

## World Wide Webを用いたヘルプデスク指向の質問応答システム

A Helpdesk-oriented Question Answering System Using the World Wide Web

三原 英理<sup>†</sup> 藤井 敦<sup>†</sup> 石川 徹也<sup>†</sup>

Eri Mihara Atsushi Fujii Tetsuya Ishikawa

## 1 はじめに

インターネットの普及により、Webには多種多様かつ膨大な情報が存在する。そのため、日常的な疑問に対し、検索エンジンを用いてWeb上の情報から回答を探す機会が増えている。しかし、既存の検索エンジンを使うためには、質問の内容をキーワードで的確に表現する必要がある。また、検索されたページから回答となる語句や文章をユーザ自身が探さなければならない。

自然言語による質問文に対して回答するために、質問応答システムが研究されている。既存の研究は、限定された領域に関する知識や規則によって推論を行うものと、組織化されていない文書集合から回答を抽出するものの2種類に大別される。近年は後者の「抽出型システム」が主に研究されている。

しかし、抽出型システムは、who、when、where、whatを中心とするWH疑問文に対して名詞で回答する「事実回答」が主流である[1]。また、行動を問うhow型質問に対応した研究は、コンピュータの利用法などに領域が限定されている[2]。

本研究は、抽出型の質問応答システムをWebに応用し、既存の研究で対象とされていない「行動を問う質問」に対して、領域を限定することなく回答することを目的とする。すなわち、困った状況にあるユーザに対して、取るべき行動を示唆するヘルプデスク指向の質問応答システムを提案する。

## 2 提案する質問応答システム

## 2.1 概要

本システムは、入力された質問文に関連するWebページを検索し、抽出した回答を分類して提示する。対象とするのは、what-if、what-to-do、how-to-do型の行動を問う質問である。これらの質問に回答するため、名詞句と動詞からなる「行動表現」に着目する。例えば、「蜂に刺されたらどうすればよい?」という質問には「針を抜く」、「流水で洗う」、「軟膏を塗る」などの行動表現と、その行動表現を含む段落を回答する。

システム構成を図1に示す。本システムでは、検索されたページの内容を解析するため、通常のWeb検索よりも応答に時間がかかる。そこで、聞かれそうな質問と

それに対する回答を事前にFAQとしてデータベース化することで、迅速な回答を可能にする。

まず、FAQを構築する事前処理(オフライン処理)について説明する。ここでは「質問抽出」、「検索」、「回答抽出」、「回答組織化」の処理を段階的に行う。

「質問抽出」は、Web上に頻出する質問を抽出する。「検索」は、既存の検索エンジンを用いて、質問に関連するページを網羅的に検索する。「回答抽出」は、検索されたページから行動表現を含む段落を抽出する。「回答組織化」は、含まれる行動表現によって段落を分類する。分類された回答と質問を対にしてFAQデータベースに蓄積する。

次に、ユーザがシステムを利用する場合のオンライン処理について説明する。ユーザがインタフェースに質問文を入力すると、FAQ内に同じ質問がないか検索する。同じ質問があればそれに対応する回答を提示する。同じ質問がなければ、検索、回答抽出、回答組織化を行い、動的に回答を生成して提示する。

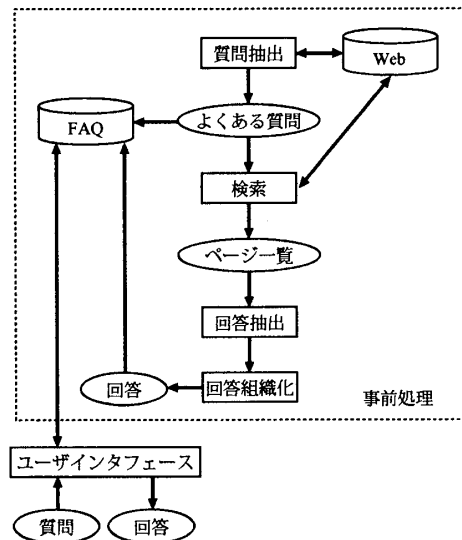


図1: 本研究で提案する質問応答システムの構成

回答は図2のような画面で提示する。これは「蜂に刺されたらどうすればよい?」という質問に対する回答画面例である。図2の「病院に行く」は代表的な行動表現で、その下の「診察を受ける」、「救急車を呼ぶ」などは、それ以外の行動表現である。これらの行動表現を含む段落と、抽出元のページへのリンクを併記し、詳細情

