

単語学習とユーザモデル形成を行う ソフトウェア部品検索ツールについて

小高 知宏 内山 明彦

早稲田大学理工学部

ソフトウェア開発における生産性・信頼性の向上を目的とした方法論の一つに、ソフトウェアの部品化・再利用がある。この際、必要な部品を検索し、ユーザに提示する手法として制限された自然言語による方法がある。今回提案するシステムは、ユーザの入力から構文情報等を手掛りに意味を検出し、対話的に部品を検索するものである。また部品の保存の際に、検索に利用するために部品に付加する情報をユーザから対話的に取り出す機能も持つ。加えて、検索に用いる自然言語の制限をなるべく減らしユーザインタフェースを向上させるために、システムは新しい単語を学習する。以上のような機能を持つ部品検索・登録システムの構成について述べた。

The natural-language query system for reuse of software components

Tomohiro ODAKA, Akihiko UCHIYAMA

Waseda University, School of Science and Engineering

3-4-1, Ohkubo, Shinjyuku, Tokyo, 160 Japan

This article explains the features of the natural-language query system that supports reuse of software components. The system constructs the user models, and learns new words contained in input texts. First, the paper explains the feature of Alice system that supports reuse of software components, then explains the natural-language query system that is a part of Alice system. The query system is used in search for software components and registration of software components to database.

1. まえがき

ソフトウェア開発における生産性・信頼性の向上を目的としてこれまでに多くの方法論や手法が提案されているが、これらのうち近年特に注目されているものにソフトウェアの部品化・再利用がある。これは既存ソフトウェアを機能単位に分解して再利用可能部分を抽出・保存し、新たにソフトウェアを開発する際にこれを用いようとするものである。ソフトウェア部品（以下部品と呼ぶ）の再利用によりソフトウェアの製造工程を短縮し、かつ用意した部品の品質を保証することでソフトウェア製品の品質を向上させることが目的であり、従来より部品化・再利用の実現のために必要となる部品の保存や検索、部品結合の手法等について研究が行われている。^[1]

筆者らは、プログラムの部品化・再利用を支援する開発支援環境“ A little interactive computing environment (A l i c e)”の構築を行っている。現在までに、プログラムの部品化を効率良く行うためのプログラム言語

“multiple Computer Language for Development (m C L D)”、言語m C L Dのための設計ツール“Language ahead of CLD (L C)”、テストツール“Hyper Debugger (H D)”、及びエディタ“looking-glass”、データベースマネジメントシステム“wonderland”の設計・試作を行った。

今回はA l i c eにおけるプログラム部品の検索手法について報告する。部品化・再利用支援システムでは既存部品の検索は、出来るだけ部品検索を行うユーザに負担をかけずに行えなければならない。その方法として従来の部品化・再利用支援システムでは、述語論理等によるもの、例によるもの、制限された自然言語によるもの等が用いられている。今回提案するシステムでは、制限された自然言語による検索方式を採用する。これは、

対話的に部品検索者（ユーザ）から文を入力してもらい、入力された文についてその構文情報等を手掛りに意味構造を生成し、これを部品データベース内の項目と比較・検索することで部品を検索するものである。なお部品データベース内で部品の意味を表現する部分は、部品をデータベースに登録する際にあらかじめ部品毎に与えてある。また部品保存の際に、後で検索に利用するために部品に付加するキーワード等の情報をユーザから対話的に取り出す機能も持つ。加えて、検索に用いる自然言語の制限をなるべく減らしてユーザインタフェースを向上させるために、制限された範囲でシステムが学習を行う。また単語・構文辞書を個人別に持つことで、これをユーザモデルとして利用できるようにする。以下ではこのシステムについて、主として構成方法に関して述べる。

