

化学構造データベースへの関数型アプローチ

4R-5

樂玉琴* 大保信夫** 北川博之** 藤原譲**

(*筑波大学工学研究科、**筑波大学電子・情報工学系)

1. はじめに

化学構造データは部分構造検索、薬物設計など、様々な化学応用で重要である。しかし、汎用関係型データベースシステムにおいては、このような複雑な構造をもったデータが平たい表の形で表現されるので、化学者にとってわかりやすい化学構造の操作ができない。本稿では、関数型データモデルを用いて、化学者の認識に近い形で化学構造データの外部ビューを定義し、それを通して化学構造データベースをアクセスする化学データ管理システムの設計及び構築について述べる。

2. 化学構造データベース

2.1 化学構造の分割

部分構造の包含関係や化学グラフの同型性判定などを行うために、化学構造を部分構造に分割して格納し、アクセスの単位としなければならない。スーパーブロック分割法がそのために開発されている。それは化学構造を環状スーパーブロックとそれに含まれる頂点をすべて取り除いた後に残る鎖状スーパーブロックに分割する方法である。分割されたスーパーブロック間の結合関係を表すために、SC-Tree(Superblock Connection Tree)を用いる。図1にスーパーブロック分割の一例を示す。

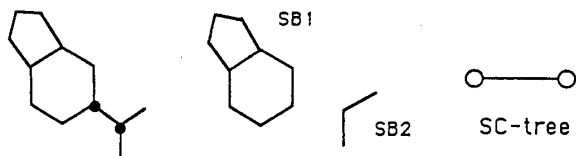


図1 化学構造のスーパーブロック分割

2.2 関係型スキーマ

化学構造をスーパーブロック分割法によって分割し、また分割された部分構造から元の化学構造を復元するために、部分構造における頂点と化学構造における頂点との間の対応をリンク情報として保存す

る。次に化学構造データベースの関係型スキーマの一部を示す。

R1: COMPOUND (C#, NAME,...)
 R2: SUBSTRUCTURE (S#, CHEMICAL-GRAPH,...)
 R3: CONNECTION (CG#, GRAPH,...)
 R4: LINK (L#, VERTEX-MAPS)
 R5: COMPOUND-CONNECTION (C#, CG#)

3. TIME化学データ管理システム

3.1 TIMEシステムアーキテクチャ

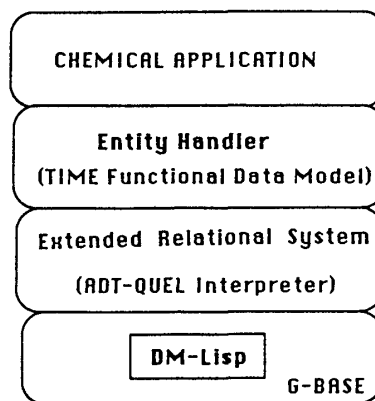


図2 TIME化学データ管理システムアーキテクチャ

TIME化学データ管理システムは、図2に示すように階層構成となっている。一番下にあるのはGBASEと呼ぶ商用関係型DBMSである。GBASEの特徴はDM-Lispと呼ぶLISPベースの言語インタフェースである。DM-Lispは通常のLISP関数にデータベース操作関数を加えた形となっている。

2.2で示したスキーマは定義域上の抽象データ型(ADTs)の使用を仮定している(下線部)。GBASEの上にADT-QUELインタプリタというADTsを扱う能力をもつ拡張関係型システム

Functional Approach to Chemical Structure Databases

Yu-qin Luan*, Nobuo Ohbo**, Hiroyuki Kitagawa** and Yuzuru Fujiwara**

* Program in Engineering Sciences, Univ. of Tsukuba

** Inst. of Information Sciences and Electronics, Univ. of Tsukuba

が開発されている。ADT-QUELインタプリタはユーザにADTsを定義する能力を提供し、またユーザ定義のADT関数を含んだADT-QUEL文をLISPのS式に変換する機能を持っている。

TIME化学データ管理システムはADT-QUELインタプリタの上に更にエンティティ・ハンドラーと呼ぶモジュールを加えることによって、外部ビューの形で、化学構造データを単一の実体として操作することができる。

3.2 外部ビュー

2.2のスキーマにおいて、一つの化学構造データは異なる関係に属する複数の組によって表現されていて、それに対する操作は複数の組にわたっているため、非常に複雑になる。一つの化学構造データを単一の実体として操作するために、関数型データモデルを用いて、2.2のスキーマの上に外部ビューを定義する。外部ビューは複数の実体型の定義から構成されている。COMPOUNDという実体型の定義を例として次に示す。

```
(def-entity-type COMPOUND()
  ((ID          CHAR)
   (SEGMENTATION SC-TREE)
   (STRUCTURE   CHEMICAL-GRAPH)
   (ELEMENT     set-of SUBSTRUCTURE)))
```

ここで、ID、SEGMENTATION、STRUCTURE及びELEMENTは関数で、CHARは基本データ型、SC-TREEとSUBSTRUCTUREは実体型、CHEMICAL-GRAPHは定義域上で定義されたADTである。実体型は属性値の集合を単一の実体として扱うために導入されている。それに付随する関数はデータベース中に存在する属性値間の1対1あるいは1対多の関係を表している。一方、ADTは関係型データベースの定義域上の特定の操作をサポートするために導入されている。

次にSEGMENTATIONという関数の記述を示す。

```
(def-attribute-func SEGMENTATION(i)
  (range of t is COMPOUND-CONNECTION
    retrieve (cvt t.CG#)
    where t.CG# = ID(i)))
```

関数記述の本体はADT-QUELの文である。iはCOMPOUNDというタイプの一つの実体を表している。ID(i)はそのデータベース中におけるサロゲートを返す。t.CG# カラムにはiのSC-Treeのサロゲートを格納しており、cvtはサロゲートから実体へのマッピングであり、従って関数の結果はSC-TREEというタイ

プの一つの実体を返すことになる。関数記述はエンティティ・ハンドラーとADT-QUELインタプリタ間の橋渡しである。

3.3 データ操作

外部ビューを通しての化学構造データベースのアクセスはfor-eachスペシャル・フォームによって行われる。次に問い合わせの一例を示す。

例：サイズが5である環状スーパーブロックを含む化合物の構造を図示せよ。

```
(for-each ((i CYCLIC-SUPER-BLOCKS))
  (where (= 5 (NUM-OF-ATOMS i)))
  (values (mapcar 'display
    (mapcar 'STRUCTURE (PARENT i)))))
```

ここで、NUM-OF-ATOMSとPARENTはそれぞれ環状スーパーブロックのサイズとそれを含む化合物の集合を返す。displayはCHEMICAL-GRAPHというADTの中で定義されている関数で、与えられた化学グラフの2次元図形を副作用として生成する。

4. おわりに

現在化学構造データベースに対する検索しかサポートされていないが、実体の挿入や削除、更に外部ビューの更新問題などが今後の課題である。

参考文献：

- 1) Nakayama, T. and Fujiwara, Y.. Computer Representation of Generic Chemical Structures by an Extended Block-Cutpoint Tree, J.Chem. Inf. Comput. Sci., 23, 1983, pp.80-87.
- 2) Jiang, S.J., Kitagawa, H., Oubo, N., Suzuki, I., Fujiwara, Y.. Abstract Data Types in Graphics Database. Proceedings of IFIP TC-2 Working Conference on Visual Database Systems (to appear), Tokyo, Apr. 1989.
- 3) Yu, X., Ohbo, N., Kitagawa, H., Fujiwara, Y.. Integration of Functional Data Model into Programming Environment: Application to Solid Database, *ibid.*