

## ヘルプデスクにおける類似文書検索システムの構成と機能について

辻 洋 間瀬久雄 津原進 衣川一久\*

(株)日立製作所 システム開発研究所 (株)日立電子サービス\*

顧客からの問合せとそれに対する回答を文書データベースとして蓄積し、新たな案件が生じたときに類似案件を検索することにより、迅速かつ的確な顧客サポートサービスを目指すシステムをWWW上で開発した。日本語で記述された問合せ内容とデータベースに格納されたの事例文書とに現れるキーワードのマッチング頻度をベースとした類似度計算を行い、値の高いものから順に出力する。新たな文書データ登録時にも類似事例の抽出により、既存事例と重複しないかをチェックすることが可能である。また、検索されたログをとることによりある期間一度も検索されない事例を抽出し不要ではないかと警告したり、利用された場合に新たな制限などのコメントを付加することにより事例の内容を洗練化して後日利用する人を支援できるようにするなどユーザインタフェースについて配慮したので、その検討内容について報告する。

## Similarity-Based Case Document Retrieval for Help Desk System

Hiroshi Tsuji, Hisao Mase, Susumu Tuhara and Kazuhisa Kinugawa\*

Systems Development Laboratory, Hitachi, Ltd. and Hitachi Electronics Services Co., Ltd.\*

This paper describes a help desk system on WWW which retrieves relevant cases to customer's new problem from document database. The input of the system as trouble shooting report is described in natural language and the system uses keyword-matching algorithm which calculates the similarity among new problem and cases in database. Our system implies whether new case can be added in database or not based on the same algorithm. There are also a couple of features for usability such as log analysis function for case maintenance and annotation function for retrieved case.

## 1. はじめに

一般にサービスを営む場合、顧客からの問い合わせがあったときに速やかに適切な回答をすることが大切である。この業務を支援するシステムは、ヘルプデスクとか窓口システムと呼ばれ、いろいろなものが知られている。従来は、電話でその問い合わせを受けることが一般的であったが、ファックスや電子メールで問い合わせがきたり、さらには直接WWWを用いて顧客自らが診断をすることも予想される。

多くの問い合わせは、再現性があるため、問い合わせ内容とその回答事例をデータベースに蓄積しておき、それを検索しようというアプローチをとることが多い。たとえば、顧客の問い合わせ内容からオペレータが適切なキーワードを選択し情報検索することがある。このとき、単なるキーワード検索では、ユーザは1個あるいは高々数個しかいれないため、ノイズが多く現れたり漏れが生じたりという問題がある [1]。また、問診タイプで問い合わせ内容を絞り込んでいこうというアプローチもある [2]。これは、絞り込みのためにステップが長くなりすぎることが多いという問題がある。最近では、事例集を文書データベースとして蓄積しておき、それに対して類似文書を判定してランク付けして出力しようというアプローチもある [3]。さらには、顧客にはなるべく独力で情報検索させ、必要に応じて遠隔のオペレータが支援するというコンセプトの基づくシステム [4] もある。

本論では、事例を文書データとして蓄積しておき、日本語で記述された新たな問い合わせがあったときに瞬時にその問い合わせに一番マッチングした事例を検索することによりオペレータを支援することを考える。類似事例の抽出には、先に報告した文書分類ツール F L U T E (Filtering Lens software system for Unclassified Texts) [5] を応用した。以下では、類似判定のアルゴリズム自体には深く言及せず、システムとして具備すべき機能と実現システ

ムの構成について議論する。

## 2. 問題の記述

はじめに想定する業務モデルについて記述する。

- (1) 顧客から問合せを受け付け、問い合わせ内容を記録する。
- (2) 必要に応じ、類似事例がないかを検索する。
- (3) 検索結果の中に今回の回答たりうるものがあるか否かを判断する。
- (4) 回答を先の質問に追記して一つの事例とする。
- (5) 顧客への回答終了後、その事例を蓄積しておくかどうかを判断する。

このモデルに基づくシステムを構築するには、次の課題がある。

- (1) 問い合わせをなるべく多くのチャンネルで受け付けるようにするには基盤部分をどのようにしておけばよいか？
- (2) 検索時にともすれば大量のものが出力され、また逆に全く検索されないこともあるが、これを平滑化するにはどうすればよいか。
- (3) 事例をどのように保守していけばよいか。重複したものが入り込まないようにしたり、または事例の過不足を判定できないか。

