

2 単語連結表現を用いた検索方式の提案

北川結香子*, 国分利直*, 桑原恒夫*, 玉城幹介*

多くの企業は、顧客からの製品やサービスに問合せや苦情に対応するヘルプデスクを設けている。ヘルプデスクのオペレータは、幅の広い内容の問合せに対し迅速かつ的確な対応を求められている。そこでオペレータの顧客対応を支援する環境として、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、新たな事例が生じたときに過去の類似事例を検索することが可能なシステムの開発を進めている。本論文では、このような類似事例検索システムにおいて、煩雑な検索式を入力することなく、多くの事例から検索者の求める事例を抽出するために、事例に対する 2 単語連結表現によるインデキシング方式およびその検索アルゴリズムの提案を行なう。

A method of 2-term linkage format for Q-A retrieval system

Yukako KITAGAWA**, Toshinao KOKUBUN**, Tsuneo KUWABARA**, Motonori TAMAKI**

Many companies have Help desk to answer various kinds of questions or claims from their customers. Service operators must answer them rapidly and correctly. We developed a help desk system which retrieves relevant cases to customer's new problem from document database. This paper describes an index style of 2-term linkage format and the retrieval algorithm in which it is used. As a result, the retrieval only of a similar case is achieved by an easy input as keywords by the operator.

1. はじめに

1.1. ヘルプデスクにおける顧客満足度

本研究では、コールセンタのヘルプデスク部門での利用を想定した類似事例の検索システムの検討を行っている。

ヘルプデスク部門とは、コールセンタの中で、製品やサービスへの問合せや苦情に対応する専門部署である。企業にとって、製品やサービスを新規購入してもらうだけでなく、購入に対する満足度を高め、再び自社製品の購入へと結びつけること、製品や企業のブランド力を向上し、維持することは重要な活動である。ヘルプデスクにおいて、顧客の満足度を向上させるためには、顧客をオペレータ間で転送や、回答に時間を掛けることなく迅速に回答すること、オペレータによって回答の違いが生じないことが必要である。

しかし、近年の提供するサービスや商品の高機能・多機能化、および顧客の利用環境や要望の多様化から、各オペレータの持つ経験、

知識だけでは迅速かつ正確に回答することが困難になっている。その一方で問い合わせ件数中その 6 割が過去に類似した問合せがあったとの報告もある[1]。

そこで、顧客からの問合せとそれに対する回答を応答ログとして蓄積・共有し、新たな事例が生じた際には過去の類似した事例を検索し再利用することが有効であると考えられる。実際に、検索システムがヘルプデスクでの類似事例再利用の目的で活用されている。

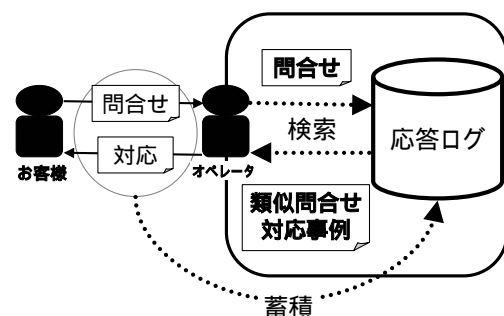


図 1 類似事例蓄積 / 検索システム

* 日本電信電話株式会社 NTT アクセスサービスシステム研究所

** NTT Access Service Systems Laboratories, NTT Corporation

1.2. 検索システムの基本モデル

検索システムの基本モデルとその性能指標について述べる。

検索システムは、検索者の検索要求とシステムに蓄積した情報(蓄積情報)との照合を行い、最も検索者の要求を満たすとみなされる情報を提示する。

しかし、検索要求、および蓄積情報を自然言語のまま計算機で照合することは困難であるので、各々を計算機で扱える内部表現に変換し照合が行われる。一般に、蓄積情報の内部表現をインデックス、検索要求の内部表現を検索式と呼んでいる(図2)[2]。