<sup>†</sup>筑波大学大学院図書館情報メディア研究科

報を取得しやすくしている。

すなわち、本システムの「回答」とは、「病院に行く」のような句であり、また同時に、段落のような文章でもある。ユーザは利用環境などによって、行動表現によるグループ一覧やグループ内の段落一覧など、表示内容を選択することができる。

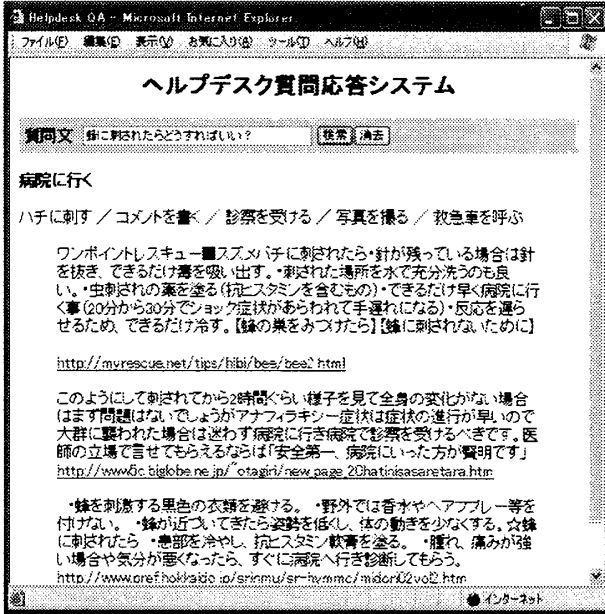


図 2: 本システムの回答画面例

以下 2.2~2.5 節で、質問抽出、検索、回答抽出、回答組織化の各処理について説明する。

## 2.2 質問抽出

Web には、FAQ サイトに用意された質問から、日記やブログに書かれた素朴な疑問まで、様々な質問表現が潜在するため、これらを収集する。

「○○が△△したら、どうする?」、「○○を△△したら?」のような表現を単語や品詞で規則化し、該当するフレーズや文を抽出する。具体的には、名詞の後に助詞(が・に・を)、動詞(～したら)、と続き、さらに疑問符や「(、) どうする」が続く表現を抽出する。

形態素も用いるため、表層表現に一致しない語句も抽出することができる。他方において、不要な表現(ノイズ)が検出される可能性がある。しかし、仮に FAQ にノイズが含まれても、ユーザが同様の質問をしなければ実害はない。そこで、質問抽出では精度(precision)よりも再現率(recall)を重視する。

抽出元データには NTCIR<sup>1</sup> の Web コレクションを使用した。これは JP ドメインから 11,038,720 のページ(約 100GB)を収集したコレクションである。その結果、6961 のフレーズが抽出された。

## 2.3 検索

Web 検索には Google を利用し、入力された質問文をそのままクエリとして用いる。しかし、実験の規模を拡

張するためには、質問応答に適した検索エンジンを独自に開発する必要がある。

## 2.4 回答抽出

回答抽出部は、検索されたページ群から回答となる段落を抽出し、スコア付けを行う。Web には PDF ファイルなども存在する。しかし、本研究では本文のレイアウト解析を行うため HTML ファイルのみを対象とする。

まず、検索された各ページから、一定の長さを持つ段落を抽出する。単語や文の単位で抽出した回答では、行動を説明するための情報が少なく、逆にページ単位では不要な情報が含まれやすいためである。段落の判別には、`<P>` と `<BR>` を用いた。

タグの使い方はページ作成者によって様々であり、タグを用いるだけでは段落の長さにはばらつきが出てしまう。例えば段落を示す `<P>` を改行代わりに用いているページでは、各文が段落となってしまう情報が不足する。逆に、非常に長い段落を読むことはユーザの負担が大きい。そこで、文字数に関する閾値を併用し、120 文字以上 200 文字以下を段落とした。