## 3. 提案するシステムの構成

前章の課題を解決すべく、図3. 1に示すシステム構成を提案する。提案システムは、従来システムに比べて次の特徴を実現することを目標としている。

- (1) 日本語で記述された問題記述文による検索

これによりオペレータは問い合わせを受けたときに、特に検索のためのキーワードを選定する必要がなくなるし、また、問題をもったユーザ自身が、問い合わせ文を作成する事も可能となる。さらに、メールなどから送られた質問文を入力として回答をメールで返信することも期待できる。

## (2) 類似事例のランキング提示

類似性についてある評価基準を設け、その類似度の高いものからリスト表示することによりユーザの選択を支援する。これによりキーワード入力型の場合に指摘されている「大量に結果が得られたり、逆に全く得られないことがある」という問題を回避する事ができる。また、検索された文書群のなかのキーワード分布をユーザに表示して、絞り込み検索を促進することができる。

## (3) 事例利用の履歴の取得

オペレータが利用した事例の履歴をとっておき、オペレータはそれに対して「事例の一部制約を超えて有用であった」「再利用には事例に記載されていない制限があった」などのコメントを付加しておけるようにする。ある一定期間全く参照されない事例は有用ではないと判定したり、よく参照される事例は有用であるとみなす。また、新しい事例を登録する際に、既に登録されている事例を対象として類似検索を行い、内容的に同一の事例が存在する場合、その新事例の登録をしないようにすることにより、事例DBを洗練化できる。

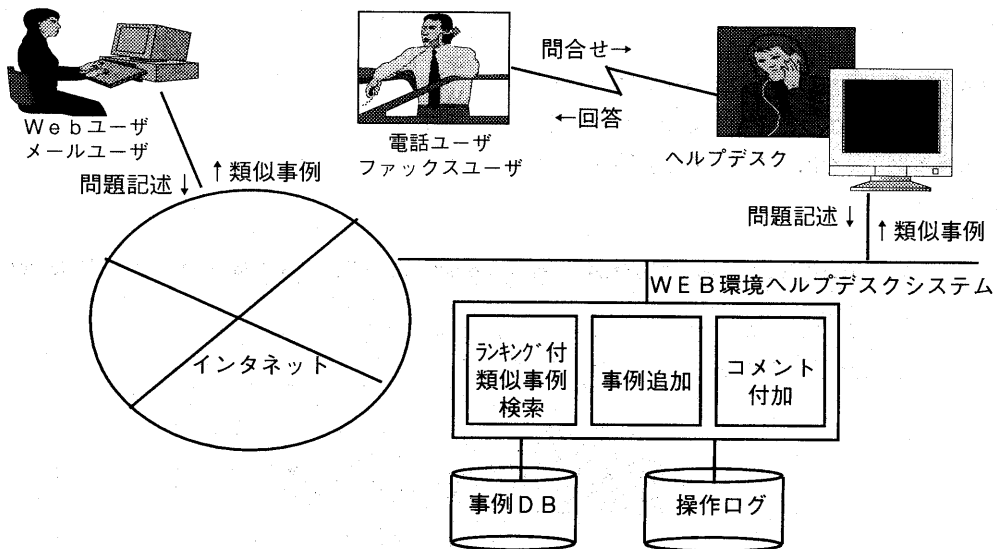


図3.1 提案システムの構成

## 4. プロトタイプシステムの開発

今回、ソフトウェアの障害をコンサルタントする業務を例に取り、図3.1に対応したプロトタイプを開発した。本システムは、WWWブラウザから新規発生障害を日本語で記述された文章を入力とし、過去の障害事例から同一、または関連すると計算されたものを検索、表示する。検索結果は、類似度の高いものから順に複数の事例が概要と共に表示され、一つの事例が指定されると、その詳細を表示する。

システムの入出力については、次のように設計した。

### (1) 事例データフォーマット

検索対象となる事例DBに格納する情報は以下の通りとする。これらをHTML (HyperText Markup Language) で記述した。

[タイトル] [問題内容] [処理内容] [回答内容] [問合せユーザ名] [利用物] [製品名] [機種 (ノード名)] [OS名] [NOS名]

