

FAQ 生成機能を持つ新しい電子掲示板の提案*

松本 和之

名古屋大学 情報科学研究科 メディア科学専攻
matsumoto@nagao.nuie.nagoya-u.ac.jp

長尾 確

名古屋大学 情報メディア教育センター
nagao@nuie.nagoya-u.ac.jp

1 はじめに

電子掲示板は知識の宝庫であるが、そのままでは再利用が困難である。もし掲示板から自動的に質問・回答を整理して抽出できれば、さらに有益な知識源となることは疑いがない。そこで、従来型の掲示板から質問・回答を抽出し、分かりやすく提示する手法を検討した。しかし従来型の掲示板を用いる場合、質問部分はどこで、要点は何か、といった深い自然言語処理を行う必要がある。それよりも、掲示板が質問・回答を自動抽出できるような仕組みを持つべきである、と考えられる。たとえば、投稿者が投稿する際、質問部分や要点部分等をメタ情報として与えることにより、収集や抽出などの処理が効率的に行えるシステムを目指している。このような手法を用いることによって、質問・回答構造に限らず、応用として様々な意図構造を抽出できると考えられる。そこで、表層的なルールに基づく仕組みの限界と、質問・回答構造を効率よく抽出するために適したメタ情報としてどのようなものがあると適切かを調査する予備実験を行い、その結果を踏まえて、意図構造の抽出が容易な新しい掲示板を試作した。

関連研究にダイアログナビ [1] があるが、これは、あらかじめ抽出された質問・回答集の中から対話的に目的の回答を含むテキストへナビゲートする仕組みである。本研究では、よりピンポイントに質問・回答の組み合わせを抽出し、提示することを試みる。

2 FAQ の自動生成

掲示板の投稿から、質問の意図を端的に表した質問文を作成するとき、投稿中のどの部分を抽出すべきかを考える必要がある。そこで、投稿中の各部分の意図表現を調べ、それに従って取捨選択した。意図表現は、実際の投稿を見ながら分類・定義した。意図表現の定義と決定方法の一例を表 1 に示す。なお [2, 3] では電子メールの談話表現を解析しており、本論文の意図表現に相当するものを「質問内容表現」と呼んでいる。

*A Proposal of a New BBS Capable of FAQ Generation by Kazuyuki Matsumoto (Graduate School of Information Science, Nagoya University) and Katashi Nagao (Center for Information Media Studies, Nagoya University)

表 1: 意図表現と決定ルールの一例

意図表現	決定ルール
挨拶	「こんにちは」「どなたか」等の形態素が全形態素数の 6 割以上
状況	末尾表現が「しまった」「しまうので」等
行為	末尾表現が「した」「したところ」等
質問	末尾表現が「ですか」「でしょうか」等
意思	意思の助動詞「たい」を含む

表 2: 意図表現による取捨選択ルールの一例

意図表現の並び	抽出する部分
質問 any	質問のみ
行為 状況 質問 any	行為・状況・質問の 3 つ
意図 any	意図のみ

各部分に意図表現が付与されたら、質問文を構成するのに必要な部分を抽出する。抽出する部分は、意図表現の並びを元に決定する。抽出部分の決定ルールは、質問投稿の意図表現の並びと、必要と思われる部分とを統計的に分析し、作成した (表 2)。この表からも分かるように、質問の要点は投稿の最初の部分に現れることが多い。

FAQ において、質問は短く端的に表されるのに対し、回答は長く詳しく表される場合が多い。そこで本手法では、回答は 1 投稿中の全文とした。回答投稿としては、質問投稿に直接返信している投稿の中で、質問者以外が返信した、(時間的に) 最初の投稿を選択した。

予備実験で用いた掲示板は、Yahoo! 掲示板における質問専用のスレッドである。すべての投稿には返信関係が付与されている。また、質問専用スレッドを用いることで、スレッド作成者へ直接返信している投稿はすべて質問投稿と見なすことができ、問題を簡単化している。まず、Yahoo! 掲示板の質問スレッド「Windows XP Q&A」の各投稿の本文を句読点や「!」「?」等の記号で区切ってできた各部分に対して意図表現を付与した。さらに、各部分の取捨選択ルールによって FAQ の質問文としてふさわしい質問文を生成した。最後に、各質問文に対する回答投稿を選択し、それらを組にした XML として保存した。得られた結果の画面イメージを、図 1 に示す。

スレッド作成者に対する返信が 652 投稿あり、そのうち 71.2%にあたる 464 投稿に何らかの返信があった。それら

表 3: 生成された FAQ の主観評価

評価	質問		回答
	短さ	正確さ	正確さ
良い	64.5%	73.8%	87.2%
普通	22.0%	5.7%	6.4%
悪い	13.5%	20.6%	6.4%
計	100.0%	100.0%	100.0%