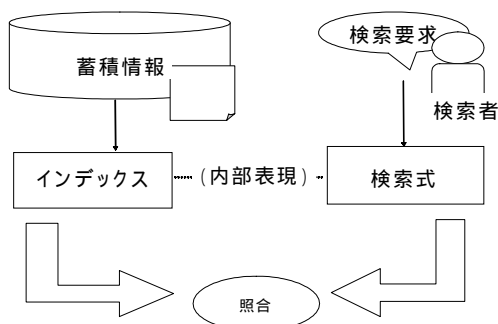


図2 検索システムの基本モデル

1.3. 検索結果の評価基準

検索システムの検索結果の指標には再現率と適合率がよく用いられる。

全検索対象文書を母集合としたときに、検索要求に一致する全文書(図3:A)と、実際に検索された文書(図3:C)の重なる部分(図3:B)が、検索に成功した文書となる。

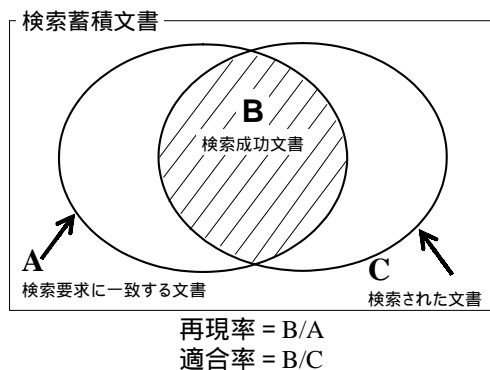


図3 再現率と適合率

再現率とは、蓄積情報に存在する検索要求に一致する文書(図3:A)のうち検索された文書(図3:B)を表す指標である。ヘルプデスクにおいては、全応答ログの中でオペレータの検索要求に一致する全事例のうち検索できた事例の割合を表す。

一方適合率とは、検索された事例(図3:C)の中で検索者の検索要求と一致した文書(図3:B)の割合を表す指標である。ヘルプデスクにおいては、検索結果として表示された事例のうち、オペレータが顧客の問合せに回答するために利用できる事例の割合を表す。

検索システムでは、この再現率と適合率ともに向上させることを目標としている。一般に適合率が情報をいかに絞り込んだかを示す指標であり、再現率がいかに関連した情報を広く出すかを表す指標であるため、相反する関係となる(図4)。

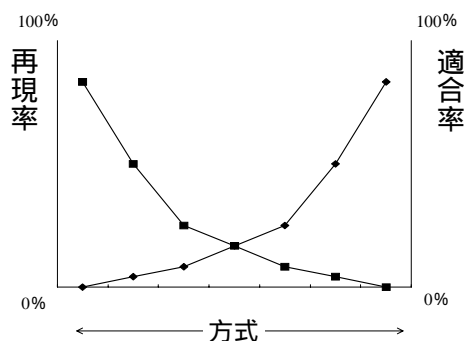


図4 再現率と適合率の関係

本事例検索システムは、オペレータが類似事例を参照して顧客への対応を補助することが目的である。

そこで、まず出来るだけ顧客に一致する事例のみを検索すること、つまり適合率の高い検索を行う。その時、完全に一致する情報が得られない場合には、検索する事例の対象を広げ類似する可能性がある事例を広く検索すること、つまり再現率の高い検索へと拡張する。

1.4. 従来の検索システム例

次にヘルプデスクで利用されてきた検索システムの類型とその課題について述べる。

1.4.1. 診断型エキスパートモデル

診断型エキスパートモデルは、検索者がシステムの提示する質問に対して順次回答していくことで、目的とする情報を検索するシステムである。応答ログの各事例が質問により分岐する決定木で表現されており、検索者は適切に回答することで類似する事例に到達することができる。しかし、応答ログの情報が多量である場合や逐次増加する場合には、その決定木を矛盾なく構築することが困難である。仮に、構築できたとしても、類似事例にたどり着くまでの質問階層が深くなり、絞込みに時間がかかることが問題となる[3]。

