

2M-8

三重組知識表現を用いた
化学データベースシステムの構築

林 敏浩 山本米雄

(徳島大学工学部)

1.はじめに

化学におけるデータベースの構築は化学への計算機利用の初期から積極的に行われた。CAS¹⁾の文献検索やCIS²⁾の化合物物性値の検索が有名である。近年はRDBを用いた化学データベースの構築³⁾が盛んである。

それに対し、本研究は知識表現の立場から化学データベースの構築を試みる。実際的なアプローチとして三重組を用いる。これはデータを人間の知識に近い形として表現可能であり、人間の思考方法に近い機能、操作を表現できる。

2.三重組⁴⁾

三重組は物(entity),属性(attribute),値(value)の三成分で構成する。三成分をe,a,vで表現すると三重組は(e,a,v)=(物,属性,値)となる。

物,属性,値の三成分に適切な解釈を与えると,三重組は様々な事象を柔軟に表現可能である。

3. TRIAS⁵⁾

本システムはシステム構築にTRIASを用いる。TRIASはパーソナルコンピュータの適応を前提に開発したデータベース管理システムであり,以下の特性を有する。

- 1)三重組データを制御する。
- 2)データはフォーマット・フリーである。
- 3)データの高速操作が可能である。
- 4)三重組データの個数に制限はない。
- 5)データ項目の付加が容易である。

4.システム概略

本システムは化学分子に関するデータ操作を基本として以下の操作を実現する。

- 1)新規化学物質の確認,登録
- 2)化学物質の検索(構造検索,性質による検索,部分検索)
- 3)化学物質の設計(置換,結合,分解)
- 4)化学物質の表示(立体表示,構造式表示)

5.データ表現

5.1.データ分類

化学分子に関するデータは以下に分類できる。

- 1)名称データ 2)性質データ 3)構造データ
- 4)所属データ 5)付属データ

本システムは構造データを化学分子データの基本とする。これは以下の理由による。

- 1)名称は分子構造から一意に決定できる(IUPAC名)。
- 2)性質は構成置換基,結合に影響される。
- 3)所属データは性質に依存する。

5.2.表現原則

本システムは以下の原則に従い,データ構造を決定する。

- (1)化学物質の構造を階層により表現する。
- (2)三重組データは冗長性を許す。
- (3)三重組はプリミティブなデータである。

以下,(1)から(3)について詳述する。

- (1)化学物質を分類して階層性を持たせる。

a)分子 b)中間物(置換基) c)原子

階層性の規則を以下に示す。

- 1)分子を上位,原子を下位レベルとする。
- 2)上位物質は下位物質,同レベルの物質から構成する。
- 3)階層性は上記の3レベルのみで表現する。

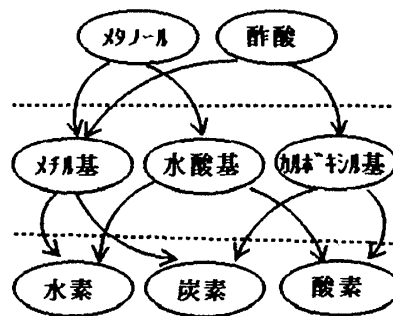


図-1 階層関係

(2)他の三重組を利用して導出可能な三重組データのうち,頻繁に参照するものはデータベースに登録する。

(3)ある三重組データが複数の三重組データに変換可能,かつ後者がより基本的な事象を表現するならば,後者をデータベースに登録可能である。

5.3.構造表現

構造に関する三重組データは以下に分類できる。

- ①物質自身を表現する三重組データ
- ②構成物質の結合関係を表現する三重組データ
- ③物質自身と構成物質の関係を表現する三重組データ
- ④結合の手を表現する三重組データ
- ⑤三次元座標を表現する三重組データ

以下,各レベルごと①から⑤について詳述する。

5.3.1.分子レベルの三重組

- | | |
|-----------|------------------|
| ① E:分子の名称 | <分子名> |
| A:レベル | level |
| 化学式 | chemical_formula |