次に、回答としての適切さによって段落にスコア付けを行う。まず、検索されたページに出現する行動表現ごとにスコアを付け、段落内に含まれる行動表現のスコアを総和する。そこで、行動表現を形式化した。

行動表現について説明する。CaboCha<sup>2</sup>を用いて、検索されたページのテキスト部分について係り受け解析を行う。そこから「名詞+助詞+動詞」という係り受け構造を抽出し、行動表現の最小単位とする。例えば「薬を患部に塗る。」という文の場合は「薬+を+塗る」と「患部+に+塗る」の 2 フレーズが得られる。ここで、「薬を塗る」ならば「薬:塗る」のように助詞を省き、これを行動表現とする。

助詞を省くのは「蜂に人が刺される」と「蜂が人を刺す」のように、格交代により助詞が変化しても「誰が何をどうする」という行動内容は変わらず、同じ行動表現とみなすためである。

本手法は、「薬を患部に塗る」のように複数の名詞句(名詞+助詞)が動詞に係った表現を行動表現として扱うこともできる。しかし、名詞句が複数あった場合「薬を塗る」と「薬を患部に塗る」が別の行動表現とみなされるという問題が生じる。そこで、実験では名詞句をひとつに限定した。

行動表現のスコア付けでは、回答として適切な行動表現を選択するため、次の 5 つの基準を用いる。

a 係り受けの距離が近い。

係り受けの距離とは、名詞句と動詞の間にある形態素数である。日本語の場合、文末に近い要素ほど動詞の意味を限定する性質がある [3]。また、距離が近いほど係り受け解析の誤りが少ないという利点もある。

<sup>1</sup><http://research.nii.ac.jp/ntcir/index-ja.html>

<sup>2</sup><http://chasen.org/~taku/software/cabocho/>

- b 「～すること」、「～しましょう」といった推奨表現と「～してはいけない」といった禁止表現を伴っている。

推奨表現は問題解決に有効な対処法を述べる時に用いられる。禁止表現は行ってはならない対処法を述べる時に用いられるため、推奨表現と同等に有用である。「蜂に刺されたら?」の質問例では「薬を塗りましょう」、「病院へ行くこと」などが推奨表現であり、「アンモニアをかけてはいけない」、「毒袋をつぶさないでください」などが禁止表現である。このような表現を正規表現によって照合し、検出する。

- c 抽出元ページの検索における順位が高い。

抽出元ページとは、行動表現が抽出されたページである。現在、検索には Google を用いているため、抽出元ページの順位は Google の検索順位である。Google は PageRank により、他ページから多くリンクされているページの順位を高くする。そのため、信頼性を部分的に取り入れることができる。なお、このスコアは以下の式で計算する。

$$\frac{\text{検索ページ数} - \text{順位}}{\text{検索ページ数}}$$

- d 質問に含まれる行動表現との距離が近い。

行動表現中の動詞と質問中の動詞の間にある形態素数である。質問に近いほど質問との関連性が強く、対処法として重要な行動表現が出現しやすい。

- e 質問者（ユーザ）がすべき行動である。

質問者が動作主となる行動を取り出すため、格助詞「ガ」によって主語が明示されている行動表現はスコアを下げる。例えば「蜂が巣を守る。」という文であれば動作主は「蜂」である。この文に出現する行動表現「蜂が守る」、「巣を守る」はともに蜂が行う行動表現であり、質問者への回答として適切でない。

行動表現  $x$  のスコア  $s(x)$  は、上記の基準を式 (1) によって結合し、計算する。実際には、同一の行動表現  $x$  が複数回出現するため、これらを  $x_i (i = 1, 2, \dots)$  とする。

$$s(x) = \sum_i \left\{ \left( \frac{1}{a_{x_i}} + b_{x_i} + c_{x_i} + \frac{1}{d_{x_i}} \right) \cdot e_{x_i} \right\} \quad (1)$$

ここで、 $a_{x_i} \sim e_{x_i}$  は、 $x_i$  に関して上記 a～e で計算された値である。 $a_{x_i}$  は係り受けの距離、 $d_{x_i}$  は質問との距離である。 $b_{x_i}$  は推奨・禁止表現の有無によって 1 か 0 を取る。 $c_{x_i}$  は上記 c で計算される値である。 $e_{x_i}$  は格の有無によって 1 か 0 を取る。