データ件数は、1万件～2万件程度を想定して、複数ユーザの同時アクセスもありうるが、その頻度はさほど多くないという前提をおき、CGIプログラムで実現した。

### (2) データ入力

新規発生障害内容を文章で記述する。図4.1に本プロトタイプの入力後の画面例を示す。問題を記述した文章は普通の自然言語で入力することができるが、キーワードを羅列するだけでも良い。文章が長いほど処理時間がかかるが、一般に精度の良い結果を得ることができる。

### (3) データ出力

検索の結果、同一、または関連する事例と計算されたものについて、あらかじめ指定された件数だけ検索結果一覧を出力する。表示の順序に次の2通りを設けた。

・段階表示：ヒットした事例をタイトル他で一覧表示し、一覧から詳細にリンクする。

・直接表示：ヒットした事例の確信度の高い上位数件は詳細表示。それ以降を段階表示。

また、検索結果の事例について、どのキーワードがヒットしたのかを各事例について列挙表示する。検索者の思惑と、実際の検索結果（過程）とのズレを補正するための助けとなる。図4.2に類似事例検索後の画面例（直接表示）を示す。上位5件については、文章内容詳細も表示している。以降の事例はタイトルのみの表示とし、全体を少ないスペースで鳥瞰できるよう配慮している。