の投稿から質問・回答を作成できた投稿は 76.1%にあたる 353 投稿であった。生成した質問文は、1 投稿あたりの平均が 70.3 文字、文字数は約 5 分の 1 になった。また出力された FAQ のうち、無作為に選んだ 141 投稿に関して評価を行った。評価項目は「質問文の短さ」「質問文の正確さ」「回答投稿の正確さ」であり、主観により 3 段階評価を行った。その結果を表 3 に示す。ただし「回答投稿の正確さ」は、回答の内容ではなく選択として正しいか否かを判定した。特に質問投稿に多かったが、タイトルが本文の一部になっているものが目立った。

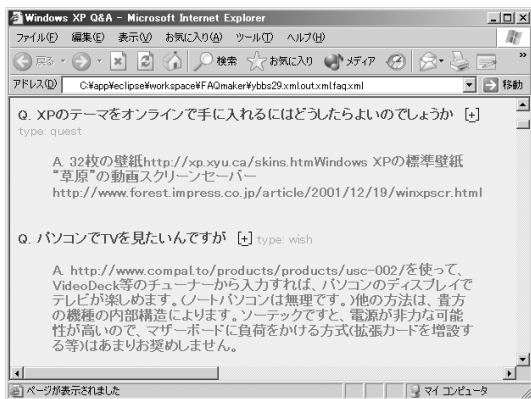


図 1: FAQ の出力結果

3 新しい掲示板の提案

予備実験によって、意図表現に関する単純なルールで可能なことの限界がかなり明らかになった。掲示板のデータをより効率的に、正確に扱うためには、掲示板の仕組みが変わる必要があると考えている。すなわち、投稿にはメタ情報が投稿者によって付与されることによって、掲示板はより知的なコンテンツとなる。そこで前章の実験を踏まえて、投稿時にあらかじめ投稿を構造化して書き込む、質問・回答専用の掲示板を作成した(図 2)。投稿内容は、各意図表現に対応したフレーズ(例えば「意思」なら「こんな風にしたい」等)によって構造化する。また、「これは についての質問」の「 」の部分を入力させることによって、適切な質問投稿のタイトルを自然に入力させようとしている。このように投稿者が投稿を構造化することに

よって、意図表現付与における誤りを大幅に減らすことができ、FAQ 等への再利用が容易になると考えられる。

この新しい掲示板では、実験によって獲得した談話構造パターンを利用している。すなわち、行為表現の後にはそれを受けて状況表現・質問表現が来る、などというパターンに基づいて、投稿者は構造化された投稿を作成する。またこの「行為 状況」などの議論の展開パターンは XML ファイルに記述されているため、パターンの追加・変更は容易にできるほか、質問以外の構造についても拡張しやすい。

将来的には、提案する掲示板によって収集された議論構造つきコーパスを利用することで、従来の非構造化掲示板データからの情報抽出も可能になると考えられる。

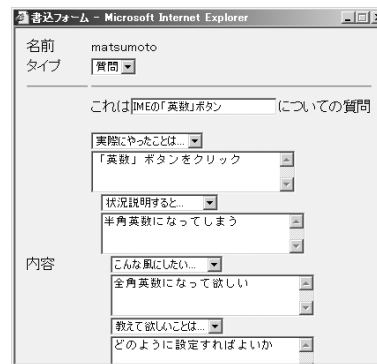


図 2: 作成した電子掲示板の書込フォーム

4 おわりに

電子掲示板の質問スレッドから FAQ を生成する手法について述べた。質問投稿の本文に対して、5 種類の意図表現「挨拶」「状況」「行為」「質問」「意思」を付与し、取捨選択ルールを用いて FAQ の質問文としてふさわしい部分を抽出した。抽出した質問文に対する回答投稿を選択し、それらを組にして表示するようなシステムを実装したところ、質問文は 5 分の 1 程度に短くなったほか、返信があったもののうち 76.1%の投稿は FAQ が作成できた。またこの実験を踏まえて、投稿者が投稿に対して行為・状況などの構造をあらかじめ与えた上で投稿する掲示板を作成した。

今回は質問・回答の構造化を行ったが、同様に、製品等の評価スレッドや雑談スレッドなど多種多様なスレッドの議論構造において、要約・検索などが高い精度で実現可能な仕組みを目指している。

参考文献

- [1] 清田陽司ほか, 大規模テキスト知識ベースに基づく自動質問応答 - ダイアログナビ -, 言語処理学会第 8 回年次大会発表論文集, pp.271-274, 2002.
- [2] 柳瀬隆史ほか, 談話構造を利用した質問応答事例の検索, 言語処理学会第 7 回年次大会発表論文集, pp.30-33, 2001.
- [3] 柳瀬隆史ほか, 質問応答事例検索のための談話構造の認定, NL145-15, pp.97-104, 2001.