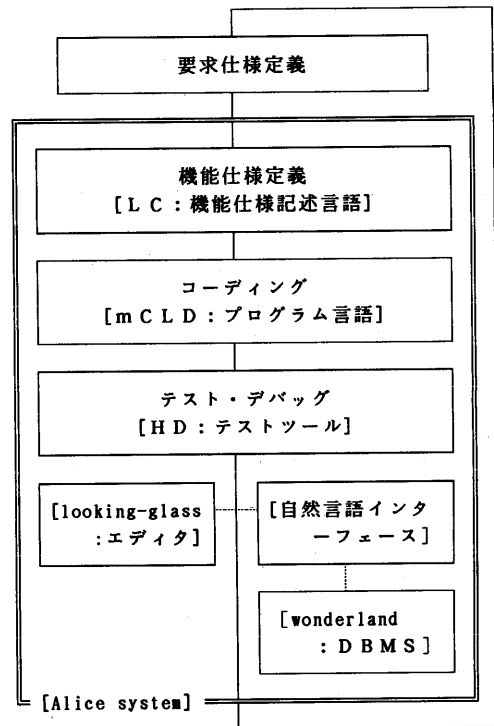


図1 開発の各段階とA l i c eのツール

2. 開発支援環境 Alice と部品検索

プログラム開発支援環境 Alice の概略を図 1 に示す。図 1 ではソフトウェア開発の各段階に対応させて Alice のツールを示しているが、今回述べる部品検索部分は部品管理用データベース管理システム（以下 DBMS と呼ぶ）のフロントエンドプロセッサとなるものである。現在 Alice の DBMS は、あらかじめデータベース登録時に登録者によって与えられた各部品のキーワードをもとにして、部品検索を行うユーザの指示により適当な部品を検索する機能を持つ^[2]。この部品を内部（機能）仕様書に記述することでソフトウェア作成を行う^{[5][6]}。今回はこれをより使いやすくするための工夫として、自然言語処理機能を導入したインタフェースを設計した。Alice においては、部品の検索は対話的に行う。従って自然言語形式の記述から部品検索を行う場合でも、Alice に組込むことを考慮すると、対話的な検索を行うシステムが必要になる。機能としては、部品の検索及び部品データベースへの登録が出来ればよい。

自然言語インタフェースを用いることの利点は、部品検索システムの操作性が向上することにある。また、別の利点として、部品登録者と部品検索者が一致しない環境では、単純なキーワード方式と比較して部品検索が容易になることが上げられる。後で述べるように本システムでは、検索を行うユーザ毎に単語辞書を切り換えることで、ユーザの用いる用語と部品データベースで用いる用語との対応関係をユーザ毎に変更することができるようシステムを構成する。このように構成すると、ユーザの質問文に出現する単語や構文と、部品データベースのサポートするものとの差異を、ユーザ毎に設定した辞書で吸収することが出来る。また部品登録時に同じ自然言語インタフ

ェースを用いることで、登録者の用いる用語と部品データベースで用いることができる（図 2）。このようにすると、検索や登録に用いることができる用語の自由度が増すので、部品検索は容易になる。

3. 検索システムの構成

次に自然言語インタフェースシステムの構成を示す。システムは部品登録部と部品検索部に分かれる（図 3）。登録部は、検索部の一部を利用して構成する。なお、本システムでは入力として分かち書きローマ字文を仮定する。従ってかな漢字混じり文の処理には、別に日本語形態素解析フロントエンドプロセッサを用