段落  $P$  のスコアは式 (2) により、 $P$  に含まれる行動表現のスコアを総和して計算する。

$$S(P) = \sum_{x \in P} s(x) \quad (2)$$

## 2.5 回答組織化

回答組織化部は、2.4 節の手法で抽出された複数の段落を分類する。ここでは、同じ行動表現を含む段落を同じグループにまとめる。ただし、スコアの低い行動表現はノイズになるので、事前に削除する。現在は経験的に、スコアが 2 以上の行動表現を使用して分類を行っている。ユーザに回答を提示する際に、グループごとに最も高いスコアの段落と行動表現を提示することで、類似した複数の段落を読む手間を省くことができ、どのような回答グループがあるのかを概観しやすくなる。

また、同一グループに含まれる行動表現は依存関係にあり、逆に、同じグループに属さない行動表現は排他的な関係にある。このような行動表現同士の関係は、今後、行動の順序関係や依存関係を抽出するための手がかりになる。

## 3 評価実験

### 3.1 方法

how 型の質問応答システムはほとんど研究されていないため、評価方法が確立されていない。そこで、人手でテストコレクションを作成し、行動表現を回答とした場合における質問応答の精度を評価した。

本システムは、高スコアの行動表現を元に段落をグループ化し、グループ内で最高スコアの行動表現を代表としてユーザに提示するため、回答として適切な行動表現のスコアが他よりも高くなることが重要となる。そこで、行動表現のスコア付けについて、「2.4 節で提案した 5 つの基準が有効であるか」という観点で評価を行った。

質問は、質問抽出で抽出されたフレーズのうち、「パスワードを忘れたら」など、頻度が高い 40 フレーズを使用した。各質問に対して、Google で検索された上位 500 ページを使用した。

これらの質問に対して、抽出された行動表現の適否を人手で判定した。さらに、適切な表現については、禁止表現かどうかも判定し、正解として扱った。

スコアによって順位付けた行動表現について、以下に示す精度と再現率を用い、F 値を計算した。

$$\begin{aligned} \text{精度} &= \frac{\text{システムが出力した正解数}}{\text{システムが出力した全行動表現数}} \\ \text{再現率} &= \frac{\text{システムが出力した正解数}}{\text{全行動表現中の正解数}} \\ \text{F 値} &= \frac{2 \times \text{再現率} \times \text{精度}}{\text{再現率} + \text{精度}} \end{aligned}$$

### 3.2 結果と考察

回答数による F 値の推移を調べるため、行動表現を段階的に増やしていき、上位 200 件までの行動表現について F 値を計算した。評価に用いた全質問について、行動表現数を変化させた各点における F 値を計算し、その平均値を用いて比較評価した。

結果を図 3 に示す。グラフの縦軸は F 値、横軸は行動表現数を表す。それぞれの系列は以下の条件で評価した値を用い、全質問の平均を取っている。

- all : 基準 a~e をすべて用いたスコアによる順位付け
- Pa : 基準 a「係り受けの距離」を使用したスコアによる順位付け
- Pb : 基準 b「推奨・禁止表現の有無」を使用したスコアによる順位付け
- Pc : 基準 c「検索における順位」を使用したスコアによる順位付け
- Pd : 基準 d「質問に含まれる行動表現との距離」を使用したスコアによる順位付け
- Pe : 基準 e「ガ格の名詞がかかる行動表現」を使用したスコアによる順位付け
- freq : 検索されたページ群における出現頻度による順位付け

