

Web を用いたヘルプデスク指向の質問応答システム あなたの疑問に答えます

1. 背景

インターネットの発展に伴い、Web を使って様々な「調べ物」をすることが日常化している。また、生活様式が多様化したことで、本来なら他人に相談するような事をオンラインで調べる機会も増えた。しかし、Web には統制や秩序がないため、欲しい情報がすぐに見つかるとは限らない。情報を選別し、複数の情報を結びつけて何らかの解を見出すといった知的な作業はユーザにゆだねられている。この作業を自動化することで、Web を用いた調べ物の手間を省くことができる。

2. 目的

本プロジェクトの目的は、ユーザが抱えている問題を解決するための「行動」や「手順」を回答する「ヘルプデスク指向の質問応答システム」を実現することである。Web を用いて「蜂に刺されたら？」のような「どうしたらよいか」という質問に回答できるようにする。

3. 開発の内容

機能概要

図 1 に基づき、～ の機能について説明する。

質問抽出

FAQ 構築のため、Web に頻出する質問を収集する。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、形態素解析を用いて、これらを収集する。

Web 検索

入力された質問文をクエリとして、質問と同じ表現、もしくは質問中の語句を含むページを Web から検索する。既存の Web 検索エンジンを使うことができる。本プロジェクトでは、Google を利用した。

回答抽出

質問に関連する重要な行動表現を含む文章を段落の単位で抽出する。各段落に含まれる「行動表現」を用いてスコアを付ける。行動表現とは「名詞 + 助詞 + 動詞」という係り受け構造を抽出し、助詞を省いたものである。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。

回答組織化

回答抽出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。ユーザに回答を提示する際には、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き、どのような回答グループがあるのかを概観しやすくした。

ユーザインタフェース

ユーザの入力した質問が、FAQ 内に存在するかを検索する。存在すれば FAQ にある回答を提示し、なければ、類似する質問を検索して提示する。回答の表示には、回答グループのリスト表示と、Java アプレットによる可視化表示の 2 種類の方法を用意した。