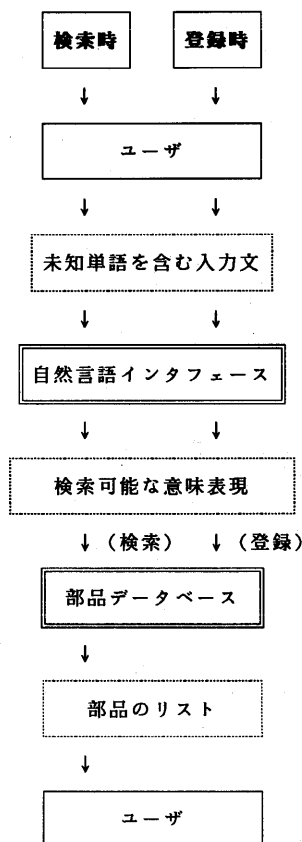


図 2 自然言語インタフェースの機能

意する必要がある。

3. 1 解析部の構成

はじめに、検索部の構成について述べる。検索部は図4に示すように3つの部分に分かれる。解析部は、ユーザの入力した文を構文を手掛りに解析し、検索の

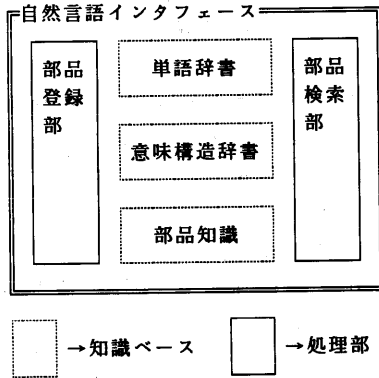


図3 自然言語インタフェースの内部構造

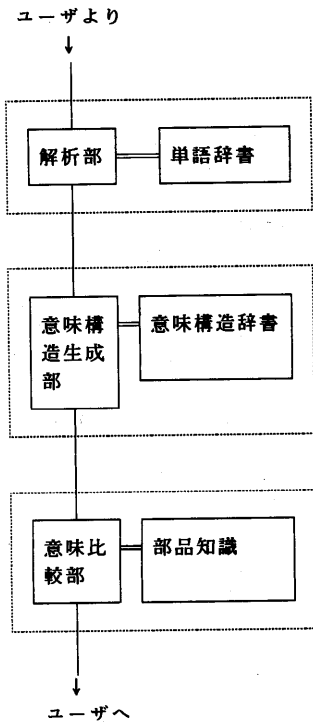


図4 検索部の構成

鍵となる単語や構文情報を抽出する^[7]

(図5)。意味構造生成部は前段から受け取った構文情報をもとに、検索に必要な意味構造を生成する(図6)。意味比較部は、あらかじめ部品データベースに蓄えておいた部品に関する情報と、前段までの解析で入力文より抽出した意味構造とを比較し、ユーザの要求する部品を検索する。

解析部は図7に示すように、構文解析部を中心に構成する。図で「単語辞書」と示した部分は知識ベースであり、それ以外の字句解析部、構文解析部、構文情報抽出部、キーワード生成部が処理部分である。字句解析、構文解析部は入力文から検索に必要な単語の情報を抽出する。構文解析は文脈自由文法に基づくものであるが、解析能力に限られる単純なものであるが、処理が高速で拡張が容易である^[3]。一般に対話的な自然言語インタ