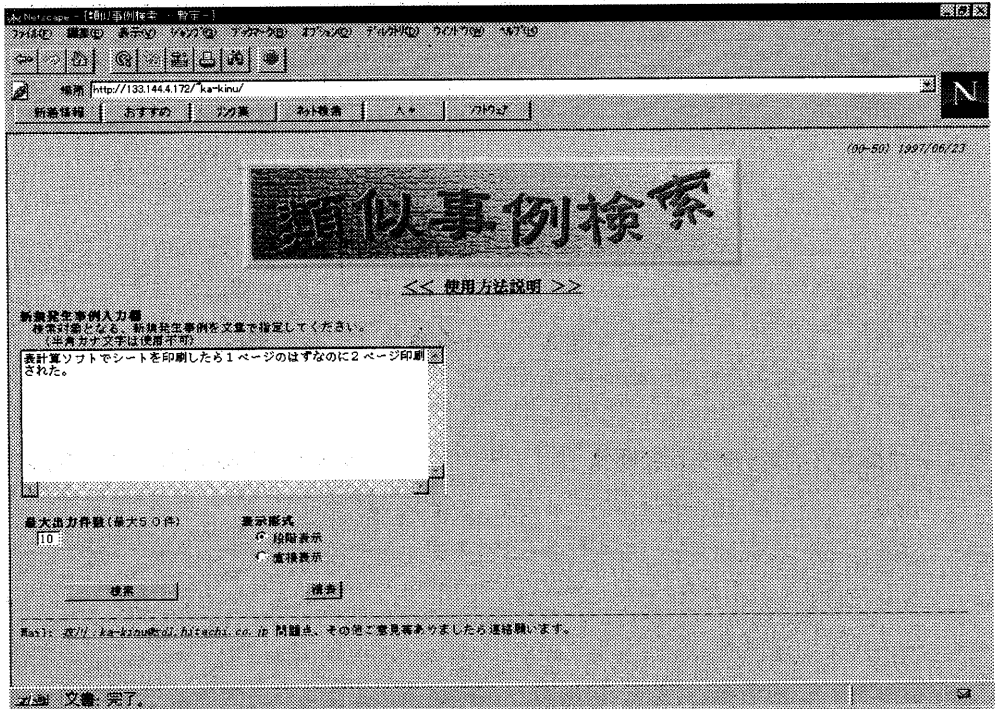


図4.1 事例入力後の画面例

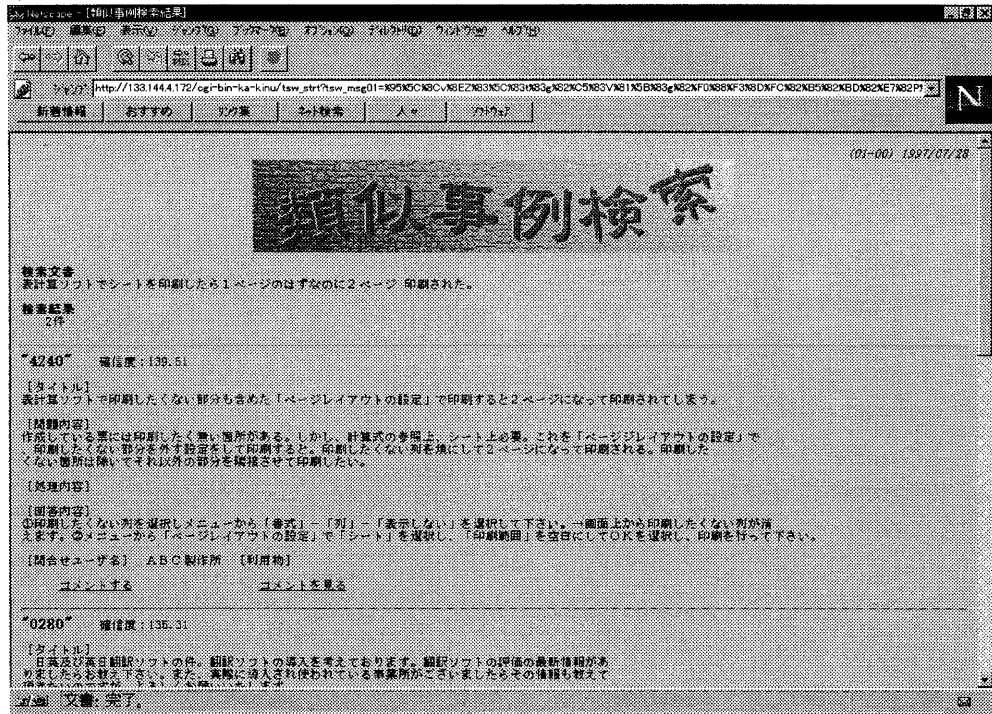


図 4. 2 類似事例検索後の画面例 (直接表示)

#### (4) 類似事例検索エンジン

FLUTEと呼ぶ、テキスト文書を予め定義しておいたカテゴリに自動分類 [6] [7] するツールを用いた。ここでは、一つのカテゴリを一つの事例とみなし、「テキストを関連するカテゴリに分類する」ことを、「テキストを関連する事例に対応付ける」ことに対応させることにより、FLUTEを類似事例検索エンジンとして応用した。

FLUTEは、各カテゴリ (ここでは各事例) を特徴付ける重み付きキーワードを分類済みのデータから自動抽出して知識ベースに格納する。ここでは、事例DBに格納されているすべての事例についてそれぞれキーワードを頻度付きで抽出し、知識ベースに「あるキーワードがあらわれ

たら、各事例に何点 (頻度分析から求めた得点 [8]) 与える」格納する。表記のゆれや同義語の扱いに付いてはすべてFLUTEが対応する。

新規事例の類似事例を検索する際には、新規事例から同様の方法でキーワードを抽出し、知識ベースのキーワードとのマッチングを行うことにより各事例毎に類似度 (得点の和) を算出し、類似度の高い事例から順に出力する。この処理概要を図4. 3に示す。少なくとも事例ベースに格納された文書と同じ問題を入力するとその事例が1位で検索される。格納されていない新規問題についての類似性判定精度については、現在評価中であり別途報告する。

なお、検索結果として上位の事例が使えなかった場合は、新たな問題とみなすことができ、その

回答が得られた時点で新たな事例を追加する。上位に使える事例があった場合には、その事例を残すか、それを削除して新しい問題を事例として置換するか、事例とあらたな問題を一般化して事例とするかが考えられるが、そのいずれの場合も重複事例がデータベースに入り込む事を防げると考える。

