススキーマは、さらに複合や併合を行うと新たに不整合を生じるという意味で最適である。複合や併合を適用することにより得られるデータベーススキーマでは、基底関係を結合した形でデータベースが構成されるので、結果として結合処理を含んだ検索質問の処理効率を向上させることができる。

後半では、オブジェクト指向データベースのスキーマ設計について述べる。たいていのオブジェクト指向データベース管理システムは、関係データベースと同様に検索処理の性能を向上させるために、文字列の属性値に対するインデックスを作成・利用する機能が提供されている。オブジェクト指向データベースではメソッドが定義できるため、メソッドの返す値もインデックスのように扱えれば、検索処理がさらに高速化できる。すなわち、メソッドがオブジェクト識別子を返す場合にも、あらかじめメソッドを実行しておき、その結果のオブジェクト識別子を保存しておくことにより、検索範囲をせばめることができる。この機能は、特に更新の少ないデータベースにおいて有効である。しかし、これを

提供しているデータベース管理システムはない。本論文では、更新の少ないデータの例としてタンパク質構造データをとりあげ、試作したオブジェクト指向データベースについて紹介する。さらに、検索を高速にするための方法、および、この作成経験で得たオブジェクト指向データベースのスキーマの設計についての知見を述べる。

タンパク質の立体構造データは、Protein Data Bank (PDB) というデータバンクにファイルとして登録されている。できるだけ簡潔にデータを検索できるようにするために、PDB のデータをデータベース化する試みが行われてきたが、特に最近 2 年間で PDB の容量は約 3 倍の 1 ギガバイトに増加しており、データを効率良く蓄積・検索することもデータベース化する際の重要な課題である。各タンパク質は複数個の鎖で構成されており、各鎖はアミノ酸群で、さらに各アミノ酸は原子群で構成されている。オブジェクト指向データベースを用いると、この構造を自然にデータベーススキーマに反映させることができるのが検索および蓄積の効率を良くするためには工夫が必要である。例えば、アミノ酸系列を検索する質問が多いので、アミノ酸名の系列だけを別に文字列としても蓄積しておき、文字列検索を行うようとする。また、オブジェクト指向データベースでは、クラス階層にうまく適合しないような検索質問に対しては、検索処理の効率が悪くなるという弱点がある。これを補完するために、検索質問やメソッドをあらかじめ実行し、その結果として得られたオブジェクト識別子の集合を、オブジェクトとして保存できるようにスキーマを設計した。さらに、一つのタンパク質を構成する原子の数は数千個にもおよび、原子座標を対象とした検索が殆んど無いことから、個々の原子をオブジェクトとせずに、まとめて一つのオブジェクトとする方が記憶領域を少なくできる。この結果、試作したデータベースではアミノ酸系列などの検索が高速に実行でき、しかも、データベースのサイズを PDB の約半分に抑えることができた。また、本論文で述べたスキーマ構成方法は、C++ 言語に基づいた別のオブジェクト指向データベース管理システムを用いる際にも適用することが可能である。

論文審査の結果の要旨

データベースは、企業における在庫データなどだけでなく科学技術データも対象として、様々な分野で用いられるようになってきている。また、個々のデータベースで扱うデータ量は指数的に増加しており、利用者からの検索質問をいかに高速に処理するかは、データベース分野の重要な研究課題である。本論文は、関係データベースとオブジェクト指向データベースに関して、特に検索処理を高速にするためのスキーマ構築方法を理論的に考察し、さらに実際のシステムでの適用可能性を検証している。

関係データベースでは、データを更新するときの不整合をなくすために、関数従属性と呼ばれるデータの依存関係に基づいて、データを正規化し、表形式のデータベーススキーマを構築する。この結果、列の少ない表が数多くでき、大量のデータを検索する場合に表の結合 (join) 処理に時間がかかるという問題がある。これを解決するために、実際のデータベースシステムの設計者は、実行頻度の高い定型業務に着目して、データベーススキーマを変形して、これらの業務の検索で用いられる表をあらかじめ結合した形でデータベースに蓄積しておくことにより、検索処理を高速化するという手法を用いてきた。しかし、このようなデータベーススキーマの変形を行っても、データ更新時に不整合が生じないかどうかの判定基準は、明確になっていなかった。本論文は、更新不整合を生じないための必要十分条件を理論的に明確にしている。この結果を用いることにより、スキーマを変形してもよいかどうかが機械的に判定できるようになる。さらに、実現の業務で用いられているデータベースシステムに対する適用可能性を示している。

本論文は、次に、オブジェクト指向データベースについて論じている。科学技術データは量が膨大で、かつ、構造が複雑であることから、そのデータベース化にはオブジェクト指向データベースを利用することが有望である。ここでは、科学技術データの一例として、タンパク質の立体構造データをとりあげ、オブジェクト指向データベースを用いたデータベース化について検討している。特に、タンパク質構造データは追加されることが主で更新が少ないという特徴に着目し、検索処理を高速化するために、(1) 実行されることの多い検索質問をあらかじめ実行して、その結果として得られるオブジェクト識別子を仮想クラスとしてデータベースに蓄積する。(2) オブジェクトの中で特に検索されること

の多い属性をまとめて蓄積する、などのスキーマを構成法を提案している。さらに、実際にデータベースを構築してその有用性を明らかにした。構築には特定のオブジェクト指向データベース管理システムを利用しているが、そのスキーマ構成法は、他の商用オブジェクト指向データベース管理システムを用いた場合にも適用できる汎用的なものである。

以上の通り、本論文の成果は、データベースの応用に関して、学術・科学の発展に寄与するところが大きく、博士論文として価値あるものと認定する。