

ソフトウェア常識集知的検索エキスパートシステム

Intelligent Information Retrieval System for Common Knowledge of Software

日立製作所では、ソフトウェア不良の作り込み防止を図るため、プログラム作成時の失敗例とその教訓をソフトウェア常識集にまとめ教育に活用してきた。この常識集を更に有効に活用するために、ソフトウェア常識集知的検索エキスパートシステムを開発した。本システムの特徴は、利用者が入力したキーワードから関連するキーワードを連想し、これらのキーワードと教訓を関連づける連想検索方式にある。利用者の試用によって、本方式が検索漏れの防止及び検索効率の向上に効果があることを確認した。

更に、連想検索方式を同種システムへ適用するために、本システムからソフトウェア常識集に関する専門知識を除くことによって、はん(汎)用的な検索形エキスパートシステム構築ツールを開発した。

橋本 肇* *Hajime Hashimoto*
神品 芳明* *Yoshiaki Kôshina*
辻 洋** *Hiroshi Tsuji*
安信千津子** *Chizuko Yasunobu*

1 緒 言

ソフトウェアの社会的重要性の増大に伴い、その信頼性に対する要求は一段と厳しくなってきた。ソフトウェアの信頼性確保のためには、構造化設計、デザインレビューなどが実施されているが、仕掛けと同時にしつけ・教育も大切である。ソフトウェアの生産は個人に依存しているのが現状であり、検討不足・認識不足などから類似不良を再発させている事例が少なくない。このため過去の不良事例を蓄積し、有効に活用していくことが重要である。

日立製作所では、ソフトウェア不良の作り込み防止を目的として、プログラム作成時の失敗例とその教訓をソフトウェア常識集としてまとめ、プログラムの教育、自己啓発などに活用している。この常識集には、プログラム作成に関する社内のノウハウが蓄積されており、更に効果的に活用するために、効率のよい検索システムが期待されていた。

一方、通常のキーワード検索では、入力キーワードと検索対象に付与されたキーワードが一致しないと検索できないこと、及び検索結果の選択をすべて利用者が判断しなければならないという欠点があった。

今回、これらの欠点を補うために、キーワードの指定及び検索結果の選択に、熟練者のノウハウを活用した連想検索方式を考案し、本方式を適用したソフトウェア常識集知的検索エキスパートシステム^{1)~3)}を開発した。今回考案した連想検索方式によれば、あらかじめシステムにキーワードを連想する知識と検索結果を順位づける知識を持たせることによって、検索漏れの防止及び検索効率の向上が図れる。このような知的検索システムに対するニーズは高く、連想検索方式が適用可能な分野は多いと考えられる。そこで、ソフトウェア常識集知的検索エキスパートシステムを母体として改善を加え、

はん用的な検索形エキスパートシステム構築ツールも開発した。

本稿では、ソフトウェア常識集知的検索エキスパートシステムの機能と実現方式、及び連想検索方式の応用例について述べる。

2 エキスパートシステム化のねらい

2.1 ソフトウェア常識集

ソフトウェア常識集は、プログラム作成時の失敗例とその教訓を集めたもので、次のような特徴がある。

- (1) 事例と教訓がイラスト入りで一件一葉にまとめられている。
- (2) 実践に基づいた多くの事例(13分類, 161件)が載っている。
- (3) 個人の自己啓発、グループ勉強会、講習会などの教育に使われ効果を挙げている。
- (4) ソフトウェア不良の大部分が、いずれかの教訓に該当する。

2.2 従来のIRシステムとの違い

ソフトウェア常識集は、ソフトウェア不良の根本原因が教訓として抽象的に記述されている。そのため、初心者にとって不良事例から該当する教訓を見極める場合、又はコーディング上注意しなければならない教訓を得る場合、検索しにくいという問題がある。

一方、従来のIR(Information Retrieval)システムでは、次のような欠点がある。

- (1) 各事例にキーワードを付与して検索するが、入力されたキーワードと同じキーワードの事例しか検索できない。そのため入力キーワードが不足している場合、検索漏れが生じる。

* 日立製作所ソフトウェア工場 ** 日立製作所システム開発研究所

(2) 事例に付与したキーワードが同じ場合でも、ある事例には密接に関連するが、別な事例にはあまり関連しないという場合が多く、検索結果を利用者の判断で取捨選択しなければならない。

以上の欠点に対し、ソフトウェア常識集を理解している熟練者について、次のように仮定した。

- (1) 個々の不良事例から幾つかの具体的なキーワードを抽出し、抽出したキーワードから同意語、関連語を重みを付けて連想できる。
- (2) 抽出及び連想したキーワードと各教訓を重みを付けて関連づけし、最も関連する教訓を選択できる。

エキスパートシステム化のねらいは、この仮定に基づいて熟練者の知識を収集し、知識ベースに蓄積することによって、初心者が入力したキーワードから熟練者が選ぶのと同等の教訓を得られるようにすることである。

3 システムの概要

3.1 システムの構成

システムの構成を図1に示す。ソフトウェア常識集知的検索エキスパートシステムは、日立クリエイティブワークステーション2050上でスタンドアロン方式で動作する。推論エンジンはエキスパートシステム構築ツールES/KERNEL⁴⁾(Expert System/KERNEL)である。知識ベースは、その保守を容易にするために、ES/KERNELが提供する知識テラ⁵⁾を用いている。ソフトウェア常識集の各教訓は、ファイル容量を少なく、追加修正を容易にするために、OFIS/REPORT(OFIS/TEXT, OFIS/SKETCHなどを含む。)を使用し、文書ファイルに格納している。

3.2 システムの特徴

本システムの特徴を次に述べる。

- (1) 操作が簡単なマウス入力である。また、出力結果は、マルチウインドウ、図形表示を活用している。
- (2) 検索方式として基本検索、キーワード検索及び連想検索の3種類をサポートしている。

