

日本語構文に基づく用語間関係情報の自動抽出法

4L-9

黒木 宏明 磯田 定宏  
(NTT ソフトウェア研究所)

1 はじめに

一般にソフトウェアの設計書は日本語文章を用いて記述される。したがって、日本語文章から用語間の関係を自動的に抽出することができれば、設計書に関するインタビューや矛盾等のチェックを効率化できる可能性がある。本稿では、日本語設計書を解析し、用語間の関係情報を自動抽出する方式について述べる。

2 関係情報の分類

(1) 関係情報

日本語文章から抽出する関係情報を次の2種類に分類する。

①階層関係

ドキュメントには章節等あるいはインデクションで規定される構造が存在する。これを階層関係としてとらえる。

②用語関係

日本語文章に現われる名詞、動詞、形容詞類を用語とする。用語関係は日本語文章から抽出する情報であり、11種類(定義、入力、出力、処理、生成、状態、同値、包含、参照、型、属性)に分類する。

(2) 関係要素

用語関係を構成する各用語はその抽出された関係に關する意味的な役割(関係要素)を有する。代表的な関係要素を表1に示す。

例えば、入力関係の関係要素には以下の4つがある。

- ①入力主体 … 入力処理を実現する機能名
- ②入力対象 … 入力処理の対象(装置名)
- ③入力要素 … 入力データ
- ④格納対象 … 入力データの格納場所

関係要素の関係を図1に示す。

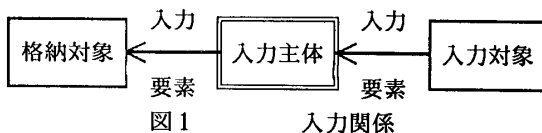


表1 関係とその要素一覧

関係名	関係要素	対応格	意味
定義関係	定義対象	主格	定義名
	定義内容	対象格	定義名の意味・内容を詳細に説明した文章
入力関係	入力主体	主格	入力処理を実現する機能名
	入力対象	始点格	入力処理の対象(装置名)
	入力要素	対象格	入力データ
	格納対象	終点格	入力データの格納場所
出力関係	出力主体	主格	出力処理を実現する機能名
	出力要素	対象格	出力データ
	出力対象	終点格	出力処理の対象(装置名)
	格納対象	始点格	出力データの格納場所
参照関係	参照主体	主格	参照を行なう主体者
	参照対象	対象格	参照される対象
同値関係	同値対象	主格	同値名
	同値内容	対象格	同値名の意味・内容を示した文章
包含関係	包含対象	主格	包含する対象
	包含要素	対象格	包含される要素

3 関係情報の抽出

(1) 日本語構文解析法

設計文章から関係情報を抽出するため、構文木を作成する。構文木の作成手順を以下に示す。

【step1】単語の切り出し

日本語文章を形態素解析し、文章を構成する単語の情報(品詞、活用形、用法等)を決定する。

【step2】構文解析

日本語構文規則に従い、係受け解析を行い構文木を作成する。

【step3】用語の意味づけ

動詞および動詞に係る助詞を手がかりに、各用語に対して文の中で果たす役割(格)を付加し、用語に対して意味づけを行なう。

(2) 関係情報抽出法

日本語構文解析により得られた構文木から用語間の関係情報を抽出する。

構文木からの関係情報の抽出手順を以下に示す。

【step1】関係づけられる用語の抽出

構文木から動詞に係る用語を抽出する。

【step2】用語間の関係抽出

構文木から用語間の関係を抽出するため、動詞がゴリ毎に抽出ルールを設定する。

抽出ルールは以下の形式で記述する。

【入力動詞の場合】

入力動詞 ( 主格: 名詞1, 対象格: 名詞2,  
始点格: 名詞3, 終点格: 名詞4)

⇒入力関係(名詞1<入力主体>, 名詞2<入力要素>,  
名詞3<入力対象>, 名詞4<格納対象>)

動詞に係る用語に対して関係抽出ルールとのマッチングを  
とり、用語の関係要素を決定する。

(3) システム構成

設計ドキュメントから用語間の関係を自動抽出するシステム構成  
を図2に示す。

本システムは、日本語文章から構文木を抽出する日本語解  
析、日本語構文木から関係情報を抽出する関係解析、用  
語間の関係情報を格納する関係情報DBおよび動詞がゴ  
リ毎に作成した関係抽出ルールより構成される。