- 構成物質の総数 construction_total
 結合の総数 connection_total
 V:属性に関する値 <A>
- ② E:結合の名称 <分子名>_connection_<n>
 A:結合対象の物質 construction_<n>
 V:結合の手 hand_<n>
 a)1個の結合の表現三重組は2個存在する。
- ③ E:分子の名称 <分子名>
 A:構成する物質 construction_<n>
 V:構成物質の名称 <構成物質名>
- ④ 三重組データは存在しない。
- ⑤ E:⑤のための名称 <分子名>_construction_<n>
 A: x 座標, y 座標, z 座標 x, y, z
 V:各座標値 <A>
- 5.3.2.置換基レベルの三重組
- ① E:置換基の名称 <置換基名>
 A:分子と同じ属性。以下は分子に存在しない属性。
 結合の手の総数 hand_total
 V:属性に関する値 <A>
- ② E:結合の名称 <置換基名>_connection_<n>
 A:結合対象の物質 construction_<n>
 V:結合の手 hand_<n>
 a) 結合の手を含む結合は三重組を1個だけ有する。
- ③ E:置換基の名称 <置換基名>
 A:構成する物質 construction_<n>
 V:構成物質の名称 <構成物質名>
- ④ E:置換基の名称 <置換基名>
 A:結合の手 hand_<n>
 V:結合 connection_<n>
- ⑤ E:⑤のための名称 <置換基名>_construction_<n>
 A: x 座標, y 座標, z 座標 x, y, z

V:各座標値 <A>

5.3.3.原子レベルの三重組

- ① E:原子の名称 <原子名>
 A:置換基と同じである。構成物質の総数は存在しない。
 V:属性に関する値 <A>
- ②から⑤ 三重組データは存在しない。

5.4.他のデータ表現

他のデータの三重組表現は構造データで示したEに対して適当な属性を付加することにより全て表現可能である。

6.むすび

本稿では特に三重組データ表現について詳述した。三重組表現はユーザ、システム設計者双方に”扱いやすさ”を提供する。また、システム作動中に三重組データを動的に定義可能である。これはRDBにない有効な特長である。

三重組のシンプルさとフレキシビリティは種々のシステムを構築する際に、システム設計の強力な手段となると考える。本システムも他のシステムのデータ部として幅広い利用が可能である。

参考文献

- 1)化学工業の技術開発の情報化に関する調査,日本産業技術振興協会(1985)
- 2)藤原謙:C I S(化学情報システム)の概要:知識システムによる分子設計研究会第18回講演記録,(1981)
- 3)小野修一郎:コンピュータケミストリー:丸善株式会社,東京,(1988)
- 4)山本,柏原,川岸:個人用データベース管理システム I TRIAS,情報管理,Vol.30 No.11,p p1057-1067(1988)
- 5)山本,柏原,川岸,塚本:個人用データベース 構築ツール TRIASの開発,情報処理学会論文誌,Vol.30 No.6,pp734-742(1989)

(炭素,level,原子)	(メチル基,hand_1,connection_4)
(炭素,chemical_formula,C)	(エーテル基,level,置換基)
(炭素,connection_total,4)	(エーテル基,chemical_formula,-O-)
(水素,level,原子)	(エーテル基,construction_total,1)
(水素,chemical_formula,H)	(エーテル基,connection_total,2)
(水素,connection_total,1)	(エーテル基,hand_total,2)
(酸素,level,原子)	(エーテル基_connection_1,construction_1,hand_1)
(酸素,chemical_formula,O)	(エーテル基_connection_2,construction_1,hand_2)
(酸素,connection_total,2)	(エーテル基,construction_1,酸素)
(メチル基,level,置換基)	(エーテル基,hand_1,connection_1)
(メチル基,chemical_formula,CH3-)	(エーテル基,hand_2,connection_2)
(メチル基,construction_total,4)	(ジメチルエーテル,level,分子)
(メチル基,connection_total,4)	(ジメチルエーテル,chemical_formula,CH3OCH3)
(メチル基,hand_total,1)	(ジメチルエーテル,construction_total,3)
(メチル基_connection_1,construction_1,hand_1)	(ジメチルエーテル,connection_total,2)
(メチル基_connection_1,construction_4,hand_4)	(ジメチルエーテル_connection_1,construction_1,hand_1)
(メチル基_connection_2,construction_2,hand_1)	(ジメチルエーテル_connection_1,construction_2,hand_1)
(メチル基_connection_2,construction_4,hand_1)	(ジメチルエーテル_connection_2,construction_1,hand_2)
(メチル基_connection_3,construction_3,hand_1)	(ジメチルエーテル_connection_2,construction_3,hand_1)
(メチル基_connection_3,construction_4,hand_2)	(ジメチルエーテル,construction_1,エーテル基)
(メチル基_connection_4,construction_4,hand_3)	(ジメチルエーテル,construction_2,メチル基)
(メチル基,construction_1,水素)	(ジメチルエーテル,construction_3,メチル基)
(メチル基,construction_2,水素)	
(メチル基,construction_3,水素)	
(メチル基,construction_4,炭素)	

図-2 ジメチルエーテルの三重組表現