

Windows マシンにおける Seep の実装

5 A a - 2

江口 肇† 森 健作‡ 竹田 正幸† 松尾 文碩†

†九州大学大学院システム情報科学研究科 ‡九州大学工学部

1. はじめに

本研究室ではデータベースとルールベースを統合して管理する情報知識統合管理システム Seep を開発している。現在、Seep は UNIX と MS-DOSのもとで動作する。九州大学大型計算機センターでは、Seep を用いて構築された核酸塩基配列データベースシステム GENAS や結晶構造データベースシステム XDT の検索サービスを行っている^{1, 2)}。また、INSPEC テープ³⁾などの大量の文献データを扱う情報検索システム AIR³⁾を、Seep の機能を一部用いて現在構築中である。

この Seep を DOS/V 機上の Windows95のもとで動作するように実装を行ったので、それについて報告する。

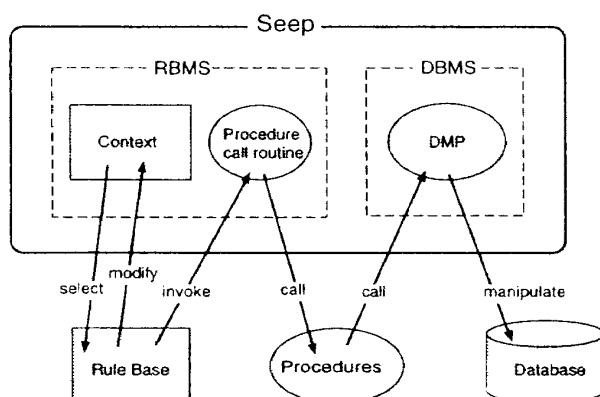


図 1 Seep の構成

2. Seep の構成

Seep は図 1 のように、データベース管理機構とルールベース管理機構によって構成される⁴⁾。ここでは、Seep のデータベース管理機構(DBMS)の構成について説明する。

Seep の DBMS の外部へのインターフェースは、DMP (Data Manipulation Primitives) である。DMP は、データベースの基本操作の集合であり、C 言語で書かれた副手続きである。これらの副手続きを応用プログラムから呼び出すことにより、データベースを操作することができる。DMP の一部を以下に示す。

Open_database … データベースのオープン処理を行う
Close_database … データベースのクローズ処理を行う
Retrieve … 指定された条件を満たすデータの組を求め
Eval … データベースキーからそれを含む行を求める。

Seep のデータベースは、図 2 に示すような関係と関数の集合である。Seep では、関数を関係の特別な場合として区別する。図 2 に示した関係 *JWDIDX*

$JWDIDX \subseteq REC \times JHEAD \times KHEAD \times CONID$

は、四つの領域 *REC*, *JHEAD*, *KHEAD*, *CONID* の直積の部分集合である。これらの領域は関係データベースにおいて属性と呼ばれる。関係のそれぞれの属性には転置索引が作成され、どの属性からも検索ができるようになっている。

一方、関数 *JWDENT*

$JWDENT : REC \rightarrow ENT$

は、*REC* から *ENT* への写像を意味する。関数では写像の値域に対応する属性（この例では *ENT*）に関しては転置索引を作成しない。これが関係との違いとなっている。

Implementation of Seep under Windows

Takeshi Eguchi†, Kensaku Mori‡, Masayuki Takeda†,
Fumihiro Matsuo†

†Graduate School of Information Science and Electrical Engineering, Kyushu University, Hakozaki, Fukuoka 812-81, Japan

‡Faculty of Engineering, Kyushu University, Hakozaki, Fukuoka 812-81, Japan

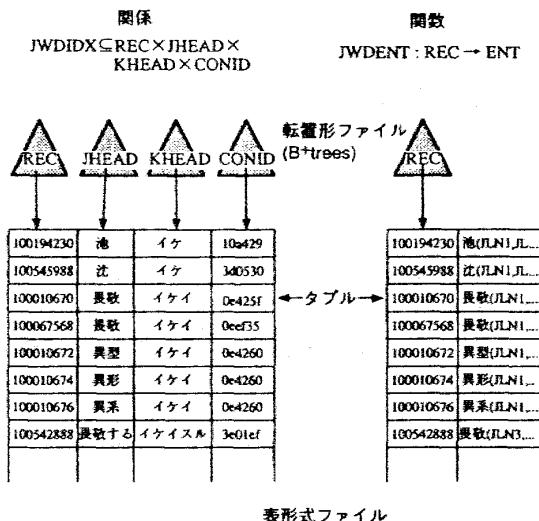


図 2 関係、関数のデータファイル構成

3. データファイルへのアクセス

関係・関数のデータファイルへのアクセスを高速に行うためには、以下の二つを実現する必要がある。

- (1) アクセスする2次記憶のデータブロックのアドレスを計算によって求める。
- (2) 応用プログラムの主記憶領域と2次記憶のデータブロック間のデータ転送をシステムのバッファを介さずに行う。

UNIXやMS-DOSは、標準には上記の機構をもたない。そこで、Seepでは以下に述べる方法によって上記の2項目を実現している。

富士通の汎用機のUNIXであるUXP/Mには、通常のUNIXのファイルシステムであるUFS以外に、富士通が独自に開発したVFLファイルシステム(VFL-FS)がある。VFL-FSでは、データの連続領域割り当て、データの直接転送が可能である。そこで、UXP/MのSeepでは、データファイルをVFL-FS上におくことにし、上記の(1),(2)を達成した。

一般のUNIXに関しては、今のところ(1),(2)を断念し、次の方法を採用している。すなわち、Seepのデータファイルのためにパーティションを一つ用意し、そ

れを1つの特殊ファイルとして扱い、その中に複数の関係・関数のデータファイルを置いている。これによりデータブロックのアドレスを計算によって求めることができる。また、データの転送にはUNIXのシステムコールであるlseek, read, writeを用いている。

MS-DOSは、システムコールにセクタ単位の入出力を行う機能があるので、それを用いてデータの直接転送を行う。また、データファイル創成時に連続領域を割り当て、それに伴ってFAT(File Allocation Table)を書き換えるようにしている。

Windows95でSeepを実装した際には、WindowsのディスクシステムはMS-DOSと同一であるので、アセンブリを用いてシステムコールを呼び出し、データの直接アクセスを行った。

4. むすび

MS-DOSのシステムコールやFATを用いてデータファイルへのアクセスを行う方法を使って、DOS/V機上のWindows95のもとで動作するSeepの実装を行った。

参考文献

- 1) 迫田洋史、竹田正幸、松尾文碩：Seepによる科学ファクトデータベースシステムの構築、電気関係学会九州支部連合大会講演論文集, pp.796(1994)
- 2) 河野重昭、竹田正幸、松尾文碩：結晶構造データベースシステムXDTの改版、九州大学大型計算機センター広報, vol.27, No.3(1985).
- 3) 篠原武、二村祥一、松尾文碩：情報検索システムAIRの改訂について、九州大学大型計算機センター広報, vol.20, No.3, pp.211-238(1987)
- 4) 松尾文碩、竹田正幸、橋本圭介、石津健：ルールベースとデータベースの統合、第45回電気関係学会九州支部連合大会講演論文集, pp.654 (1992).