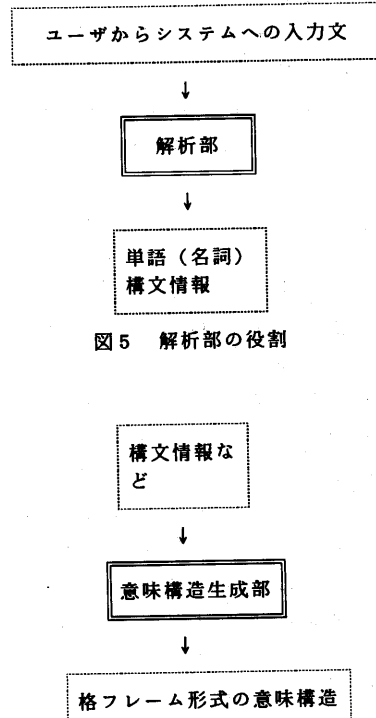


図5 解析部の役割

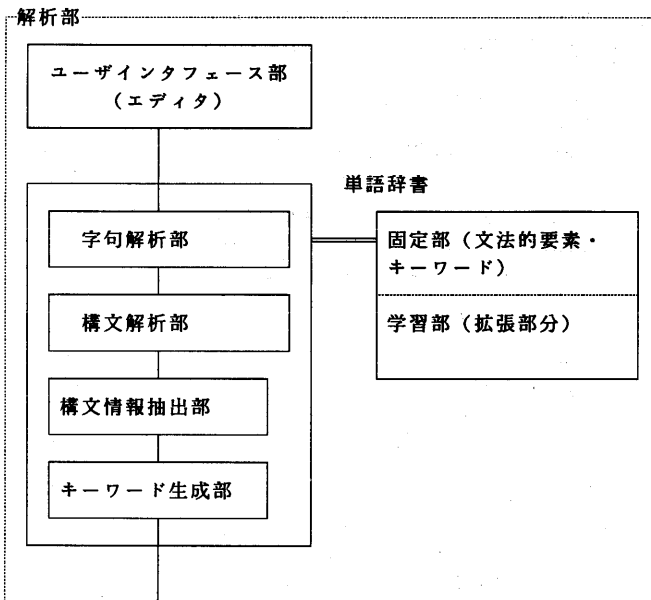
図6 意味構造生成部

フェースでは、高速性が重要な要素になる^[8]。このため、構文解析部はUNIXのコンパイラジェネレータであるyaccを用いてプロトタイプを作成した。yaccは本来プログラミング言語のコンパイラを作成する場合に用いるツールであるが、機能上プログラミング言語に限らず文脈自由文法で規程される言語一

般に対して適用が可能である。解析部を構成するyaccの記述の一部を図8に示す。

構文情報抽出部及びキーワード生成部では、次段の意味構造生成部に渡す情報として、キーワードとして抽出した単語や構文情報を生成する。

解析部では、上のいずれの解析にも単語辞書を用いる。単語辞書には、助詞等の文法的な要素と、部品検索に用いる動詞、名詞が混在している。単語辞書の大局的な構成を図9に示す。図9において固定部には文法的要素と、部品データベ



次段へ
図7 解析部の構成

```

%token <id> NOUN
%token <id> ADJ
%token <id> JYOSHI
%token <id> VERB
%token <id> ENDOFCONV
%token <res> BLANK

%type <id> ssp
%type <id> sp
%type <id> vp
%%
statement      : sp vp
                {
                    printf("arisa : find statement\n");
                    YYACCEPT ;
                }
                | vp
                {
                    printf("arisa : find statement\n");
                    YYACCEPT ;
                }
                | sp
                {
                    printf("arisa : find statement\n");
                    YYACCEPT ;
                }
                | ENDOFCONV
                {
                    printf("arisa:end-of-conversation\n");
                    end_of_processing_flag = TRUE ;
                    YYACCEPT ;
                }
                | error
                {
                    printf("arisa :syntax error\n");
                    YYACCEPT ;
                }
;
sp              : ssp {
  
```

図8 構文解析部の実際 (一部)

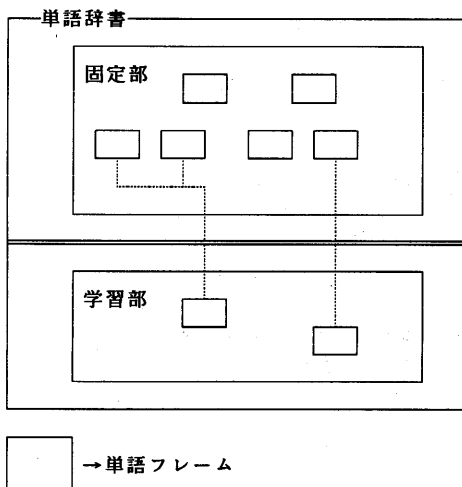


図9 単語辞書の構成

ースで検索のキーワードとして用いることのできる単語が格納してある。学習部には、ユーザからの入力文中に出現した未知の名詞が格納される。単語を学習して新たに単語辞書に追加する際には、はじめに部品データベースで使用可能な単語との関係をユーザに問い合わせる。次に登録する単語と、問い合せた結果である検索のキーワードとして用いることのできる単語との関係を両方とも単語辞書の学習部に登録する。単語辞書の内部構造は、図10に示すようなフレーム形式である。フレームのスロットは単語、品詞の分類、キーワードとなるかどうかの別、他の単語との関係、意味上の分類等からなっている。他の単語との関係が一对多の場合、関係を直接的な連鎖とする方法と、ファジィ的に結合させる方法が考えられる。前者は図11に示すように、一つの単語が幾つかのキーワードと直接に結び付けられるものである。後者は一つの単語がどの単語に結び付けられるかを0から1の実数値で記述したもので、いかえれば新しく追加する単語に、データベース検索に用いることのできるキーワードへの所属度^[9]を与えることで関係を記述するものである。現在のシステムでは直接的な結合のみをサポートしているが、本来単語同士の関係は一对一に決定することは無いので、関係の扱いについては後者の採用が望ましい。