(5) 利用者のコメント入力

検索時にヒットした事例について、利用者がその有効性、補足事項等を自由にコメントとして追加できるようにした。入力/参照方法は、

(a) 詳細表示画面に“コメント入力”ページへのリンクを設定し、これを指定することによりコ

メント入力画面を表示する。

(b) 入力したコメントは事例データと関連付けて保存し、一覧表示、詳細表示時に各事例単位に設定する“コメントを見る”を指定することにより参照する。

(c) 既にコメントが入力済の場合にコメント入力を指定した場合、既存のものに追加する(入力済のコメントは、一般ユーザには編集/削除できない)。

である。

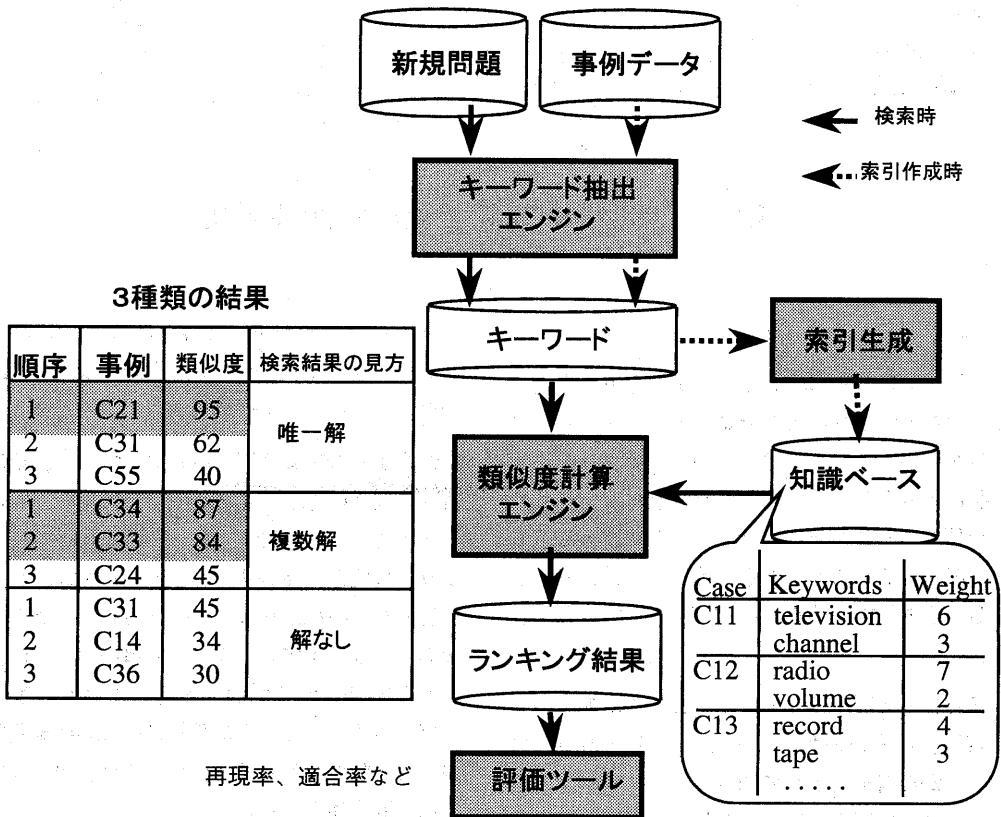


図4.3 FLUTEの処理概要

コメントを付加するメリットは、登録済障害事例入力者の認識誤り、対策ミス（偶然解決しただけで原因は別である等）の訂正、またよりよい対策ノウハウ、関連事項等、登録者以外の大勢の知識やノウハウを共有できることにある。さらに、特定の障害に対する障害事例が、別件の同一障害で得られたコメントにより、より一般的な事例となることが期待できよう。その一方、コメントが本文（または他のコメント）を否定する場合、利用者はどちらを信じるべきか等の問題が生じるかもしれないが、これは試行結果からのフィードバック時に対処を考えたい

その他の特徴として、システム利用状態のログ（利用者、参照事例等）取得がある。実際の利用者数、利用の多い部署等を知ることにより、性能評価、チューニング等の参考になる。ヒット率、参照数の多い事例等を分析することにより、発生障害、登録事例の傾向を知る。また、長期間ヒットの無い事例について、メンテナンス時の削除対象抽出の参考にできる。

