

拡張関係データベースADAMの類似検索機能の実現

2Q-5

宇田川 佳久 · 辻 秀一
(三菱電機機・情報電子研究所)

1. はじめに

定型事務処理をはじめとする従来のデータベース応用では、レコード(タプル)に意味があることが多かった。他方、CAD、OAやエキスパート・システムなどでは、一連のレコードの集まりが意味のある情報単位(オブジェクト)となることが多い[1]。ここで、複雑なレコードの構造を持つオブジェクトをどのように検索するか、ということが問題となる。従来のデータベースのように、属性値を直接指定する方法が適当であるとは限らないように思える。

本文では、拡張関係データベースADAM[2]のデータ・ディレクトリ(DD/D)を使って、利用者が指定したオブジェクト(回路図)と類似したオブジェクトを検索する手法と実現結果について述べる。

2. 類似検索の手法

2.1 回路図の階層とデータ・ディレクトリ

回路図は、一般に、階層を成す。例えば、10進リッパルカウンタは4個のJ-Kフリップ・フロップと1個の2入力NANDゲートによって作られている(図1)。

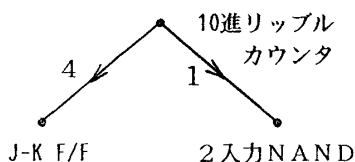


図1. 回路図の階層の例

ADAMのデータ・ディレクトリは、回路図名、バージョン名、生成時刻などを管理するリレーションと、回路図の階層情報を管理するリレーションとから構成されている。階層情報を管理するリレーションは

(回路, 構成回路, 引用個数)

という属性で構成されている。一方、回路図そのものは

- (1) COMP(DIV-ID, DIV)
- (2) TERM(T-ID, DIV-ID, CHR, XC, YC)
- (3) CONN(L-ID, S-ID, DIV-S, D-ID, DIV-D)

というリレーションによって表現されている。

上記の回路図の階層情報のうち、構成回路と引用個数は、以下のデータ操作によって得ることができる。

```
SELECT  DIV, COUNT(DIV)
FROM    COMP
WHERE   DIV-ID≠'D0'
GROUP  BY DIV ;
```

また、回路名は、

```
SELECT  DIV
FROM    COMP
WHERE   DIV-ID='D0' ;
```

によって得ることができる。回路図の階層情報は、回路図を登録するときにデータ・ディレクトリに記憶される。

2.2 回路図の類似検索手法

回路図の類似検索は、次のステップから成る。

- (1) 回路図エディタEVEで、利用者の意中にある回路図(ルアー回路図)を書く。
- (2) 類似検索システムAPPLE(A Prime Partner for Leading Experts)を呼び出す。
- (3) 類似度、相違度を参考にしながら、適切な回路図を選択し、画面で図面として確認する。
- (4) 最適な回路図を決定し、EVEにもどる。

以下では、(2)の類似検索機能について述べる。この機能は、以下の処理から成る。

- (2.1) (1)で与えられたルアー回路図の要素回路を取り出す。
- (2.2) それぞれの要素回路に対し、その要素回路を引用している回路をデータ・ディレクトリに記憶されている階層情報を使って見つける(類似回路候補)。
- (2.3) ルアー回路図と類似回路候補との類似度、相違度を計算する。

ADAMでは、回路図をリレーションによって表現しているため、リレーションの差を類似度、相違度の基準とすることができる。図2は、その原理を示したものである。

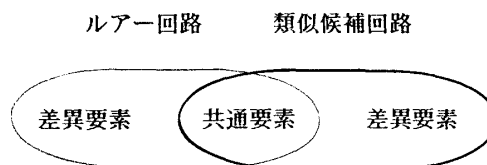


図2. 類似度、相違度の計算原理

Implementation of Similarity Retrieval Function of Extended Relational Data Base ADAM

Yosihisa UDAGAWA and Hidekazu TSUJI

MITSUBISHI ELECTRIC Corp.

