

## 事例ベースに基づく仕様記述環境における 類語辞書を用いた利用者要求獲得

唐橋 拓史<sup>†</sup>      吉村 晋<sup>‡</sup>      白鳥 則郎<sup>†</sup>

<sup>†</sup>東北大学 電気通信研究所 / 情報科学研究科

<sup>‡</sup>(株) 東芝 情報・通信システム技術研究所

通信ソフトウェア開発において、過去の仕様、知識の再利用を行なう局面は、開発の全工程において存在する。特にソフトウェア開発の上流工程を支援する目的で、仕様記述言語 HSC の設計とその記述支援環境である AOBA システムが開発された。AOBA システムは事例ベースに基づき利用者に再利用支援を提供するが、利用者が用いる語の管理を行なう類語辞書の構造に難があり、事例ベースに格納した事例をうまく活用できなかった。本稿では、事例をより有効に活用できるような類語辞書の構造と、新しい知識を得たときの更新方法を提案する。

## User-Requirement Aquisition on the Specification Description Environment using Keyword Bank

Takuji Karahashi<sup>†</sup>      Susumu Yoshimura<sup>‡</sup>      Norio Shiratori<sup>†</sup>

<sup>†</sup>Research Institute of Electrical Communication/Graduate School of Information  
Sciences, Tohoku University, Sendai, Japan

<sup>‡</sup>Information & Communication Systems Lab., TOSHIBA CORPORATION

In the development of communication software, one of the effective methods for advancing the productivity of software is specification reuse. To support the specification development, we developed a specification description language HSC(Hierarchical Sequence Chart) and implemented a support environment, called AOBA System. AOBA System support searching of similar components and describing new specification based on CBR techniques. But AOBA System effectively cannot use the reusable components because of the structure of Keyword Bank and the method of renewing it. In this paper, to use the reusable components more effectively, we propose new structure of Keyword Bank and the renewing method.

## 1 はじめに

近年、情報通信技術の発展にともない、さまざまな局面で情報通信システムが利用されている。このようなシステム上で提供されるサービスは利用者から要求が出され、開発者の手によってシステム上に実装されるが、効率の良い開発を行うためには過去のソフトウェア資産をいかにうまく利用するかが重要である。しかし、通信ソフトウェアの共通利用の実現には様々な困難がある。我々は通信ソフトウェア資産を蓄積して、自由に再利用できる仕様記述支援システム AOBA[1][2][3]を開発し、通信ソフトウェア開発へ適用する研究を行った。

AOBA は仕様を記述するときに、利用者の要求に応じて過去に記述された仕様部品を利用者に提示し、また必要に応じて修正箇所を指摘、記述された仕様は新しい部品として蓄積する機能を持つ。このシステムによって利用者は仕様記述における負担を軽減できるが、再利用部品の検索時に用いられる類語辞書の構成が単純であるため場合によっては適当な部品が検索されないことや、類語辞書の更新が管理者の手にまかされるため長期間の利用において管理者の負担が大きいことが問題とされていた。

本稿では、前者の問題を解決するための新しい類語辞書の構成について述べる。また、後者の問題を解決するために、システム上で利用されるキーワードの意味を定義し、それに基づきシステムによる類語辞書の更新法について述べる。

本稿の構成は以下のとおりである。まず最初に2章で AOBA システムとその問題点について述べる。3章で新しい類語辞書の構成を提案するとともに、4章ではシステムによる類語辞書の更新方法を提案する。