(a) 基本検索

教訓の分類、題目を選択する検索であり、書籍と同様に目次から必要ページを検索できる。

(b) キーワード検索

一般的なIRシステムと同じように、入力したキーワードを含む教訓を検索する。

(c) 連想検索

入力されたキーワードから関連するキーワードを連想し、キーワードと関連の強い順に教訓を検索する。

- (3) 連想検索では、熟練者のノウハウを活用しているため、検索精度が向上する。
- (4) システムの保守はルール、フレームの保守であり容易にできる。
- (5) システムは、文書ファイルを含め、3.5インチフロッピーディスク2枚に格納してあり、知識ベースを媒体とした知識の流通が容易である。

4 連想検索に関する機能

連想検索に関してシステムが提供する機能を検討するに当たり、利用者とシステムの役割分担を次のように決めた。

- (1) キーワードの指定は利用者が行う。そのため、入力キーワードが不適切である場合がある。また、入力操作を簡単にする必要がある。
- (2) キーワードの連想はシステムが行う。ここでは、連想の重みをどのように決めるかが課題となる。
- (3) 教訓への関連づけはシステムが行う。ここでも、関連づけの重みをどのように決めるかが課題となる。
- (4) 最終的な結果は、利用者が満足できる教訓を検索した時点で決まる。

この役割分担に従って、本システムが連想検索に関して提供する機能を次に述べる。

(1) キーワード入力機能

利用者が連想検索を選択すると、入力画面例(図2)に示すキーワード一覧が表示される。利用者はマウスで複数個のキーワードを選択することができる。400以上のキーワードがあるため、種々のスクロール機能をサポートし、利用者の便宜を図っている。

(a) 上下・ジャンプスクロール機能

(b) 五十音順でのジャンプスクロール機能

(c) 分類別(プログラミング用語別)キーワード表示機能

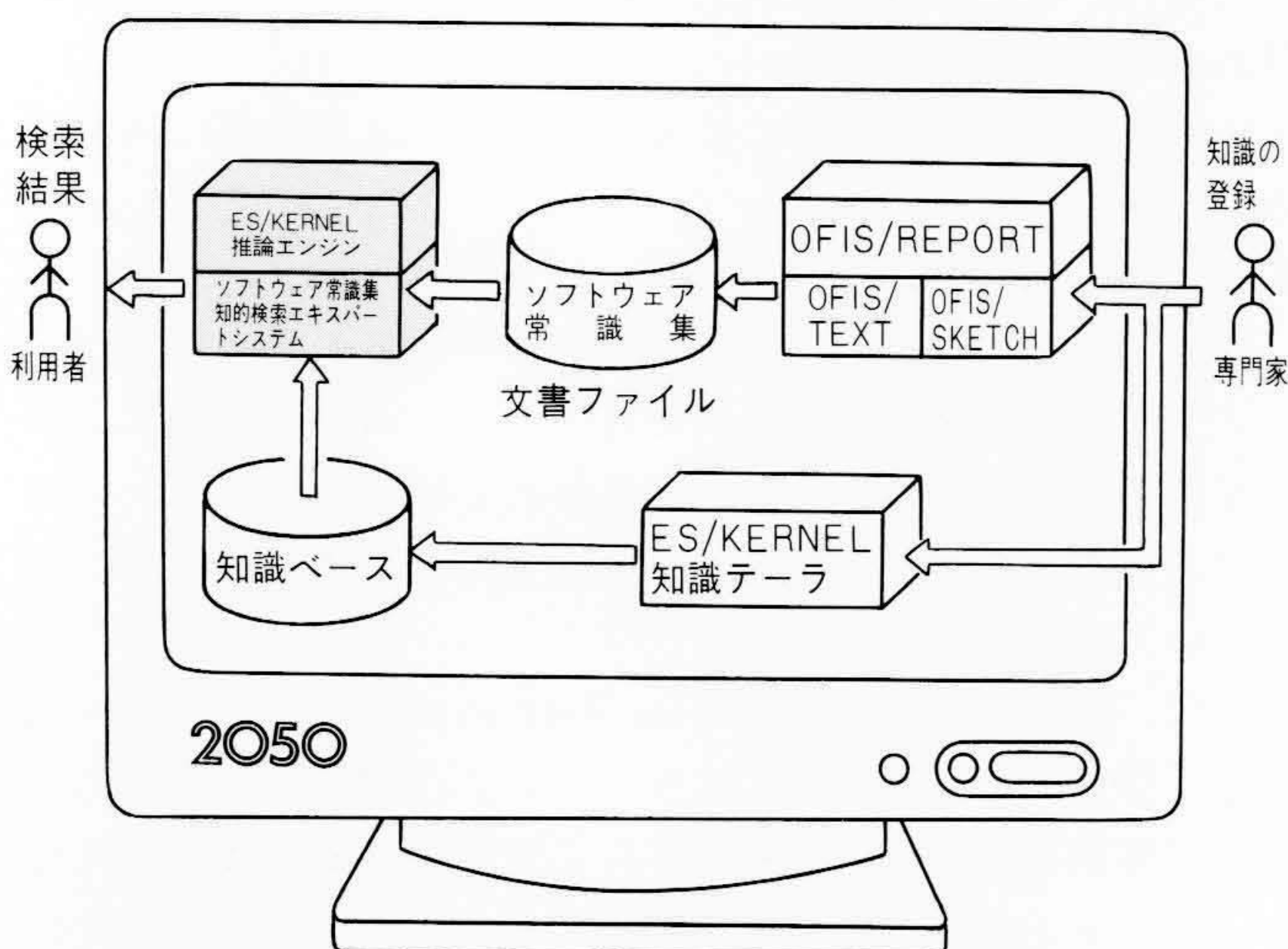
(d) キーワードの取消し機能

(2) キーワード連想機能

(a) キーワード連想機能

利用者が入力したキーワードから連想されるキーワードを追加し、表示する。キーワードの連想では、熟練者の知識を利用する。

(b) 不要な連想キーワードの削除機能



注：略語説明 ES/KERNEL (Expert System/KERNEL)
OFISシリーズ (Office automation and Intelligence support Softwareシリーズ)