## 5. プロトタイプ of 拡張に関する考察

以下に、プロトタイプを基盤として拡張可能な特徴について述べる。

### (1) 電子メールによる検索依頼と結果取得

プロトタイプでは全ての操作を Web ブラウザで行うが、検索依頼/結果取得を電子メールにより行うことも可能である。この場合、検索依頼/結果取得共に電子メールで行うケースと、検索依頼のみ Web で行い、結果を電子メールで受け取る（この場合 Web への結果表示と併用も可とする）方法が考えられる。これにより、Web の環境が無い場合でもサービスが利用できる。結果を後日再考したい（自分に転送）、また第三者に見せたい場合等に対応できると考える。先に述べたコメント機能は、メールによるサービス利用を適用する場合、参照のみ指定可能となろう。

### (2) 過去に入力した障害内容記述文の再利用

過去(前回等)に入力した検索対象文章を再利用できるように、単位時間(例えば一日等)に入力のあった発生障害の文章を参照、利用できるようにする。方法としては、過去に入力された文章を掲示板的に一覧表示し、この中から任意の文章を指定することにより、検索の再実行、または当該文章を検索画面のデフォルトデータとして設定、表示するなどが考えられる。これにより、操作を完了したあとに再び同一問題を解決したい、類似する文章を一部変更で流用したい場合、全文入力し直す手間を省ける。

### (3) 検索精度向上に固定キーワードを設定

問題記述の文章以外に特定のキーワードを指定する欄を設け（プルダウンメニュー的な多岐選択も可）、指定されたキーを検索の絞り込みに利用する。指定されたキーワードを必須とみなすことにより問題によってはノイズを大幅に削減でき、その副作用として漏れも削減できると考える。

### (4) 表記のゆれの解消、同義語辞書登録の支援

自然言語処理でキーワードマッチングする場合精度を上げるには、さけて通れない道である。辞書に登録しておけば解消できるが、どのように辞書に登録するかがいつまでも残ろう。運用の状況を見て、どこまで配慮すべきかをつめていきたい。

## 6. おわりに

顧客からの問合せ事例を蓄積して有効活用し、迅速かつ的確な顧客サポートサービスを目指す事例検索システムのプロトタイプを WWW 上で開発した。自然言語処理に基づいて問合せ内容と類似する過去の事例を類似度を測り値の高い事例から順に出力することにより回答情報の収集を促進する。

本システムは、従来システムに比べて次の効果がある。

### (1) 類似事例のランキング提示による検索効率向上

(2) 利用者のコメント付加による逐次事例ベース洗練化

(3) 類似事例なしの確認に基づく事例追加

今後は、現場のオペレータの協力を得ながら精度・性能評価を行うとともにログ解析結果のフィードバックによる検索効率化を図っていく。

for Automatic Patent Categorization, Proceedings of Advances in Production Management Systems, 377-382 (1996-11)

[8] G. Salton : Automatic Text Processing, Addison Wesley (1988)

### 謝 辞

本研究の機会と、本研究への貴重な御意見、御助言を頂いた(株)日立電子サービスの関係者の方々に感謝致します。

### 参考文献

[1]加藤寛次、他：大規模文書データベース検索用テキストサーチマシンの開発、情報処理学会シンポジウム予稿集(1991)

[2] Ackerman, M. and Malone, T. AnswerGarden. A tool for growing organizational memory. In Lochovsky, F. H. (ed.) COIS-90: Proceedings of the Conference on Office Information Systems, 1990 ACM Press, Cambridge, MA. (1990)

[3] C. Yasunobu, et al. : Document Retrieval Expert System Shell with Worksheet-based Knowledge acquisition facility, Proceedings of IEEE COMPSAC, pp.278-pp.285(1989)

[4] 辻 洋、他：ネットワーク時代の情報流通サービス、I E R e v i e w, Vol.37, No.3, PP39-44 (1996)

[5] 間瀬 久雄他：テキスト分類支援ツール FLUTE の開発 (1) (2) 情報処理学会第 52 回全国大会講演論文集 (第 3 分冊, 3-303~306, 1996.3)

[6] 森本 由起子他：新聞記事自動分類システム構築の検討と評価：情報処理学会第 53 回全国大会講演論文集 (第 3 分冊, 3-205, 1996.9)

[7] H. Mase, et al. : Experimental Simulation