## 2 要求記述支援環境 AOBA

本章では本論文で対象としているシステムとして、要求記述支援環境 AOBA システムについて述べる。

### 2.1 要求記述言語 HSC

HSC[9] はシーケンス図を基本として拡張した言語で、以下の特徴を持つ。

1. 階層化とモジュール化
2. 繰り返しと条件分岐
3. メッセージ間の関係

HSC は仕様全体を仕様枠組、フェーズ、ケースの階層で表現するため、各階層において仕様の一部分を部品化することが可能である。そのため、部品の

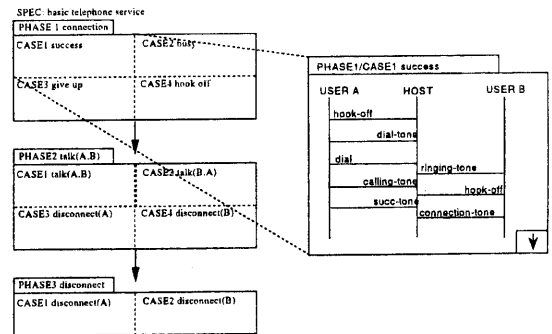


図 1: 基本電話サービスの HSC による記述

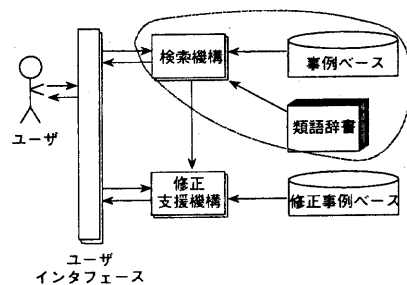


図 2: AOBA システム

特徴に応じてデータベースに登録することで、後に部品を再利用して新しい仕様を記述することが容易となる。HSC の記述例を図 1 に示す。これは基本的な電話のサービスをあらわしたもので、接続・通話・切断のフェーズをもつ。

### 2.2 AOBA システム

AOBA システムは HSC の特徴である部品の再利用に基づいた開発がしやすいという点を生かして HSC 仕様の記述を支援するシステムである。AOBA システムはユーザインタフェース、2つの機構、2つのデータベースおよび類語辞書から構成される。(図 2)

AOBA システムは、事例ベース推論の手法を用いて再利用記述を支援している。支援の内容は以下の通りである。

まず、AOBA システムは利用者の要求を要求獲得情報の形で獲得する。検索機構はこの情報に基づいて、事例ベースから利用者の要求に近い事例を幾つか検索し、利用者に提示する。利用者はこの候補の中から一つの事例を選択し、それを基にして新しい

事例を作成する。このとき修正支援機構が利用者から得た要求獲得情報と選択された事例の情報との間の差を基にして修正箇所の指摘を行なう。最後に利用者の入力した要求獲得情報と記述した仕様を基に事例を作成、格納する。以上の支援がHSC仕様構造の各階層において再帰的に適用される。

### 2.3 AOBAシステムにおける問題点

AOBAシステムは、SUNワークステーション上にCBRツールART-IM [10]、X-windowシステムおよびC言語を用いて試作された。このシステムを用いて記述実験を行なったところ、以下のような問題点が指摘された。

- AOBAシステムは複数の利用者が同じ事例ベースを用いて仕様を記述する環境を提供している。AOBAシステムを利用して複数の利用者が様々な分野の仕様を記述するとき、同じキーワードを様々な意味で利用する状況が考えられる。しかし、従来のAOBAシステムの類語辞書は、一つのキーワードは一つの意味を持つものとして構成されていたため、このような状況には対応できず、利用者が意図した事例が検索されとは限らない。
- 類語辞書ではキーワード間の類似性が定義されている。AOBAシステムを有効に利用するためには、利用者が新しいキーワードを入力したり、事例を記述するごとに類語辞書を更新しなければならない。現在は管理者が更新しているが、キーワード、事例の数が増えると、管理者にとって多大な負担となり、いずれ管理が行き届かなくなる。

## 3 類語辞書の構成の変更

本章では、2.3で挙げた1番目の問題点の一つを解決するために、AOBAシステムで用いられている類語辞書の新しい構成を提案する。

### 3.1 意味の多様化

前述したように、AOBAシステムにおいて利用者が要求を記述するために用いるキーワードは一つの意味で用いられるとは限らない。しかし、従来の類語辞書では各々のキーワード間の類似度を定義しているだけなので、この問題に対処できなかった。そこで類語辞書の構成を変更し、一つのキーワードに対して複数の意味を対応付けることを考える。

複数の意味を扱う時には、どの意味をどの場面で用いるかを定義しなければならない。そこでキーワー

ドを用いる場面を区別するために、状況情報  $SI$  を導入する。 $SI$  は要求獲得情報において入力される分野情報  $F_c$  と付加情報  $A_c$  の2項組とする。

$$SI := \langle F_c, A_c \rangle$$

これを用いて類語辞書内ではキーワード  $K$  と状況情報  $SI$  の組  $SK$  同士の類似度を定義する。すなわち、

$$SK := \langle K, SI \rangle$$

としたとき、類語辞書は  $SK_i$  と  $SK_j$  および  $SK_i$  と  $SK_j$  間の類似度  $Sim(SK_i, SK_j)$  の3項組の集合とする。これにより、一つのキーワードを複数の意味で扱うことができる。