1.4.2. 全文検索(単語検索)

単語による全文検索では、検索者が探したい情報を表すと考える単語を入力し、システムは全蓄積情報の中からその単語を含む情報を検索する。

多くの全文検索では蓄積情報のインデックスをシステムにより自動的に生成する。システムによりインデックスが自動生成されるため情報の蓄積には比較的負担が少ないこと、検索者も単語を入力するだけで利用が簡単という利点がある。

しかし、システムは情報の意味や意図を理解してインデキシング処理するわけではないため、インデックス自体が適切でない問題がある。

また、検索者が入力した単語群による指定では、同じ単語を含むが問合せには関連しない情報も出力されるなど、出力結果を絞り込むために適切な検索式を作成することが困難であるという問題もある。

従って、ヘルプデスクにおいて求められるような、短時間に、顧客の問合せにできるだけ類似する事例を絞り込んで検索を行うことは困難であると考えられる。

上記の検討結果を踏まえ、従来の検索技術の課題を解決するために、本研究では2単語連結表現を用いたインデキシングと検索アルゴリズムの提案を行う(図5)。

	インデックス		検索式	
	蓄積	維持	絞込み	自由度
エキスパート	×	×		×
全文検索			×	
提案手法				

図5 検索システム比較表

2. 知識表現形式の提案

2.1. 人手によるインデックス

本研究では、インデックスに情報の意味内容を的確に反映したインデックスを付与するために人手によるインデキシングを行う。

従来の自動インデキシングの技術では、文書の意味を理解して処理を行うのではなく、単語の出現回数や頻度から重要と思われる単語を抜き出す。このため、文書の本質的な意味や意図を完全に抽出することはできない。

本研究で対象とするのは、ヘルプデスクにおける対応ログであり、再利用されることを目的として蓄積される情報である。そこで、事例の蓄積の際に人手によりインデックスを付与する方式を採用する。

人手によるインデキシングは従来のヘルプデスクにはなかった作業であり、そのコスト増加が問題となる。これに対しては、インデックス付与作業をシステム支援することで作業増加を抑制することを考えている。

2.2. 「2単語連結表現」と「主部/補足部表現」

本節では、関連する単語を連結することにより意味内容をより限定して表現する「2単語連結表現」と、検索式やインデックスの中での重要性を表す「主部/補足部表現」の二つの構造化を提案する。

本研究では、事例に付与するインデックスおよび、検索式に「2単語連結表現」および「主部/補足部表現」形式を用いる。形式に従った入力を行うためには訓練が必要であるが、本研究での検索者はヘルプデスクのオペレータを想定しており、入力のための訓練を行うことが可能であると考えている。

2.2.1.2 単語連結表現

情報の意図や意味を表すために顧客の問合せにできるだけ一致する事例のみを素早く検索するために、関連する語を連結させてひとつの単位とする。

2 単語連結表現：(単語₁ - 単語₂)

事例の内容を示す単語を、修飾 - 被修飾の関係が成り立つ組で取り出す。

このように2単語の修飾 - 被修飾関係を利用することによってより詳細な内容を表すことができ、検索事例を絞り込むことができると考える。

下記に、単語の組合せによる表現と2単語連結表現の例を示す。例文はそれぞれ異なる主題を持っているが、単語の組合せではほとんど同じ表現となり例文の主題を十分に示すことができない。一方、提案する2単語連結表現では同じ単語数であっても、各例文を異なる表現により区別することが可能であり、より意味内容を詳細に表すことができる。

例文 1) 英語教育におけるコンピュータの利用
単語表現：

英語，教育，コンピュータ，利用

連結表現：

(コンピュータ - 利用)，(英語 - 教育)