図1 システムの構成 本システムは、日立クリエイティブワークステーション2050上で動作し、ES/KERNELで実現されている。

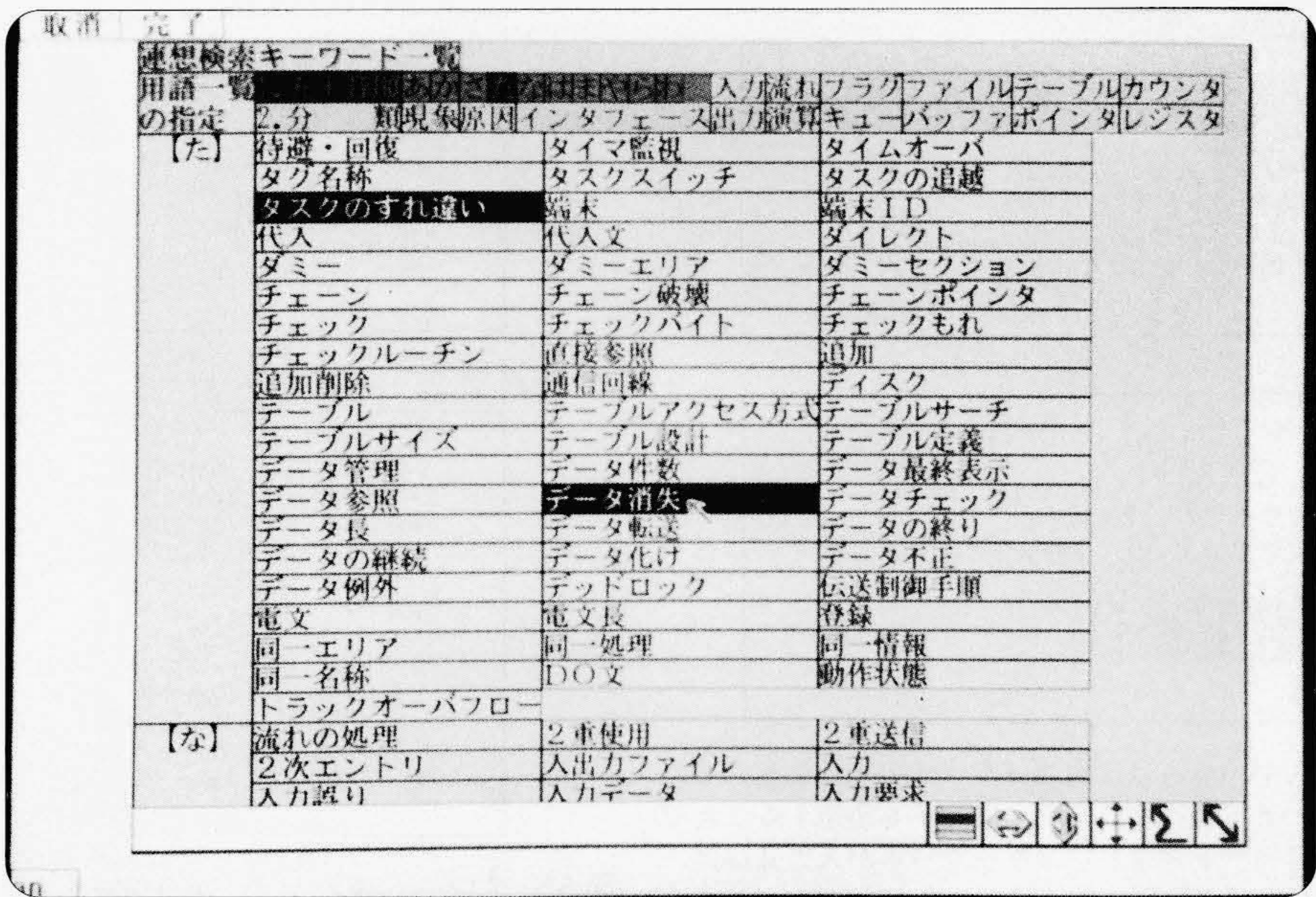


図2 入力画面例 キーワードを入力する画面の例を示す。画面上部の「五十音」又は「分類」で画面をスクロールし、キーワードをマウスで選択する。

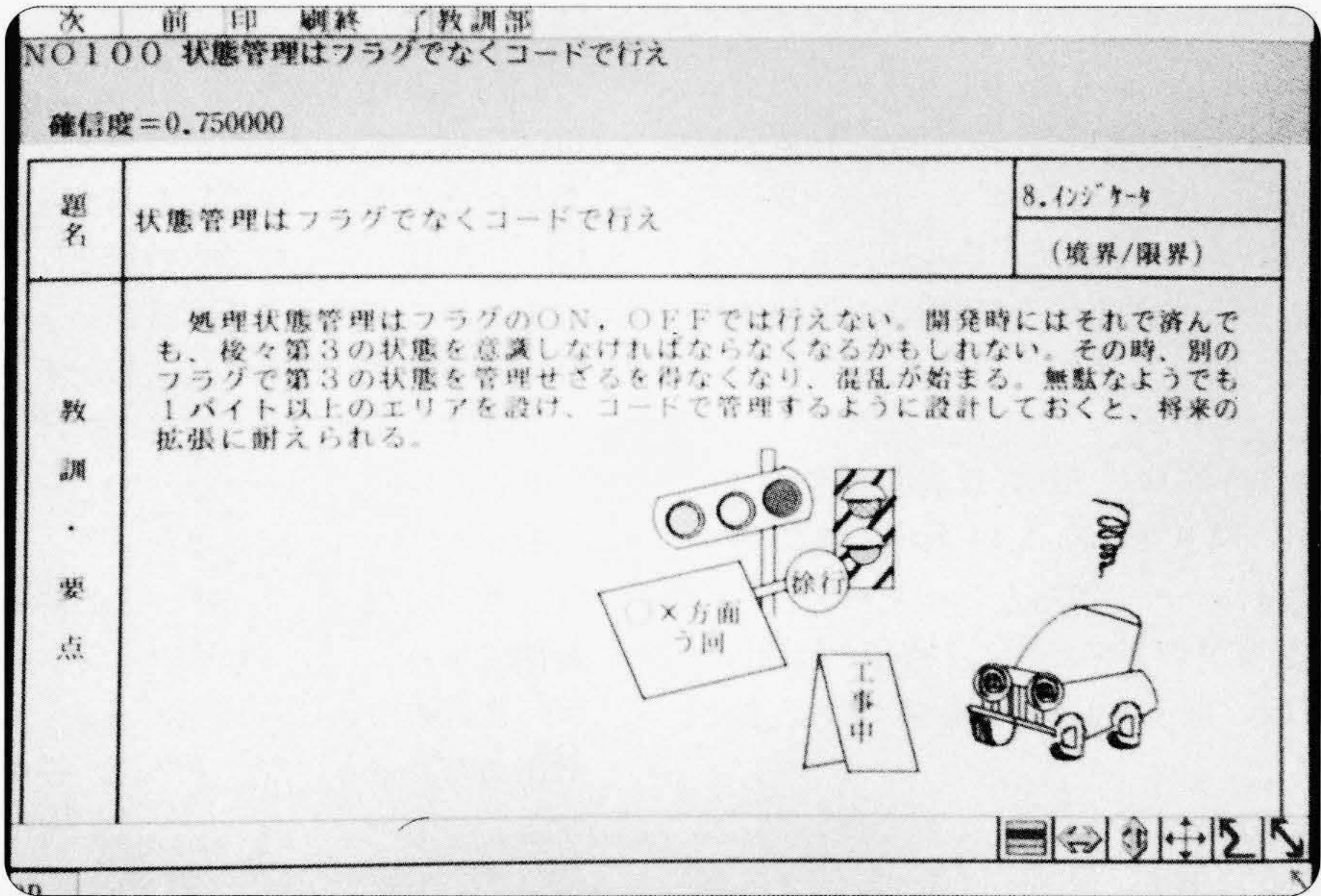


図3 出力画面例 検索結果はソフトウェア常識集と同じ様式で、確信度とともに出力される。

利用者にとって不要なキーワードを連想すると、検索範囲が広がり検索精度が悪くなる場合がある。これを救うために、不要なキーワードを削除できる。

(3) 検索結果の出力機能

(a) 検索結果の順位づけ機能

入力キーワード及び連想キーワードと教訓の関連の強さを算出し、関連の強い順に表示する。検索結果の順位づけでは熟練者の知識を利用する。出力画面例を図3に示す。

(b) 次・前コマンドによる教訓の表示機能

(c) 印刷コマンドによるプリンタ出力機能

(d) 検索件数の打ち切り機能

関連の強さがある値以下の場合、及び検索件数がある値

以上の場合、検索を打ち切ることができる。

(e) 推論説明機能

各教訓の題目と関連の強さ及び検索件数を表示する。

5 連想検索の実現方法

連想検索を実現する推論方式として、前向き推論及び後向き推論があるが、キーワード入力に関する操作性が優れていることから、前向き推論方式を採用した。

5.1 知識表現

知識表現はルールとフレームを用いた。キーワードを連想する知識をキーワード連想ルール、関連度を計算する知識を関連度計算ルールと呼び、ルール形式で表現した。また、キ

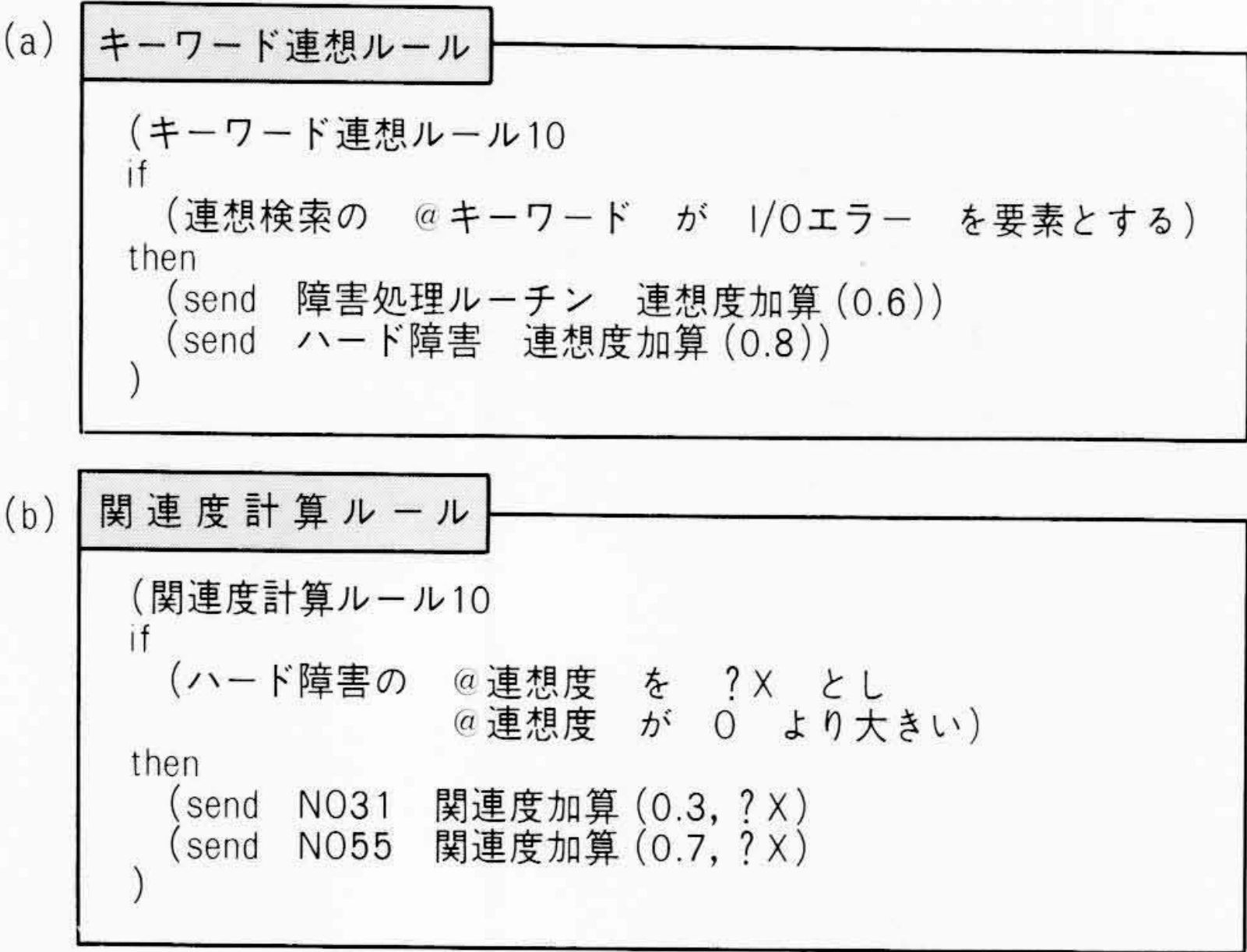


