

## 事例データベースを利用した質問応答システムの構築

5V-6

佐藤理, 太田竜男, 佐藤雪乃, 松本桂子, 田澤和彦  
富士通株式会社 情報化推進部

## 1 はじめに

一企業で扱う商品は、年々、多様で高度なものになっている。これに伴い、顧客等に対する QA サポート業務においても、広範な専門知識が要求されるようになった。しかも、企業活動の拡大に伴って、質問件数は増加する一方である。

しかし、電話とサポート要員の知識にたよった従来の作業方法では、回答品質、処理能力ともに限界にきている。そこで、我々は、まず社内の SE に対する技術サポートのインフラとして、質問応答システムを構築し、93年12月より本格的に運用を開始した。

本稿では、その概略を報告する。

## 2 システム概要

日々発生する質問の中には、過去の QA と同じ内容のものも多い。この点は従来から着目され、QA 事例のデータベース化が進められていたが、サポート要員が補助的にこれを利用するにとどまっていた。質問応答システムは、QA 事例データベースをシステムの中核に位置付け、利用者の質問に対して半自動的に回答を行なうシステムである。

質問応答システムにおける自動回答および QA 事例蓄積の方法は、以下のように、事例ベース推論 [1] の基本的枠組に沿ったものである。

- 利用者の質問と類似する過去の QA 事例を自動的に検索し、回答として提示する。
- 利用者が回答を不適切と判断した場合には、QA サポート部門へ質問を転送することができ、それに対して、サポート要員が回答を行う。
- 新規内容の QA は、そのつど QA 事例として蓄積し、以降の自動回答に再利用する。

利用者との間の情報伝達手段として、近年急速に普及している UNIX(TCP/IP) ネットワークを採用すること

A Q&A System using Case Database  
Osamu Sato, Tatsuo Ohta, Yukino Satoh, Keiko Matsumoto,  
Kazuhiko Tazawa  
Information Service & Support Department, Fujitsu Ltd.

により、QA サポートの窓口業務を自動化できる。

そこで、我々は、以下の4つのコンポーネントからなるネットワーク上のクライアント・サーバー・システムとして質問応答システムを構築した(図1)。

- QA クライアント  
質問の送信や回答の受信を行う利用者用クライアント
- QA サーバー  
QA 事例の検索等、利用者への各種サービスを提供する質問応答システムの中核となるサーバー
- CB サーバー  
QA 事例データベースを管理するサーバー
- CB マネージャー  
回答の入力や QA 事例の編集等を行うサポート要員用クライアント

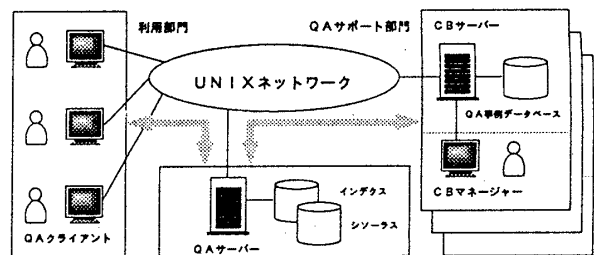


図1: システム構成

## 3 データ形式

質問や回答など、質問応答システムの中を流れる文書データはすべて SGML[2] を用いて記述されている。これは次のような理由による。

- 文書構造が明記されているため、必要部分の取り出しや形式変換が容易である。
- テキストであるため、可読性、拡張性に優れている。

図2は、利用者が入力する質問文書とその内容に類似する QA 事例を例示したものである。

## 質問文書

```
<!DOCTYPE 質問 SYSTEM "QUESTION.DTD">
<質問>
<ユーザID> 6800370
<カテゴリ> FMシリーズ
<サブカテゴリ1> FMRシリーズ
<サブカテゴリ2> MS-DOS
<サブカテゴリ3> NetWare
<質問概要> OASYS文書の共有方法は?
<質問文>
NetWare上でOASYS文書を共有する方法を教えてください。
</質問>
```