### 3.2 キーワード間の類似度

類語辞書の構成の変更によって事例の検索時のキーワード間の類似度の決定方法も変更される。ここでは、利用者から得られた要求獲得情報  $I_r$  において用いられるキーワード  $K_r$  と検索時に対象となる事例の要求獲得情報  $I_c$  において用いられるキーワード  $K_c$  との間の類似度の決定方法を示す。

1. 利用者が入力した要求獲得情報  $I_r$  の中から状況情報  $SI_r$  を取り出す。
2. 検索対象の事例の要求獲得情報  $I_c$  の中から状況情報  $SI_c$  を取り出す。
3. 類語辞書に登録されているキーワードと状況情報の組の中で、キーワード  $K_r$  と、状況情報  $SI_r$  に最も類似している状況情報の組を探す。
4. 3で見つけた組と、 $K_c, SI_c$  の組との類似度をキーワード  $K_r$  とキーワード  $K_c$  間の類似度とする。

### 3.3 状況間の類似度

キーワード間の類似度の検索を行うために、状況情報間の類似度を定義する。状況情報  $SI$  は分野情報  $F_c$  と付加情報  $A_c$  の組であり、また  $A_c$  は複数の情報の組である。例えば、電話サービスにおいては付加情報は転送モード、前状態、上位構造、設計者の組となっている。これらの情報をもとに状況情報  $SI_a, SI_b$  間の類似度  $Sim_{situation}$  を以下のように定義する。

$$Sim_{situation} = g(F_{c_a}, F_{c_b}) \times \alpha + g(A_{c_{a_i}}, A_{c_{b_i}}) \times \beta_i$$

$$g(word1, word2) = \begin{cases} 1 & \text{if word1=word2} \\ 0 & \text{otherwise} \end{cases}$$

$$\alpha, \beta_i (i = 1, 2, \dots) : \text{それぞれの情報の重み付け}$$

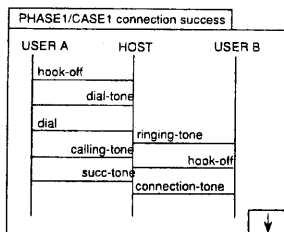


図 3: HSC のケースの例

## 4 類語辞書の更新方法

本章では 2.3 で挙げた 2 番目の問題を解決するために、類語辞書の更新方法を提案する。類語辞書はキーワード間の類似性を定義するものである。従来は、この類似性は管理者の経験によって定義されていたが、システムによる更新を考えるにあたり、まず類似性の基準を定義する。さらにそれに基づいて類語辞書の更新方法を提案する。

### 4.1 類似性の定義

**ケースにおけるキーワードの類似性** ケースにおいて用いられるキーワードは、利用者がケースを連想するために用いられる。ケースは各々のフェーズにおいて発生する可能性のある動作を表現する。この動作はシーケンスチャートによって表現される。例えば、図 3 は電話サービスにおいて USER A が USER B に電話をかけ、繋がった場合の動作である。このようなケースをキーワードで表現する場合、過去の実験から、AOBA システムの利用者はシーケンスチャートのノード間でやりとりされるメッセージからキーワードを連想することが多い。そこで、ケースのキーワードはシーケンスチャートのメッセージに対応すると考えることができる。

しかし、システムの利用者から得られる情報でキーワードとメッセージ間の対応を特定することは困難である。ここでは、あるシーケンスチャートの中で用いられるメッセージを表現する方法として、関係のあるメッセージによって表現するという手法をとる。すなわち、ある事例内のケース用のキーワードに対して、対応するケースの中で用いられているメッセージの組を対応付ける。これは、キーワードとメッセージが対応するという考えから、あるキーワードは同一の要求獲得情報内で用いられているキーワードと対応付けることと同義である。