図4 ルール表現例 図中(a)の例は入力キーワードがI/Oエラーならば障害処理ルーチン及びハード障害というキーワードを連想することを意味する。同じく(b)の例はハード障害というキーワードが入力又は連想されたならば、NO31及びNO55の教訓を関連づけることを意味する。

ーワード及び教訓を事実形知識としてフレームで表現した。

(1) キーワード連想ルール

入力されたキーワードに対し、同意語、関連語を連想するルールで、連想の重みづけ(連想度と呼ぶ。)を行う。具体例を図4(a)に示す。この例では、I/O(入出力)エラーがキーワードとして入力されたとき、キーワード連想ルール10によって障害ルーチンが連想度0.6、ハード障害が連想度0.8で追加されることを示している。

(2) 関連度計算ルール

入力キーワード及び連想キーワードから教訓を関連づけるルールで、関連の重みづけ(関連度と呼ぶ。)を行う。具体例を図4(b)に示す。この例では、ハード障害というキーワードが入力又は連想された(連想度が0より大きい。)場合、教訓NO31が関連度0.3、教訓NO55が関連度0.7で関連づけられることを意味する。

(3) キーワードフレーム及び教訓フレーム

フレームの具体例を図5に示す。同図(a)はキーワードフレームを表し、キーワード数だけ作成し、連想度を設定するスロットを設けている。同図(b)は教訓フレームを表し、教訓数だけ作成し、関連度を設定するスロットを設けている。

5.2 知識ベースの構築

知識ベースの構築は以下に示す手順で実施した。

- (1) キーワード抽出(関連するキーワードの抽出)
- (2) キーワード分類(抽出したキーワードの分類)
- (3) ルールの作成(キーワード間の連想関係及びキーワードと教訓の関連の定義)
- (4) 連想度の設定(キーワード間の関連の強さの設定)
- (5) 関連度の設定(キーワードと教訓の関連の強さの設定)