## QA事例(確信度71%)

```
<!DOCTYPE QA事例 SYSTEM "CDB.DTD">
<QA事例>
<QA事例番号> 1302
<編集日時> 1994/02/08 火 14:05
<編集者> CB MANAGER
<カテゴリ> FMシリーズ
<サブカテゴリ1> FMRシリーズ
<サブカテゴリ2> MS-DOS
<サブカテゴリ3> 日本語MS-DOS
<質問概要> OASYS文書を共有する方法は?
<質問文>
OASYS文書の文書共有は出来ますか。
<回答文>
NetWareのサーバ上の書類を共用して、OASYS文書の文書
共有が出来ます。また、文書共有は本ソフトウェアの他に、OASYS専用機(PCサ
ーバ連携)、OASYS/Win、IFOS統合文書処理システム等でアクセスできま
す。
</QA事例>
```

図 2: 文書データの例

## 4 QA事例検索

質問応答システムでは、利用者の質問と類似の内容を持つ QA 事例を検索しなければならない。このためには、単純なキーワード検索では不十分である。そこで、我々は、質問文書全体を検索キーとして QA 事例を検索する類似文書検索技術(以下、キードキュメント検索と呼ぶ)を開発した。

キードキュメント検索は QA サーバーで行われる。QA 事例本体は各 QA サポート部門ごとに CB サーバーで分散管理されているが、検索用のインデックスおよびソーラスは QA サーバーで集中管理されている。このため、CB サーバーにアクセスすることなく高速な検索が可能である。

キードキュメント検索の手順は以下のとおりである。

質問文書は、まず、キーワード集合に変換される。キーワード集合とは、質問文書から抽出されたキーワードとその関連キーワードの集合である。各キーワードには重要度に応じて重み付けがなされている。キーワード集合を生成する際には、文書構成要素ごとに異なるキーワード集合生成規則が適用される。重要度やキーワード集合生成規則は質問文書の DTD[2] に各文書構成要素の属性として定義されており、変更が容易である。

キーワード集合生成規則は、パラフレーズング、自動キーワード抽出、関連語展開等の組み合わせとして定義されたものである。パラフレーズングとは、質問表現を

他の表現に置き換えることであり、あらかじめ決められた表現を用いて質問文書の内容を要約した“質問概要”に適用される。関連語展開は、上位語、下位語、類義語、異表記(同義語)の範囲で行われる。関連語展開用のソーラスはキーワード辞書も兼ねており、これを用いて自動キーワード抽出が行われる。

次に、キーワード集合とインデックスとのマッチングがとられ、各 QA 事例の確信度が計算される。確信度とは、その QA 事例が回答としてどのぐらい適切かを表したものであり、インデックスとマッチしたキーワードの重みの合計として算出される。さらに、確信度算出の際には、利用者が質問した日時と QA 事例の編集日時との時間差を用いて、新しい事例ほど確信度が高くなるような配慮がなされている。

最後に、適当な閾値以上の確信度を持つ QA 事例の質問概要のリストが、確信度の高い順にソートされた後、利用者に提示される。

## 5 おわりに

事例データベースを利用した質問応答システムを社内適用することによって、QA サポートの効率化、回答の均質化等の効果が得られている。また、ネットワーク上のクライアント・サーバー・システムとしたことで、QA サポート窓口の一本化や 24 時間サービス等が実現できた。

QA 事例データベースには、運用開始以来、約 1 万件の QA 事例が蓄積され、有効に再利用されている。しかし、QA 事例検索による回答率を向上させるためには、より大量の QA 事例の蓄積が必要である。そこで、社内に流通する各種技術文書から自動的に QA 事例を抽出する技術を研究中 [3] である。また、QA 事例の蓄積件数が増加するにつれて、QA 事例検索の高速化が必要になると予想される。このための並列アルゴリズム等の技術開拓が今後の課題である。

## 参考文献

- [1] 小林重信: 事例ベース推論の現状と展望, 人工知能学会誌, Vol.7, No.4, pp.559-566(1992)
- [2] ISO8879, Standard Generalized Markup Language(1986)
- [3] 斉藤由香梨ほか: 質問応答システムにおけるマニュアルからの技術情報抽出, 情報処理学会第 49 回全国大会 5V-07(1994)