

プログラム自動生成システムPGENにおけるデータ関連図の意味理解

土田 賢省, 西谷 泰昭, 岩元 晃二
日本電気 株式会社 生産技術研究所

1. はじめに

自然言語や図表は人間と親和性の高い表現手段であり、ソフトウェア開発においても仕様書等のドキュメントは日本文、図、表の組み合わせで記述されることが多い。ソフトウェアの生産ツールをより人間に近づけるためには、これらの人間にとってわかりやすい表現である自然言語、図、表などを計算機が理解することが重要である。

我々は、日本文、図、表で書かれたプログラム仕様書を理解し、プログラムを自動的に生成する方式について研究しており、その実験システムPGEN-1の概要については、既に報告済みである[2]。プログラム仕様書の中で図は、データの関連を表わすものがほとんどであり、日本文と合わせて理解しなければならない。PGENにおいては、ユーザは図の意味情報など考えず書かれたままのイメージで図を入力し、図形解析でそれを図を参照している日本文と合わせて理解しプログラムに変換する方式をとっている。本稿では、その方式について、図の意味解析を中心に説明する。

2. 仕様記述における図形

実際のプログラム仕様書に出現する図のほとんどが図1に示されるような、データおよびデータの関連を表わすものである。従ってここでは、処理対象の図形を以下の3つの構成要素からなる図形として定義した。

(1) エリア図・・・データの構造を表わす矩形図で、中は水平・垂直線分で分割され、そこにテキストが置かれる。分割された最小の矩形領域(中に水平・垂直線分を含まない)をボックスと呼ぶ。

(2) テキストユニット・・・エリア図の外にある枠で囲まれていない文字列(エリア図のボックスに置かれたテキストとは区別される)。

(3) 接続線・・・エリア図のボックスやテキストの間に関係があることを示す線分(有向/無向)。

プログラム仕様書においてこのような図形は、代入や条件を表している。例えば、図1を「明細エリアに図1のように値をセットする。」というような日本文

で参照するときは、初期設定における代入の関係を表わしていると解釈する。

3. 図形解析

図2に図形処理部のシステム構成を示す。図形意味処理部はPGEN全体の意味処理部からコールされ、図形を参照している日本文の意味を表わすMネットがインクフェースとして読まれる。図形意味処理部では、図の入力に対応する構造を表現したGネット(図形構造解析の出力)とあらかじめ定義された変数辞書を参照しながら、図の意味表現であるMネットを生成する。その際、図の中の線分などで表わされる関係の種類がeq(等号)、set(代入)のいずれかを決定するが、それは読まれた日本文のMネットが条件を示すものであればeq、実行部を示すものであればsetと判定する。このようにして生成されたMネットは、PGEN全体の意味処理部により一つのMネットに統合され、PGENのプログラム生成部は、統合されたMネットを解釈してプログラムに変換する。以下、各部分について説明する。

(1) 図形入力部

図1のような図形をグラフィックディスプレイで会話的に入力するシステムである。特徴として、次の2点がある。

- ・図形の書き順の自由度が高い。
- ・定義に含まない図形は書けないように適切な制限をもうけている。

これらの特徴は図形の規則に対応した柔軟なデータ構造で図形を管理することにより達成される。特にエリア図については、水平線や垂直線に対応したノードをもつ特殊な木構造データで管理されており、入力・編集操作に対してリアルタイム(水平・垂直線分の数に対して線形時間)でデータの更新がなされる。図1の下方のエリア図に対する内部データを図3に示す。図中、tはテキスト情報、oは水平又は垂直線分の座標値を表わす。例えば、垂直線削除という操作に対しては、それに対応するノードを削除すればよい。

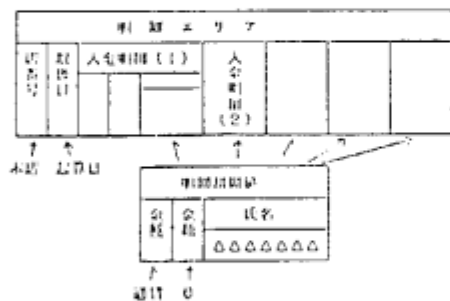


図1. 明細エリア

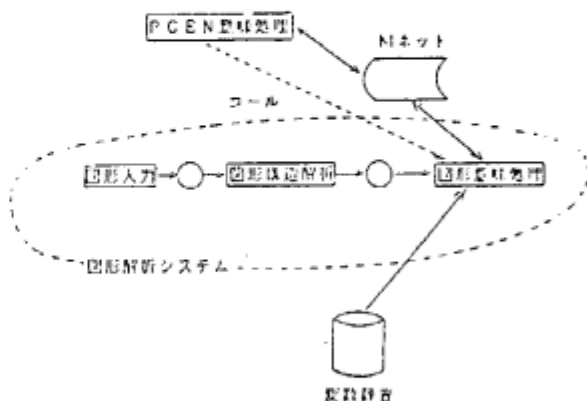


図2. システム構成

Semantic Analysis of Data Relation Diagrams
in the PGEN Automatic Program Generation System
Kensei TSUCHIDA, Yasuaki NISHITANI, Kanji IWAMOTO
NEC Corporation

(2) 図形構造解析部

入力された図形が表現しているのは、エリア図、テキストユニット、接続線それぞれの座標値とエリア図の構造である。それらから以下のような意味処理のための情報を抽出するのが図形構造解析部である。