連想度及び関連度の設定について以下に詳述する。

本システムの連想度及び関連度は、次に述べる設定基準を設け、4人の熟練者(10年以上の経験者)が5段階で設定した。

(1) キーワードの連想度設定基準

- (a) キーワード間の関連の強さ(連想の確かさ)を〔0, 1〕

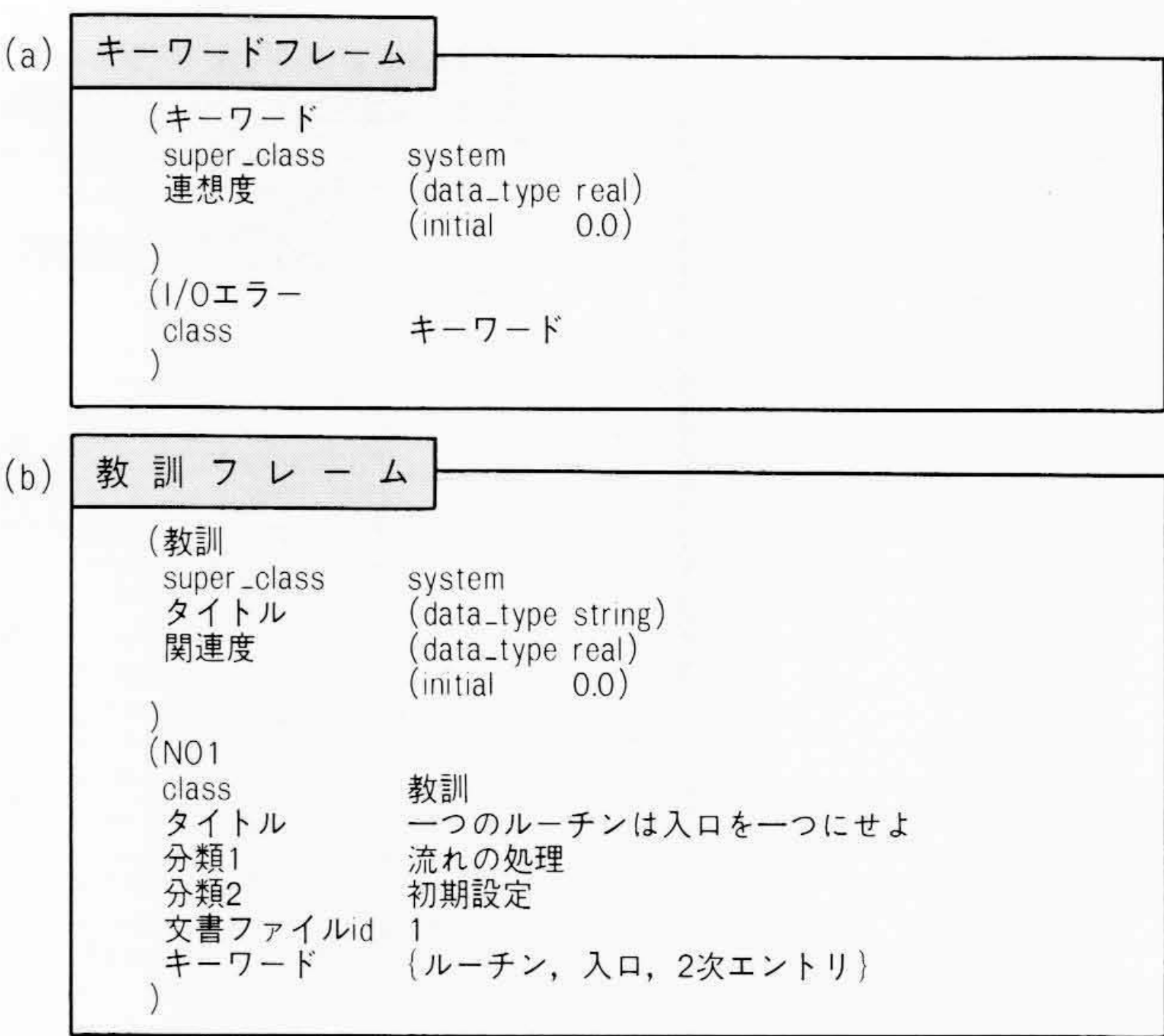


図5 フレーム表現例 キーワードフレームのクラスフレームは、キーワードの数だけ作成する。教訓フレームのクラスフレームは、教訓の数だけ作成する。

の範囲で重みづける。

- (b) 現象、原因から教訓への連想を重視する。また、適合教訓がある場合、連想度を高くする。

- (c) 下位語から上位語への連想は、検索範囲を広げることになるので連想度を下げる。

- (d) 特定のキーワードに連想が集中する場合、連想度が高くなりすぎ、そのキーワードだけでは上位に検索されないよう連想度を下げる。

(2) 教訓の関連度設定基準

- (a) 教訓に対するキーワードの関連の強さを〔0, 1〕の範囲で重みづける。

- (b) 教訓の主題に着目しキーワードの主従を判断する。1教訓当たり主キーワード2個、従キーワード4個を目安とする。

- (c) 特定の教訓が集中して選ばれないように、各教訓が持つ関連度の総和のばらつきを少なくする。1教訓当たり3.5~4を目安とする。