なお単語辞書は、これに対する操作とは独立した形式でのデータベースとなっているため、ユーザ毎に切換えて利用することができる。また後で述べるように同一の単語辞書を部品データベースへの登録にも利用する。これにより部品データベースで用いる表現を規格化すると同時に、各個人の使う自然言語の表現への制限をできるだけ少なくすることができる。

また、ユーザ別に作成した辞書はユーザの知識範囲を示していると考えること

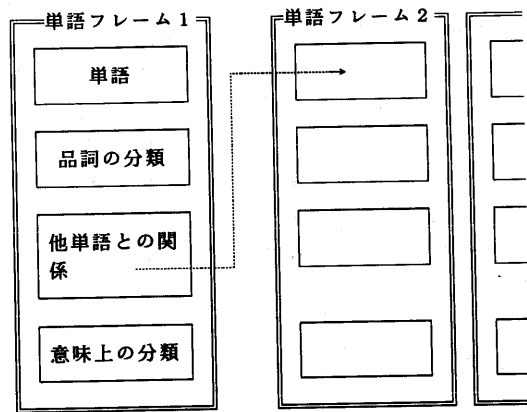
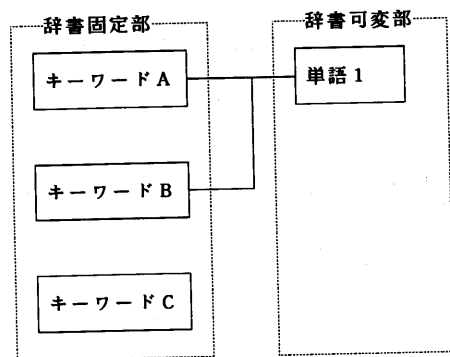
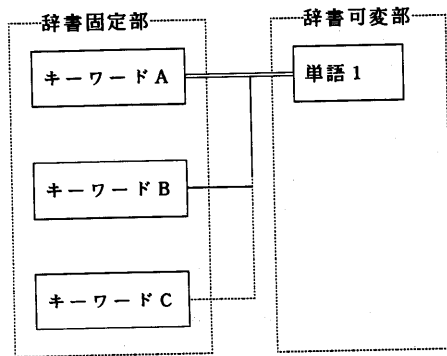


図10 単語辞書の内部構造



(1) 「新出単語1はキーワードAとキーワードBに対応する意味を持つ」ことを表現する



(2) 「新出単語1はキーワードAと特に関係し、キーワードB、キーワードCについてはその順番で関係する」ことを表現する

図11 単語辞書内での、単語間関係

もできる。図12において、ユーザと辞書の間に引いた点線は、辞書がユーザのインタフェース内部での投影であることを示す。辞書はユーザの利用する単語を記録するので、ユーザのプログラミングに関する知識を反映していると考えられる。そこで、辞書の内容を適当に評価することでユーザの知識を評価し、これに基づいてユーザインタフェースを変更してユーザ毎に操作環境を変化させることで、操作環境を向上させることができる(図12の、右向きの矢印)。すなわち辞書をユーザモデルとして利用することが可能であると考えられる。

3. 2 意味構造生成部および意味比較部の構成

意味構造生成部では、解析部で抽出したキーワードと成り得る単語と、動詞を中心とした構文情報を基に意味構造を生成する。意味構造は格フレームで記述する。意味構造生成の過程を図13に示す。意味比較部では、生成した格フレームを、同様の形式を持つ部品知識データベースと比較し、該当する部品を検索する。(図14)