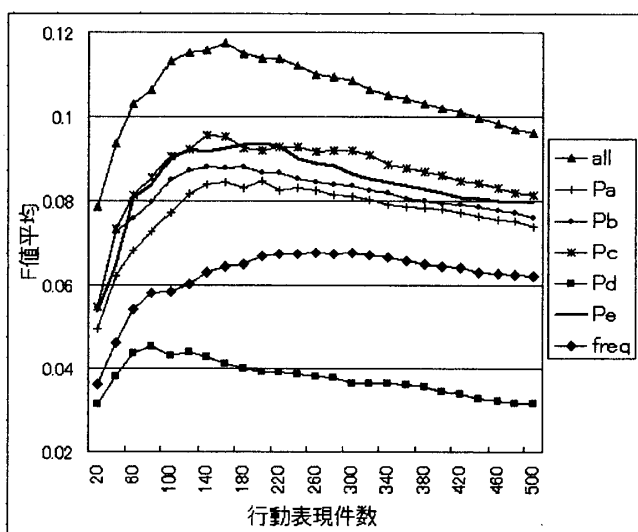


図 3: 行動表現を回答した場合の実験結果

まず、freq とその他を比較すると、基準 d 以外に関しては、いずれかの基準を用いることで結果が良くなっており、基準 d 以外は有効な基準といえる。

次に、all と Pa~Pe を比較すると、出力する行動表現の数によらず、all の方が良い結果であった。すなわち、基準の組み合わせによる相乗効果があった。

また、検索には Google を用いているため、基準 c を用いた Pc は Google の結果に相当する。ここで、all と Pc を比較すると、all の方が良い結果となった。すなわち、本手法は Google よりも正解となる行動表現をより多く、より高順位で検索できることが分かった。

## 4 関連研究

人工知能における質問応答の研究は 1960 年代に遡る。具体的には、エキスパート・システムや対話システムであり、推論によって回答する「推論型システム」である。これらは、システムとユーザとの対話を通して質問・回答を洗練させたり、推論によって複数の事実から回答を導き出したりするという特長がある。しかし、領域知識

や推論規則を手手で構築する必要があるため、回答できる分野が限定され、拡張性に乏しいといった問題がある。

自然言語処理や情報検索では、1990 年代後半から TREC<sup>3</sup> や NTCIR で質問応答の研究が盛んになった。人工知能分野における推論型システムと異なり、抽出型システムが主な研究対象である。組織化されていない文書集合を情報源として用いるため、領域知識や推論規則の構築が不要な点で推論型システムとは異なる。

TREC や NTCIR における質問応答システムの回答対象は客観的事実 (factoid) である。ユーザが入力した、who、when、where、what を中心とする WH 疑問文に対し、名詞で回答する。客観的事実を回答とするシステムに、goo ラボ日本語自然文検索<sup>4</sup> や MURAX [1] がある。MURAX は入力された質問文を解析することで、何についての質問で、回答はどのようなものかという仮説を立てる。その仮説に基づいてクエリ構築や検索を行い、文書集合から該当箇所を探して回答する。

渡辺ら [2] は、行動を問う「how 型質問」への回答を目的とし、知識獲得手法を提案した。linux ユーザのメーリングリストに投稿された質問メール・回答メールの本文から、重要文を抽出することで、方法や対処法を問う質問に回答するための知識を得た。しかし、対象となる領域が限定されている。

それに対して本研究は、領域を限定することなく「how 型質問」に回答する質問応答を実現した。

## 5 おわりに

本手法は、原理的には where-to-do 質問や when-to-do 質問にも応用可能であり、「行動を問う質問」に対して幅広い観点から回答することができる。

今後の研究課題は、行動表現ではなく段落の単位で回答抽出や組織化を評価することである。また、被験者を用いた実用性の評価が必要である。

## 謝辞

本研究の一部は、独立行政法人情報処理推進機構「2004 年度第 2 回未踏ソフトウェア創造事業」からの支援で行われました。

## 参考文献

- [1] Julian Kupiec. MURAX: Finding and Organizing Answers from Text Search. In T.Strzalkowsky, editor, *Natural Language Information Retrieval*, pp. 311-332. Kluwer Academic Publishers, 1999.
- [2] 渡辺靖彦, 園和也, 岡田至弘. メーリングリストを利用した質問応答システムのための知識獲得. 情報処理学会研究報告, 2004-NL-162, pp. 131-138, 2004.
- [3] 徳永健伸, 岩山真, 乾健太郎, 田中穂積. 日本語語順の推定モデルとその応用. 情報処理学会研究報告, 91-NL-81, pp. 9-16, 1991.

<sup>3</sup><http://trec.nist.gov/>

<sup>4</sup><http://labs.nttrd.com/>