- (d) 一つのキーワードが多数の教訓に関連する場合、同順位とならないよう関連度の差を明確につける。

以上の手順で構築した本システムの知識ベースの構成を図6に示す。システムの規模は、ルール数864、フレーム数569、手続き1,900ステップ数である。

6 活用方法と効果

本システムは、ソフトウェア開発支援ツールとして、次に述べる活用方法がある。

(1) プログラム作成時のレビューでの活用

初心者はプログラム作成中に経験者なら当然気づく常識を忘れたり、プログラム作成作業と常識集が結び付かない場合がある。今回、身近にあるワークステーション上で、プログラム作成中に必要なキーワードを入力することによって、初

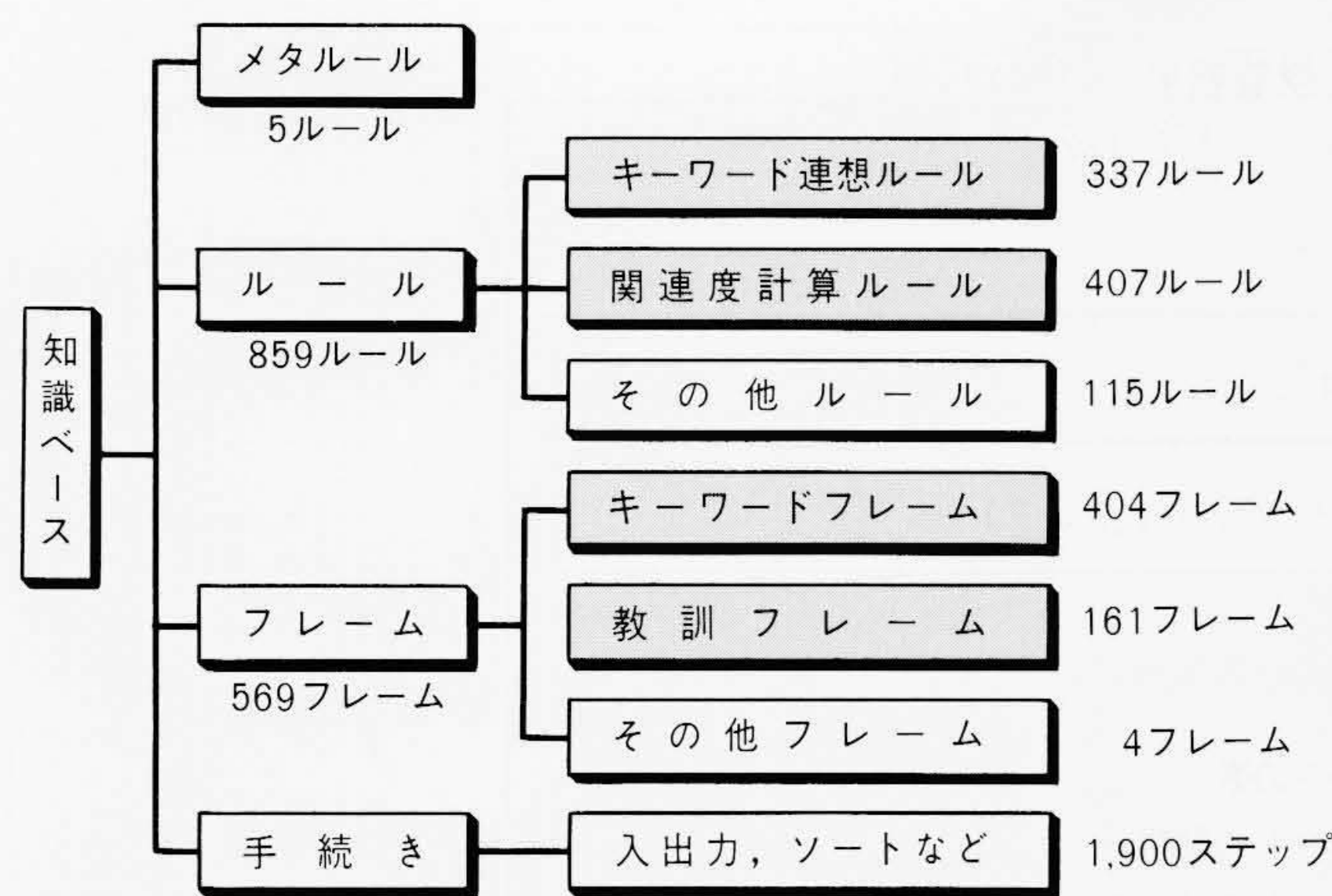


図6 知識ベースの構成 知識表現はルール、フレームなどを用いている。薄い網目の部分が専門的知識であり、他は制御用の知識である。

心者でも経験者と同じレベルで検索可能となった。

(2) 類似事例の検索での活用

ソフトウェア不良が発生した場合、過去に類似事例があるかどうかを見極め、再発防止策を検討する際活用できる。また、よく忘れられる常識、有効な常識が明らかになり、知識ベースの改善も図れる。