3. 3 部品登録部の構成

部品登録部は、意味比較部が単語登録部に入れ替わることを除いて部品検索部と同様の構成となっている。図15に部品

登録部の構成を示す。図中、解析部、意味生成部および付随する辞書群は、部品検索部と同じものである。前にも述べた

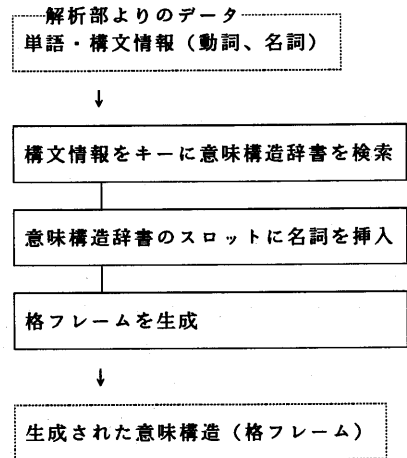


図13 意味構造生成の過程

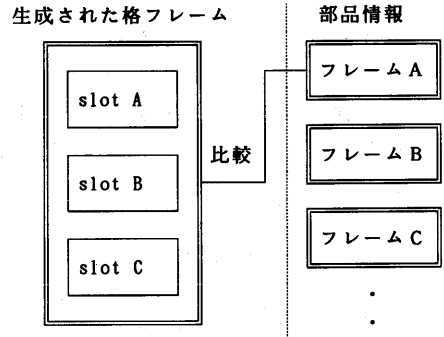


図14 部品データベースの検索

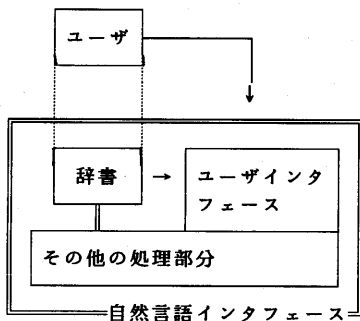


図12 ユーザモデルとしての辞書

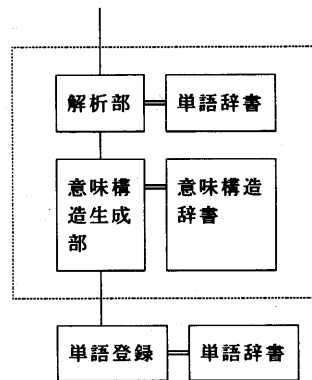


図15 部品登録部の構成

ように、これにより部品データベース内での意味表現の統一を図る。また、部品のデータベースへの登録に際して必ず作成する必要のある意味表現を対話的に生成できるので、登録が容易になる利点もある。

4 あとがき

以上、単語学習を行い、ユーザ毎に学習結果が切換え可能な自然言語インタフェースの構成方法について述べた。現在、解析部の試作を終え、意味構造生成部、意味比較部の試作を行っている。解析部については、自然言語に対する構文上の制限は大きい、十分な解析速度を得ることができたため、速度に関しては実用上の問題は生じないことが確認できた。また検索の際に各ユーザ毎に別々の辞書を用意することで、ユーザ間の検索についての個人差を吸収でき得ることが確認された。

今後試作を進め、部品化・再利用支援ツールとしての有効性を確かめる予定である。

参考文献

- [1] 古宮、原田：“部品合成による自動プログラミング”，情報処理，Vol.28 No.10，pp.1339-1340.
- [2] 小高、内山：“ソフトウェア製造ツールへの一提案 — Alice によるソフトウェア開発 —”，情報処理学会第35回（昭62後期）全国大会，7W-8，pp.995-996.
- [3] 長尾、野村：“日本語情報処理”，pp.129，コロナ社（1984）.
- [4] 小高、内山：“ソフトウェア製造ツールへの一提案 — Alice のツール LC と HD —”，情報処理学会第34回（昭62前期）全国大会，5S-5，pp.1037-1038.
- [5] 小高、内山：“ソフトウェア

の部品化・再利用を支援するプログラミング言語について”，信学論 D，Vol.J 70-D No.8（1987-08）

[6] 小高、内山：“ソフトウェアの部品化・再利用向きプログラミング言語について”，昭63信学総全大，D-376.

[7] 小高、内山：“ソフトウェア製造ツールへの一提案 —ソフトウェア部品検索の一手法—”，情報処理学会第36回（昭63後期）全国大会，7L-4，pp.959-960.

[8] B.A. Burton and R.W. Aragon, S.A. Bailey, K.D. Koehler and L.A. Mayes：“The Reusable Software Library”，IEEE J. Software, July 1987, pp.25-33.

[9] M. Zemankova-Leech and A. Kandel：“FUZZY RELATIONAL DATA BASES”，pp.37-38，（1984）[邦訳 向殿 政夫訳：“ファジィ・リレーショナル・データベース”，啓学出版（1987）].