**フェーズにおけるキーワードの類似性** フェーズにおいて用いられるキーワードは、利用者がフェーズを連

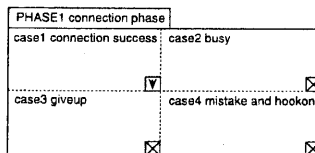


図 4: HSC のフェーズの例

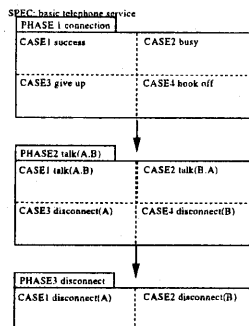


図 5: HSC の仕様枠組の例

想するために用いられる。フェーズは仕様枠組内における一つの局面を表現している。これは HSC においては複数のケースの組とケース間の関係で表現される。例えば、図 4 は電話サービスにおいて USER A が USER B に電話をかける局面である。このようなフェーズをキーワードで表現する場合、過去の実験から、AOBA システムの利用者は、フェーズ内で用いられるケースからキーワードを連想することが多い。そこで、フェーズのキーワードはフェーズを構成するケースに対応すると考えることができる。

以上からフェーズのキーワード間の類似性はフェーズとケースの関係を用いて定義することができる。ここでシステムの利用者から得られる情報としてそのケースに対応する要求獲得情報を利用することができる。

**仕様枠組におけるキーワードの類似性** 仕様枠組において用いられるキーワードは、利用者が仕様枠組を連想するために用いられる。仕様枠組は利用者が記述しようとしているサービスそのものを表現している。これは HSC においては複数のフェーズの流れで表現される。例えば、図 5 は基本電話サービスを表わしている。このような仕様枠組をキーワードで表現する場合、過去の実験から、AOBA システムの利用者はその仕様枠組が提供するサービスが持つ機能の名前からキーワードを連想することが多い。そこで、

仕様枠組のキーワードは仕様枠組内で記述される機能に対応すると考えることができる。

以上から仕様枠組のキーワード間の類似性は仕様枠組と機能の関係を用いて定義することができる。ここで個々の機能を表現するものは、仕様枠組内で記述されるフェーズおよびケースの組であると考えることができる。さらに、ケースのもつ情報がフェーズのもつ情報よりも極端に小さいことを考慮すると、仕様枠組の類似性を決定するものとして、その機能を表わすフェーズに対応する要求獲得情報を用いることができる。

## 4.2 キーワード間の類似度

HSCのそれぞれの階層について、4.1で定義した類似性に基づいて類似度を定義する。

ケースに用いられるキーワード間の類似度 ケースに用いられるキーワード  $K_{c_1}$  と  $K_{c_2}$  の間の類似度  $Sim_c(K_{c_1}, K_{c_2})$  は以下のように決定する。

$K_{c_1}, K_{c_2}$  と同じ要求獲得情報内のキーワードがそれぞれ

$$\mathbf{K1} = (k_{11}, k_{12}, \dots, k_{1m})$$

$$\mathbf{K2} = (k_{21}, k_{22}, \dots, k_{2n})$$

であるとする。このとき、

$$Sim_c(K_{c_1}, K_{c_2}) = \frac{\frac{1}{m} \sum \omega_1(k_{1i}, \mathbf{K2}) + \frac{1}{n} \sum \omega_1(k_{2j}, \mathbf{K1})}{2}$$

$$\omega_1(a, \mathbf{K}) = \max S(a, b) \quad b \in \mathbf{K}$$

ただし、 $S(a, b)$  はキーワード  $a, b$  間の類似度を表わす。

フェーズに用いられるキーワード間の類似度 フェーズに用いられるキーワード  $K_{p_1}$  と  $K_{p_2}$  の間の類似度  $Sim_p(K_{p_1}, K_{p_2})$  は以下のように決定する。

$K_{p_1}, K_{p_2}$  を構成するケースの組がそれぞれ

$$\mathbf{C1} = (c_{11}, c_{12}, \dots, c_{1m})$$

$$\mathbf{C2} = (c_{21}, c_{22}, \dots, c_{2n})$$

であるとする。このとき、

$$Sim_p(K_{p_1}, K_{p_2}) = \frac{\frac{1}{m} \sum \omega_2(c_{1i}, \mathbf{C2}) + \frac{1}{n} \sum \omega_2(c_{2j}, \mathbf{C1})}{2}$$

$$\omega_2(a, \mathbf{C}) = \max S(a, b) \quad b \in \mathbf{C}$$

ただし、 $S(a, b)$  はケース  $a, b$  間の類似度を表わす。

仕様枠組に用いられるキーワード間の類似度 仕様枠組に用いられるキーワード  $K_{s_1}$  と  $K_{s_2}$  の間の類似度  $Sim_s(K_{s_1}, K_{s_2})$  は以下のように決定する。