例文 2) コンピュータ教育に関する英語の利用
単語表現：

コンピュータ，教育，英語，利用

連結表現：

(コンピュータ - 教育)，(英語 - 利用)

例 3) 教育における英語やコンピュータの利用
単語表現：

教育，英語，コンピュータ，利用

連結表現：

(英語 - 利用)，(コンピュータ - 利用)，
(教育 - 利用)

2.2.2. 主部 / 補足部表現

検索式およびインデックスにおいて、元の検索意図や事例における主題とそれ以外を識

別するために「主部 / 補足部表現」を提案する。

主部 / 補足部表現：

主部	要望，問題症状
補足部	顧客の環境情報

主部 / 補足部はそれぞれ前節で提案した2単語連結表現により表現される。

主部 / 補足部の区別と検索時の照合方法について述べる。

これまでの予備検討において、ヘルプデスクの問合せは、What 型、How 型、Symptom 型に分類できると考えている。

What 型は、商品やサービスの名称などを尋ねるものである。例えば「とはなんですか？」などの形をとり、回答には問われた語句の情報があればよく、ひとつの「2単語連結表現」のみで表現可能であると考えられる。

How 型 顧客が要望実現の手段を求めるもので「××するためにはどうしたらよいですか？」などの形をとる。

Symptom 型 故障、動作が不安定などの問題症状を訴え、その対処法を求めるもので「ができない」などの形をとる。

これら How 型、Symptom 型の問合せに対しては回答のために、顧客の要望や解決すべき問題症状だけでなく、顧客の使用機器・サービスやその設定などの環境情報も必要となることが多い。

ただし、応答ログにはその事例が蓄積された時点で明らかであった情報のみが記述されており完全な情報ばかりとは限らない。逆に、本質的でない情報が含まれていて制約が厳しすぎる情報である可能性もある。同様に、検索式の元となる顧客の申告も、情報が欠けていたり、事象とは直接関係のない環境や設定情報や、間違った情報が含まれていたりする。

そのため、応答ログに蓄積された事例の中から顧客の問合せと完全に一致しその事例を元に顧客への回答が可能な事例のほかにも、完全に一致はしないが顧客の問合せと関連する事例であり回答の参考になる事例があると考えられる。

そこでこのような How 型、Symptom 型的事例、問合せでは要望や問題症状といった主題を表す部分を主部、顧客の環境情報の部分を

補足部とに分けて表現する。

検索時には、「主部/補足部」ともに一致する事例を最優先で表示する。もし完全に一致する事例が得られない場合には、少なくとも問い合わせや問題症状を表す主部が一致するが補足部には違いがある事例を表示する。

このように「主部/補足部表現」の一致度や補足部における相違を考慮した照合をおこなうことで、参考事例が存在する可能性、応答ログの不完全性、情報の間違いや過不足といった不適切な検索式に対し、柔軟に検索することが可能となる。

2.3. 多インデックス方式

インデックスおよび検索式生成での個人間の揺らぎから生じる再現率の低下を防ぐため一事例に対して複数のインデックスを割り当てる、「一事例・多インデックス方式」を提案する(図6)。

通常検索システムでは、一つの文書に対して一つのインデックスを割り当てる(図7)。なお、全文検索のインデックスも複数の単語で検索できるため、多インデックスのように見えるが、実際には、文中における各単語の出現頻度の組によりインデックス化されているため、一事例には一つのインデックス(組)が付与されている。

しかし、応答ログにインデックスを付与したり、検索要求から検索式を生成したりする作業には各人の視点や背景知識が影響するため、生成されるインデックスや検索式には揺らぎが生じる。その結果、検索者の検索式とインデックスが一致しない場合には、検索者の本来の要求に適合する事例であっても検索されず、結果再現率が低下する。