単純な集合であれば、共通要素、差異要素の数が、それぞれ類似度、相違度の基準となるが、2.1で述べたように、同じ要素が複数引用されることもあるので、類似度、相違度の算出には若干の工夫が必要となる。

ここでは、共通要素の引用個数の差の合計を類似度の基準とする計算法について述べる。すなわち、共通要素の数をN、ルアー回路と類似候補回路の共通要素iの引用個数を、それぞれ $C(x_i)$ 、 $C(y_i)$ とするとき、

$$\text{類似度の基準} = \sum_{i=1}^N |C(x_i) - C(y_i)|$$

となる。同様に、相違度の基準は、

$$\text{相違度の基準} = \sum_{i=1}^M D(x_i) + \sum_{i=1}^N D(y_i)$$

ただし、 $D(x_i)$ 、 $D(y_i)$ は、それぞれ、ルアー回路と類似候補回路の相違要素iの引用個数、M、Nは相違要素の数である。

3. 類似検索機能の実現

2.2で述べた類似検索機能を実現するのにFORTRANで約2000行を要した。ADAMのデータ操作機能、図形表示機能、データ・ディレクトリ機能を駆使した結果、高いモジュラリティと高速な処理を達成することができた。主要なモジュールとステップ数は、以下のとおり。

- * 類似候補名を検索する 約 250行
- * 類似度、相違度を計算する 約 300行
- * 類似候補をソートする 約 200行
- * 類似候補名を表示する 約 50行
- * 類似候補回路図を表示する 約 300行

- * コマンド・メニューを管理する 約 200行
- * 表示画面を管理する 約 700行

図3は、1個の2入力NANDゲートをルアー回路図として類似検索をした結果の表示画面である。類似検索システムAPPLEを呼び出すと、右端のコマンド・メニューと類似回路候補の類似度、相違度と候補の名前が表示される。利用者は、類似度、相違度を参考にしながら類似回路候補を選択し、画面に表示し、図面の内容を確認することができる。図面の詳細度は、必要に応じてコントロールすることができる。

4. おわりに

複雑な構造を持つデータを検索するためには、類似検索機能が重要な役割をする。本文では、拡張関係データベースADAMの類似検索機能について述べた。主な特徴は以下のとおり。

- (1) 利用者はデータの構造を知る必要がない。単に、欲しいオブジェクト（回路図）の特徴を指定するだけでよい。
 - (2) オブジェクト同士の類似度、相違度は類似モデルに基づいて計算される。従って、類似度を個々のオブジェクトに対して定義する必要はない。
 - (3) オブジェクトの階層構造を利用しているため、高速かつ検索漏れがない。
- 今後は、高度な類似モデルを開発する予定である。

参考文献

- [1] Peterson, W. R: Object-oriented DB Design, AI Expert (March, 1987).
- [2] 宇田川, 他: 拡張関係データベースADAMのデータ・ディレクトリの実現, 情報処理学会第36回全国大会, 1E-2 (1988).

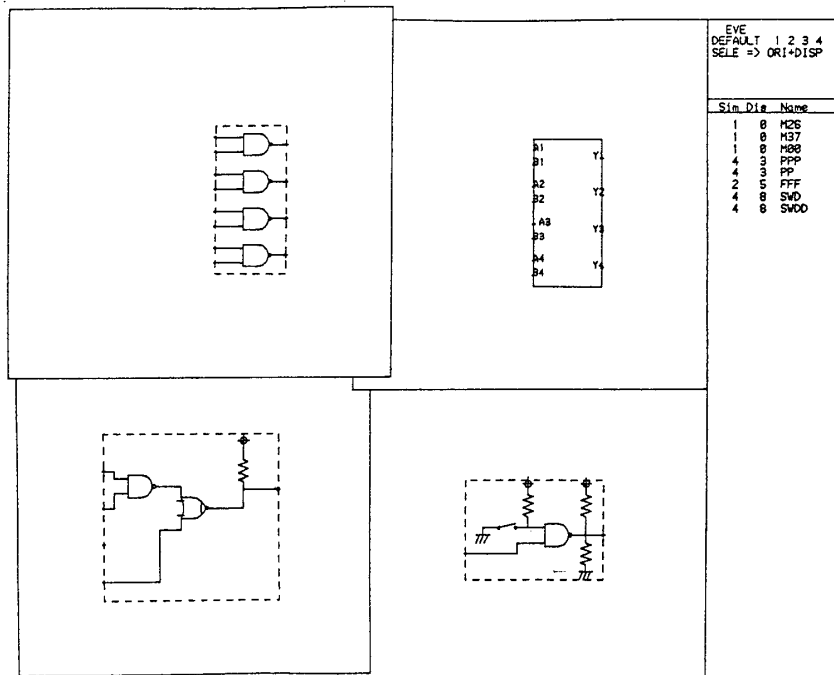


図3. 類似回路図検索の画面例