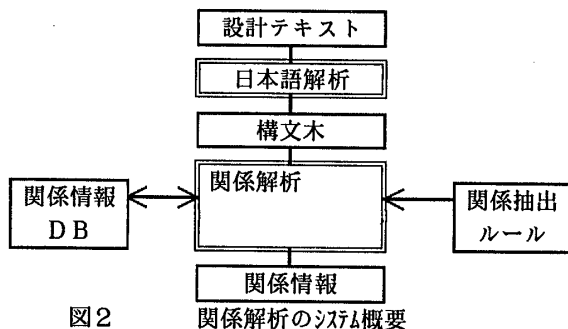


図2

4 関係情報に基づく設計支援機能

用語間の関係情報および階層関係を用いることにより、  
データフロー図作成、データ構造図作成、設計仕様の無矛盾性チ  
ェック、完全性チェック等が行える。

(1) データフロー図の作成

入力関係および出力関係等を用いてデータフロー図を作成す  
る。

(2) データ構造図の作成

包含関係を用いてデータ構造図を作成する。

(3) 無矛盾性チェック

設計文章中の用語および用語間の関係に矛盾がないこ  
とをチェックする。

(4) 完全性チェック

設計文章中の用語がすべて定義されていることをチェック  
する。

図3に日本語文章からデータフロー図およびデータ構造図を作  
成する例を示す。

5 おわりに

本稿で述べた関係情報の自動抽出法は、ソフトウェアの設計  
時のみでなく保守時にも有効である。今後、本方式に基  
づく設計支援および保守支援機能を試作し評価する予定  
である。

参考文献

- [1]黒木他「ソフトウェア関係情報に基づく再利用方式」  
情処第31回全国大会
- [2]黒木他「設計情報ネットワークに基づくソフトウェア設計法」  
情処第34回全国大会

設計仕様

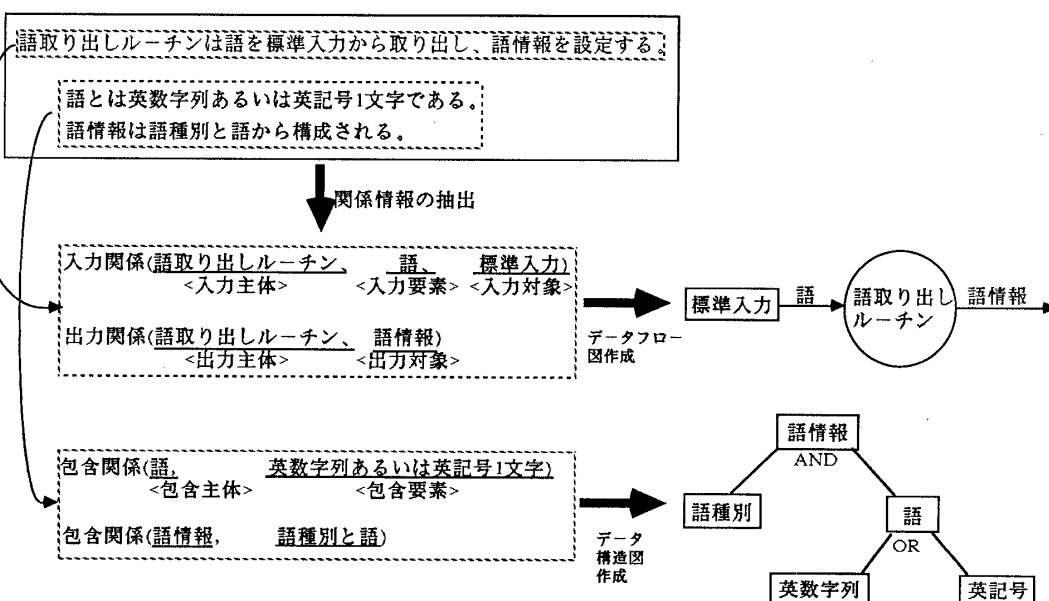


図3 データフロー図およびデータ構造図作成例