そこで、一つの事例に対して複数のインデックスを許すことにより再現率の低下を防ぐことを狙う。そして、予備実験等の結果から見てもこの揺らぎは索引者、検索者によりある程度のパターン化するものと考えられる。特に、本研究ではヘルプデスクにおけるオペレータがインデックスの付与、検索を行うことを想定していることから、多インデックス化で懸念される適合率の極端な低下は生じないと考える。多インデックスの作成方法については次章で述べる。

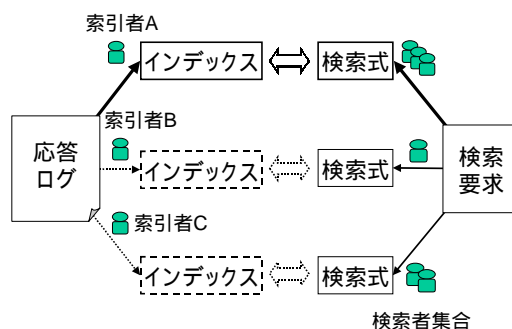


図6 一事例/多インデキシング

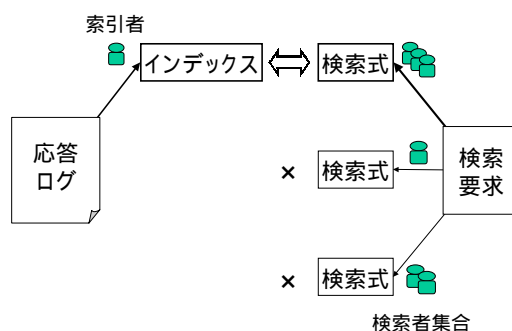


図7 一事例/一インデキシング

3. 類似事例検索システム

3.1. システム全体構成図

本研究で検討を進めているヘルプデスク向けの検索システムの構成図を図8に示す。2章で提案した技術が検索時、蓄積時に用いられる。

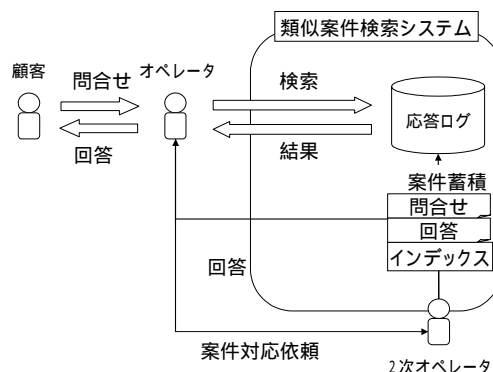


図8 システム構成図

下記に本システムの利用フローを示す。

顧客からの問合せを受け付ける
「2語連結表現」と「主部/補足部表現」により検索式を投入
検索結果確認
事例を参考にし、顧客へ回答

ただし、検索結果に類似事例が見つからなかった場合は、新規事例とみなし、

二次オペレータへ対応依頼
二次オペレータによる回答作成、蓄積
受付オペレータへの蓄積通知

のあと、 の顧客の回答を行う。

3.2. 検索-フィードバック アルゴリズム

本研究では、「2語連結表現」、「主部/補足部表現」を用いてインデックスおよび検索式を表現することで、事例や検索要求の意味内容をよりの確に表現できると考えており、その結果検索性能が向上すると考えている。しかし、インデックスおよび検索式が複雑になるほど完全に一致することが少なくなる。

そこで、インデックスと検索式の照合時には、検索者が入力した検索式とは一致するインデックスが存在しなかった場合に動的に検索式を拡張し再検索を行うアルゴリズムを提案する。また、検索結果に対する検索者の評価によりインデックスを補正し検索システムの性能向上を狙うフィードバック機能も提案する。