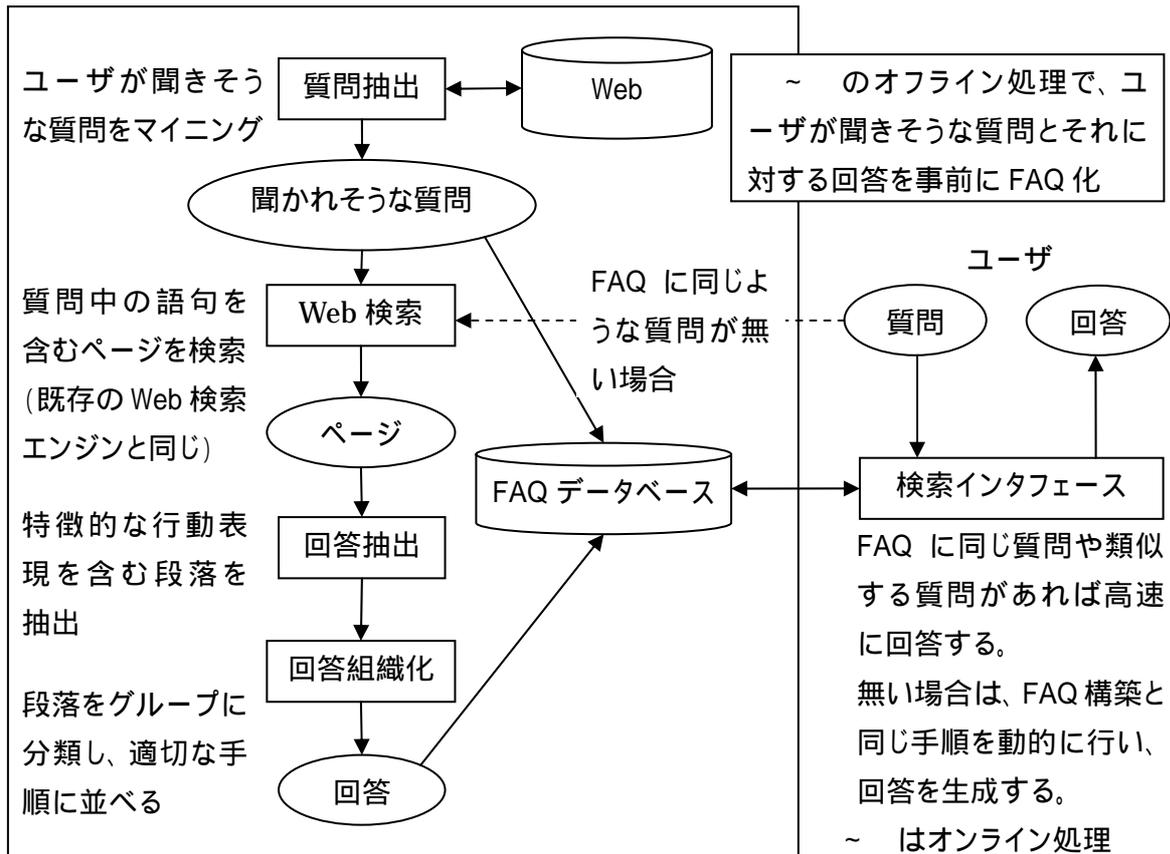


図 1: システム構成

動作環境

WWW サーバ: Linux Red Hat 7x, 8x, 9x (apache, perl 必須)

インタフェースを介した質問検索及び回答閲覧: Web ブラウザ

4. 従来技術(または機能)との相違

質問応答システムに関する従来研究は、who (誰)、when (いつ)、where (どこ)、what (何) など、factoid 質問に回答するものがほとんどである。本システムは、あまり研究が進んでいない「～した場合にどうすればよいか」という how 質問に焦点をあて、回答するものである。

5. 期待される効果

「グローバルブレイン」としての Web 利用

社会の急激な変化に伴って、人々が抱える問題は複雑かつ多様化している。誰に相談すればよいか分からない疑問、人には相談できない悩み、相談相手が身近にいない人(高齢者など)が抱える問題に対して、Web は有用な情報を提供する。しかし、既存の検索エンジンや人手で構築されたポータルサイトでは急速に生産されるコンテンツに追従することができず、有用な情報が埋没してしまっている。本システムは、Web 情報を組織化し、秩序を与えることで、ヘルプデスクやコールセンターを自動的に構築するための布石となることが期待される。

Web マイニングによる問題の予見と予防

Web には新規情報が常に掲載されており、近年は Weblog の登場によってそのスピードが加速している。そこで、社会の動向をいち早く知るための Web マイニング技術が盛んに研究されている。本システムでは、よくある質問や誰かが抱える悩みを Web から発見し、それに対する回答を生成することが目的である。これは、ユーザが将来直面するであろう問題を事前に予見し、かつ問題が発生する前に未然に予防するための危機管理技術としての発展が考えられる。

自然言語理解に関する技術の発展

質問応答は、自然言語理解に関するアプリケーションとして近年注目されている。しかし、質問応答の目的は千差万別であり、単一のモデルで全ての問題を解決することは困難である。想定される質問や回答の形式によって対象の問題を個別にモデル化し、それに対する解決を見出す努力を継続することが重要である。本システムは、質問応答問題の新しいモデルとして位置付けられ、質問応答の枠組みをさらに拡張することが期待できる。

また、質問応答システムというキラーアプリケーションを研究開発する過程で、それに必要な自然言語処理や情報検索の基礎理論と関連技術が発展することも期待できる。

問題解決エンジンとしての発展性

「テキスト全文検索エンジン」は、技術として確立され一定の成果を収めている。そこで、現在では様々な文書に利用できる汎用的なソフトウェアパッケージとして商品化されている。

本研究で開発した成果を技術シーズとすれば、Web に限らず、社内文書、新聞記事、特許情報などの様々な文書群に応用できる汎用的な「問題解決エンジン」として発展する可能性がある。これを実用化することで産業上の波及効果が期待できる。

また、問題解決エンジンを用いて、Web 上のポータルサイトに質問応答機能という付加価値を持たせることで、ユーザや顧客数を増加させる効果がある。

6. 普及(または活用)の見通し

現在、筑波大学の WWW サーバを利用して、デモシステムを公開している。ログの収集や質問抽出の改良により、回答できる質問パターンを増やし、さらに利便化していっそうの普及を図る。現在のオープンメイン(分野を問わない)のまま、一般ユーザのさまざまな質問に回答する、あるいは、質問内容をある製品の使用方法に限り、その製品のユーザを対象に、より高精度な質問応答を提供する、という2方向での展開を検討している。

7. 開発者名(所属)

三原英理(筑波大学大学院 図書館情報メディア研究科)

馬場こづえ(筑波大学大学院 図書館情報メディア研究科)

(参考)デモシステムURL

<http://cyclone.slis.tsukuba.ac.jp/mie/>