(3) 教育での活用

新人教育時に、親しみやすく学ばせることができる。
次に、本システムが知識ベースを利用する効果を、従来のキーワード検索と比較することによって測定した。
5人の初心者(1年ないし2年の経験者)、7人の中級者(3年ないし10年の経験者)、4人の上級者(10年以上の経験者)に対し7個の例題についてキーワードを選択させた。評価結果を表1に示す。

(1) キーワード連想ルールの効果

従来のキーワード検索が、112ケースのうち適切な教訓を検索できない場合が13件あったのに対し、本システムでは4件であった。しかも、その4件は初心者によるものであり、キーワードの選択が適切でなかった。キーワード連想ルールは、キーワード入力漏れによる検索漏れ防止に役立つ。

(2) 関連度計算ルールの効果

従来のキーワード検索の適合事例件数は平均14.4件であるのに対し、本システムでは、キーワードの追加によって、適合事例件数は平均26.9件と増える。しかし、その中で最も適合する教訓の検索順位は平均2.6位であり、特に上級者では2.3位であった。関連度計算ルールは多くの検索結果の中から、目的とする教訓を早く見付けるのに役立つ。

本システムのルールは、そもそも熟練者の頭の中にだけあったものである。これを知識ベースに格納することによって、多くの人々の共有財産とすることができた。

7 検索形エキスパートシステム構築ツールへの展開

エキスパートシステムの構築を容易にするためには、扱う問題の形を限定したツール(分野別シェル)が有効である⁶⁾。しかし、それぞれの分野別シェルに共通の基盤がないと、扱う問題が複数の形の組み合わせとなる場合、及び分野別シェルの機能を超える場合、利用者の負担が大きくなる。

そこで、はん用的エキスパートシステム構築ツールを用いて、問題の形に応じた知識を分野別シェルに埋め込み、分野別シェルが規定する構造に従って、個々の問題の知識を定義し、この二つの知識を合わせてエキスパートシステムを完成させるアプローチを試みている⁷⁾。このアプローチによると、はん用ツールを利用しているので、異なる分野の知識ベースを併合したり、分野別シェルとは別の機能を容易に追加できるという利点がある。

この考えに基づいて、はん用ツールを用いて開発したソフトウェア常識集知的検索エキスパートシステムから、検索形エキスパートシステム構築ツールを開発した。具体的には、ソフトウェア常識集知的検索エキスパートシステムからソフトウェア常識集に関する専門知識を除くことによって、検索分野に共通する一般知識だけとし、この一般知識に改造を加えはん用化を図った。

7.1 ツールの特徴と適用分野

検索形エキスパートシステム構築ツールは、キーワードによって文書を検索するエキスパートシステムの構築を支援する。本ツールを用いて検索システムを構築する場合、利用者は表2に示すユーザー定義知識だけを作成すればよい。

本ツールの特徴を次に述べる。

(1) 利用者の負担軽減と開発期間の短縮

表1 評価結果 初心者5人、中級者7人及び上級者4人に対して、7事例の検索をキーワード検索及び連想検索で実施し、連想検索による効果を測定した。キーワード検索では14.4件見つけるのに対し、連想検索では26.9件と増えるが、最も適合する教訓は2.6番目に見つけることができる。

被験者の区分	キーワード検索			連想検索			
	平均入力キーワード数	平均適合事例件数	検索不可のケース	平均連想キーワード数	平均適合事例件数	平均検索順位	検索不可のケース
初心者	2.0	11.9	$\frac{9}{35}$	6.0	25.1	2.8	$\frac{4}{35}$
中級者	2.7	16.4	$\frac{3}{49}$	7.6	28.8	2.6	$\frac{0}{49}$
上級者	2.4	14.0	$\frac{1}{28}$	6.7	25.8	2.3	$\frac{0}{28}$
平均	2.4	14.4	$\frac{13}{112}$	6.9	26.9	2.6	$\frac{4}{112}$