検索アルゴリズムとインデックスに対するフィードバック機能について説明する(図 9)。

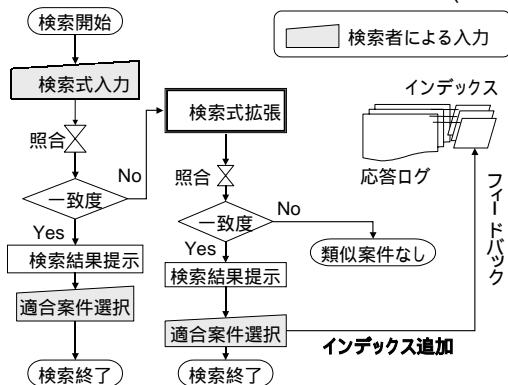


図 9 照合・フィードバックの流れ

オペレータが顧客の問合せをもとに検索

式を「2語連結表現」、「主部/補足部表現」を用いて生成し、検索を行う。応答ログに付与されたインデックスと検索式を照合する。検索式に一致するインデックスがあった場合には、類似事例として提示する。検索者は一致する事例および参考となる事例を選択しその結果をシステムに入力する。

ただし、照合では一致する事例が一件も得られなかった場合にも参考事例が存在するケースが考えられる。そこで、検索者が入力した検索式に一致するインデックスが存在しなかった場合には下記アルゴリズムにて式の拡張を行う。

検索者が入力した検索式を拡張し、その拡張式を用いて再検索を行う。式の拡張には検索者が入力した検索式に含まれる単語による And / Or 検索式などへの展開が考えられる。再検索の結果オペレータの要求と一致した事例があれば、最初に投入した検索式を該当事例の新たなインデックスとして付与する(フィードバック機能)。

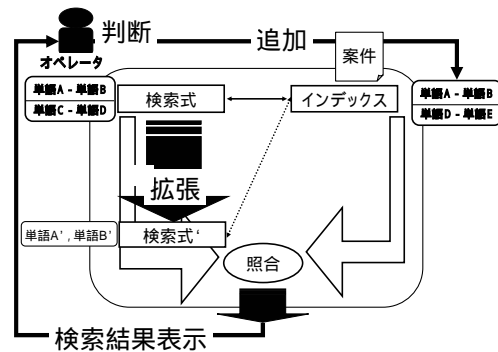


図 10 検索式の拡張とインデックス追加

以上のアルゴリズムにより、インデックス付与時には想定されなかった視点からのインデックスを追加していくことで、徐々に検索性能を向上していくことが期待できる。

4. おわりに

顧客からの問合せとそれに対するオペレータの回答の組からなる応答ログを蓄積し、新たな問合せに類似する事例を検索し、参照す

ることによってオペレータの迅速かつ確な対応を目指す類似事例検索システムの研究を進めている。

本論文では、事例に対して「2 単語連結表現」および「主部/補足部表現」による構造化されたインデックス形式、およびその検索アルゴリズムを提案した。さらに、各オペレータにより生じるインデックスおよび検索式の揺らぎを吸収し、再現率を向上させるために一事例に対し複数のインデックスを付与する多インデックス方式と、それを実現するフィードバック法の提案も行った。

これらの提案により、検索式の入力の手間を抑えながら、より類似した事例のみを検索することが可能であると考えられる。

今後、プロトタイプシステムを用いた被験者実験による提案手法の評価を進めていく計画である。

5. 参考文献

- [1] パデュー大学(2002):リサーチハイライト「顧客満足度を左右する要因分析」, 月間コンピュータテレフォニー No36, pp. 80.
- [2] P. イングベルセン(1994):「情報検索研究 認知的アプローチ」, トッパン.
- [3] 辻洋, 大川剛直(2000), 「経営情報処理のための知識情報処理技術」, コロナ社.
- [4] 小船, 津高, 小中, 有田, 橋田(2001): 「意味的情報検索」人口知能学会研究会資料, SIG-J-A101-6.
- [5] J. ヴァーレイス編(1993):「情報の要求と探求」, 勁草書房.
- [6] 北川, 桑原, 玉城, 国分(2002):「連結2語形式による問題インデキシング方式の提案」, 信学会全国大会.