- ・エリア図の構造（エリア図のボックスの親子関係）
- ・接続線により関係付けられている対象（エリア図のボックスやテキストユニット）の組（意味的には正しくないこともある。これは、図形意味処理部で修正される。）

例として図4に図1の構造解析の結果（Gネット）を示す。

(3) 図形意味処理部

図形意味処理部は、次のような問題を扱う。(a) エリア図がどの変数を表わすか。(b) エリア図のボックス内やテキストユニットに書かれているテキストはどのように使われるのか（定数や変数を表わすのかあるいは単なる注釈に過ぎないのかなど）。(c) 接続線の意味は何か（代入や等式を表わすのかあるいは注釈のためのものかなど）。(d) 接続線はどこどこを結んでいると解釈するのが妥当なのか。図形意味処理部では、以上の意味を解析し、代入（あるいは等式）の関係が成り立つ可能性のある組（変数と変数あるいは変数と定数のペア）を抽出し、日本語の意味表現と照合し、図形の意味表現であるMネットを生成する。以下に図形意味処理部で用いられる規則のうちおもなものを挙げる。

(a) エリア図に対応する変数の同定ルール

エリア図のボックスに書かれた変数の日本語名などを手掛かりとして、エリア図の最上位のボックスに対応する変数の候補を選んだとき、エリア図の表わす構造とその変数の変数辞書内に記述されているデータ構造が対応し（但し、エリア図はその変数の表わすデータ構造に対して省略した形であってもよい）、エリア図の各ボックスに書かれている日本語名がそれぞれ対応する変数名と矛盾しないとき、エリア図の各ボックスをそれぞれ対応する変数と同一視する。

(b) エリア図だけからの代入（あるいは等式）関係抽出ルール

エリア図のボックスとそこに書かれたテキストの関係（代入、等式）を抽出する。テキストの表わすものが定数のとき、その定数のデータ型とそのボックスに

対応する変数のデータ型が代入（あるいは等式）に関して矛盾がなければ、それらの間に代入（あるいは等式）の関係が成り立つ可能性があるとする。

(c) 接続線による代入（あるいは等式）関係抽出のルール（エリア図とテキストユニットの場合）

接続線で結ばれているテキストユニットとボックス及びその上位ボックス（Gネット上で上位に位置する）の関係（代入、等式）を抽出する。テキストユニットの内容がどのボックスの表わす変数（規則(a)により同一視されたもの）の日本語名や型名でもなく、かつ定数または変数を表わす（注釈等でない）とき、テキストユニットの内容がそれらボックスの表わす変数と代入（あるいは等式）に関してデータ型等に矛盾がなければ、それらの間には代入（あるいは等式）の関係が成り立つ可能性があるとして抽出する。但し、複数個の可能性があるときは、最も下位のボックスとの関係を優先する。エリア図間の関係についても同様に同定する。

このような規則によって抽出された組がe q（等号）、set（代入）のいずれの関係を表わすかは、日本語の意味表現を参照し、条件部で図を参照しているか、実行部で参照しているかによって決定する。

4. おわりに

このように、プログラム仕様書の図形を理解し、それをもとに日本語解析部等と合わせてプログラムを自動的に生成することができた。ここで定義した図形はデータを表わす図形として一般的なものであり、例えば図形によるデータ宣言などということも可能である。さらにテキストユニットにいろいろな日本語が書けるよう拡張することにより、図形の表現能力を高めることができる。これらは、今後の課題である。

文献

- [1] 西谷、和田、岩元「知識利用のプログラム自動生成システム」昭60情報学会全大第30回
- [2] 岩元、西谷、和田、庄司、土田「PGEN-1：仕様書理解をベースとしたプログラム自動生成システム」昭61情報学会全大第32回

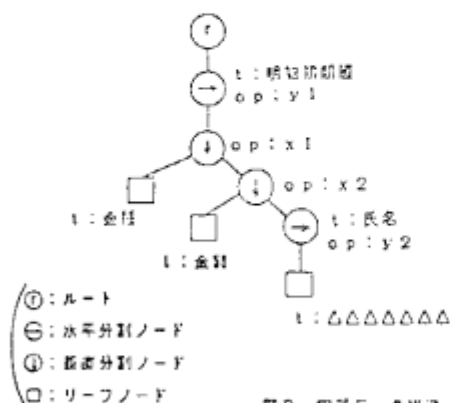


図3. 内部データ構造

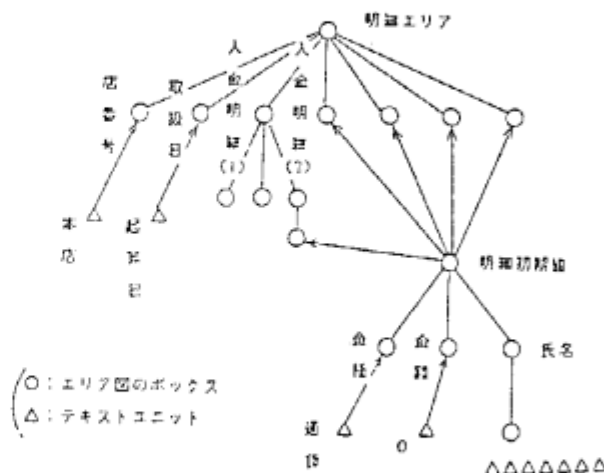


図4. Gネットの例