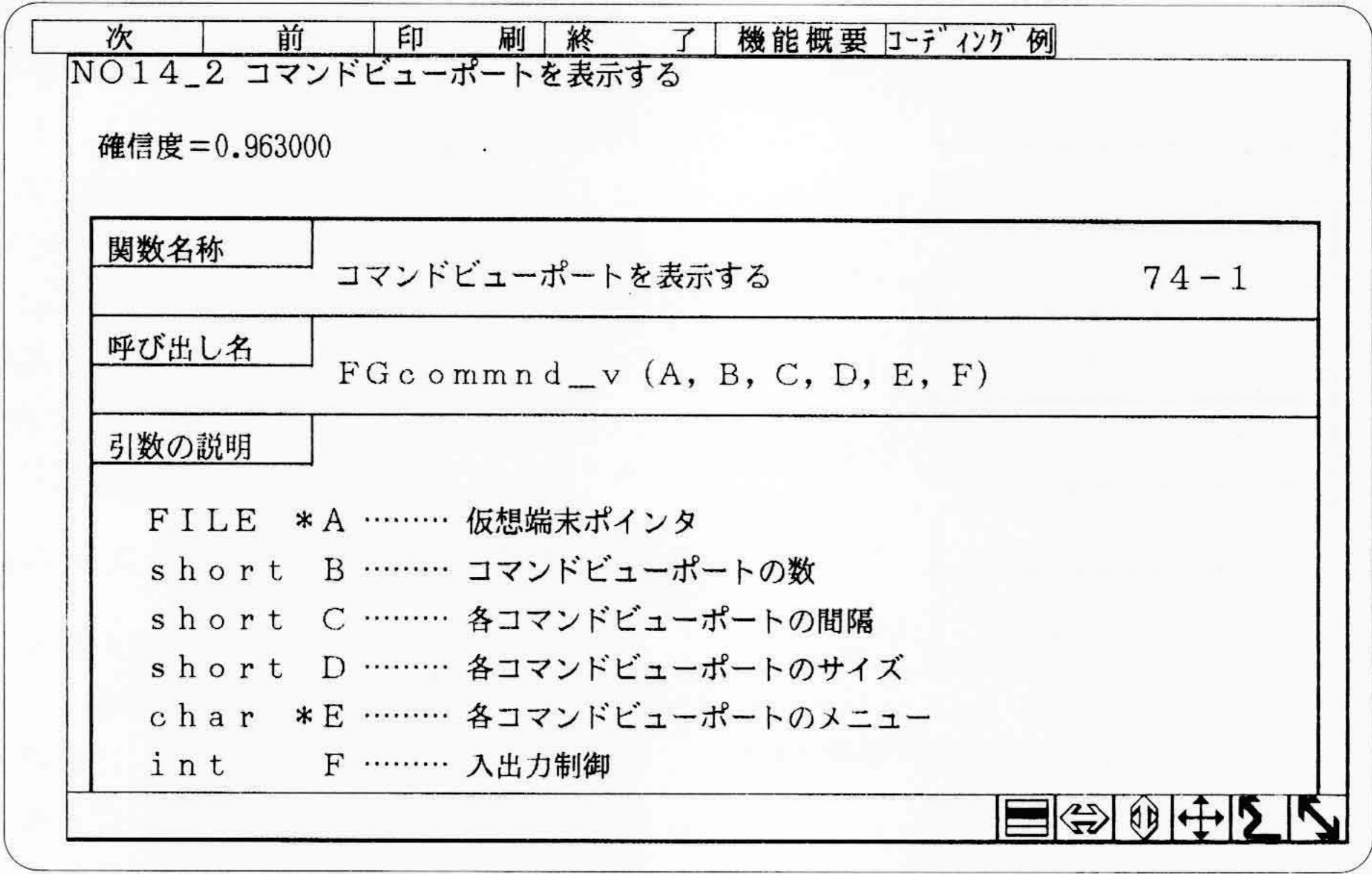


図7 ソフトウェア部品知的検索システムの出力画面例 関数名、引き数の説明のほかに、機能概要やコーディング例などを表示する。

表2 検索形エキスパートシステム構築ツールのユーザー定義知識システム組込み知識とユーザー定義知識を合わせて、実行可能な知識ベースとなる。

項 目	内 容
文 書 フ ァ イ ル	検索対象となる文書の実体
文 書 フ レ ー ム	検索対象となる文書の管理情報
キーワードフレーム	キーワードの属性
キーワード連想ルール	キーワード間の連想関係とその強さ(連想度)
関連度計算ルール	キーワードと文書の関連とその強さ(関連度)

本ツールは、システム組込み知識として検索形エキスパートシステムの基本構造を提供しているので、利用者の負担軽減、開発期間の短縮が図れる。

(2) 連想検索方式による検索精度の向上

本ツールを利用して構築する検索形エキスパートシステムは、ソフトウェア常識集知的検索エキスパートシステムと同様に、知識ベースを持たない検索システムに比較して、検索精度の向上が期待できる。

本ツールは、次に示すようなシステムに適用できる。

(1) 手続きガイダンスシステム

官公庁に提出する書類や社内で規定されている手続きに関して、手続きの漏れがないように、必要な手続きばかりでなく、関連する手続きもガイダンスする。

(2) 失敗事例検索システム

失敗事例を登録しておき、起こりそうな失敗事例を検索して再発防止に役立てる。参考にすべき経験を忘れていた場合及び他人の経験を活用する場合に有効である。

7.2 応 用 例

検索形エキスパート構築ツールを用いて、ソフトウェア部品知的検索システムを試作した。C言語で記述した14分類、74個の共通部品(関数)の部品説明書を検索するシステムである。

システムの規模は、ルール数460、フレーム数380である。開発に要した期間は、共通部品の選定、部品説明書の作成及び文書ファイルへの登録を含めて3箇月である。

本システムの出力画面例を図7に示す。

8 結 言

本稿では、ソフトウェア常識集知的検索エキスパートシステムを開発し、ソフトウェア開発支援ツールとして実用化していることを述べた。検索方式として連想検索方式を考案し、利用者の試用によって、本方式がキーワードの選択漏れによる検索漏れの防止及び検索結果の順位づけによる検索効率の向上に有効であることを確認した。最近、ワードプロセッサなどの普及によって文書検索に対するニーズは高くなってきており、ビジネスの分野でも本方式が広く応用できると考える。このためのツールとして、検索形エキスパートシステム構築ツールも開発した。今後、適用事例の拡大を通して、機能の拡充を図っていく考えである。

参考文献

- 1) 橋本, 外: ソフトウェア常識集IRシステムSOCKS(1)—概要—, 情報処理学会第35回全国大会, 1505~1506(昭62-9)
- 2) 辻, 外: ソフトウェア常識集IRシステムSOCKS(2)—連想検索方式について—, 情報処理学会第35回全国大会, 1507~1508(昭62-9)
- 3) 辻, 外: ノウハウの知識ベース化によるシステム/ソフトウェアの品質管理, 品質, Vol.18, No.2, 40~47(昭63-4)
- 4) 金森, 外: ES/KERNELでの知識表現方法と高速推論方式, 日立評論, 69, 3, 219~224(昭62-3)
- 5) 辻, 外: メタ知識定義による知識ベースの保守方式とその適用例, 人工知能学会誌, Vol.3, No.3, 319~328(昭63-5)
- 6) T.Bylander, et al.: CSRL-A Language for Expert System for Diagnosis, Proc. of IJCAI(1983)
- 7) 安信, 外: 検索型エキスパートシステムとその基本構造に関する一考察, 人工知能学会第2回全国大会, 341~344(昭63-7)