$K_{s_1}, K_{s_2}$  を構成するフェーズの組がそれぞれ

$$\mathbf{P1} = (p_{11}, p_{12}, \dots, p_{1m})$$

$$\mathbf{P2} = (p_{21}, p_{22}, \dots, p_{2n})$$

であるとする。このとき、

$$Sim_s(K_{s_1}, K_{s_2}) = \frac{\frac{1}{m} \sum \omega_3(p_{1i}, \mathbf{P2}) + \frac{1}{n} \sum \omega_3(p_{2j}, \mathbf{P1})}{2}$$

$$\omega_3(a, \mathbf{P}) = \max S(a, b) \quad b \in \mathbf{P}$$

ただし、 $S(a, b)$  はフェーズ  $a, b$  間の類似度を表わす。

## 4.3 更新手順

類語辞書の更新手順を以下に示す。

1. 通常のAOBAの手順通り、利用者が入力した要求獲得情報をもとにして、事例ベース内の事例を検索、類似度の高い候補を利用者に提示する。利用者はこの中から一つの事例を選択。
2. 利用者が入力した要求獲得情報に存在して利用者が選択した事例の要求獲得情報に存在していないキーワードを特定する。
3. 利用者が選択した事例をもとにして修正を行なった後、その修正内容から、2.で得られたキーワードに関する情報を得て、両者の組を類似情報として記録する。
4. 類似情報をもとにして、2.で得られたキーワードを既存のキーワードとの間の類似度を(再)計算する。
5. 4.の計算結果を類語辞書に登録する。

## 5 まとめ

本稿では通信ソフトウェアの再利用を目的とした仕様記述支援システムAOBAにおいて、キーワードに複数の意味を持たせ、またシステムによる類語辞書の更新方法について述べた。今後は実装を完成させ、それを用いて記述実験を行い、その効果を評価する予定である。

## 参考文献

- [1] Ching-Fa HUANG, Susumu YOSHIMURA, Takuji KARAHASHI, Norio SHIRATORI: "A New Specification Environment for Communication Systems Based on Specification Reuse"

- by the Application of Case Based Reasoning," *IEICE TRANS. INF. & SYST.*, VOL. E787-D, NO.10, October, pp.1269-1281(1995).
- [2] C.F.Huang, T.Karahashi, E.S.Lee, N.Shiratori: "A Flexible Service Development Support System for Communication Systems by Reuse Methodology," *ICPADS94* Tiwan, Republic of China, December, pp.426-431(1994).
- [3] C.F.Huang, T.Karahashi, E.S.Lee, N.Shiratori: "An Expert System for Flexibly Developing Services on Communication Systems by the Application of Case Based," *Proc. of ICOIN-9*, Japan, December, pp.541-546(1994).
- [4] Dieter Merkl, A Min Tjoa, Gerti Kappel: "Learning the Semantic Similarity of Reusable Software Components," *Proceedings of Third International Conference on Software Reuse: Advances in Software Reusability*, Rio de Janeiro, Brazil, November, pp.33-41(1994).
- [5] Evangelos Simoudis: "Using Case-Based Retrieval for Customer Technical Support," In *IEEE EXPERT*, OCTOBER, pp7-11(1992).
- [6] Janet Kolodner: "Case-Based Reasoning," Morgan Kaufmann Publishers, Inc.
- [7] M.R.Genesereth, N.J.Nilsson: "Logical Foundations of Artificial Intelligence," オーム社
- [8] Norio SHIRATORI, Kenji SUGAWARA, Tet-suo KINOSHITA, Members and Goutam CHAKRABORTY, Nonmembers: "Flexible Networks: Basic Concepts and Architecture" *IEICE TRANS. on Communications*, VOL. E777-B, NO.11, NOVEMBER, pp.1287-1294(1994).
- [9] 黄錦法、白鳥則郎: "シーケンス図に基づく通信システム仕様記述法 HSC とその支援環境," 情報処理学会論文誌、 Vol.34, No.6(1993).
- [10] "Case-Based Reasoning in ART-IM". Inference Corporation, California, USA(1991).