

FAQ 生成のためのメーリングリストの QA 構造化

2M-03

三好 右一

東京工業大学 電子機能システム専攻

住田 一男

(株) 東芝研究開発センター

1 はじめに

近年、WWW 上のテキストを自動要約するシステムが注目されている (文献)。本論文では、メーリングリスト等の電子メールコミュニケーションを対象にした要約として FAQ を自動生成するために、その討議スレッドから質問-回答 (QA) の構造を抽出することを試みる。

2 知識協創システム CIKLE

今回の実験を行うために、「まとめ機能付掲示板」を利用した。この掲示板は、従来のスレッド型掲示板に、「まとめ」を作成・投稿できる機能を組み込んだものである。なお、この掲示板はメールによる投稿・購読も可能である。

一つのスレッド内で議論を交しているときとして膨大な投稿量におよび、途中からの参加がしにくくなったり、過去の話題の結論が埋もれてしまったりすることがある。そこで、スレッドで出た結論を、適宜、人手でまとめてスレッド内容を分かりやすく提示できることが特徴となっている。現状 CIKLE が提供しているまとめ作成支援機能は、スレッドの返信部分やシグネチャ部分を自動的に削除する程度である。このため、まとめ作成に手間がかかることが問題となっている。

今回の実験では、まとめ作成の自動化を目的に、スレッド内の応答関係から QA 構造になっているものを選び出し、それを自動的に取得するための手がかりを調べた。また、その QA 構造がどれくらいまとめに取り込まれているかを調べた。

3 実験

東芝で実験運用中の CIKLE におけるコミュニティ「CIKLE ユーザ会」の、図 1 に示すようなスレッド構造から、連続する投稿の関係を分析した (個人が特定できないように個人名とメールアドレスを置き換えた)。なお、「CIKLE ユーザ会」では、CIKLE のユーザがシステムの不具合情報や改善案を開発者とやり取りするコミュニティである。

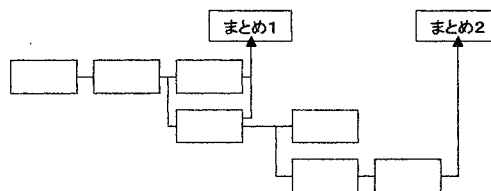


図 1 : スレッド構造

今回対象とした討議スレッドは、まとめ機能付掲示板の作成者とユーザの意見交換スレッドである。図 2・図 3 に例を示す。

Subject: 加入申込をすると確認ダイアログを出すようにしてみました

> でも、申し込んだ瞬間に管理人にメールが送られてしまう
> ので、やはり確認画面を出した方がいいのかもしれない
> ね。

ということで、加入申込みボタンを押すと確認のダイアログが表示されるようにしてみました。どうでしょう？

図 2 : スレッド内の発言 1

 Subject: Re: 加入申込をすると確認ダイアログを出すように
 してみました

> ということで、加入申込みボタンを押すと確認のダイアロ
 > グが表示されるようにしてみました。どうでしょう？
 対応ありがとうございます。よいとおもいます。

図3：発言1に対する返信

この場合、QA 構造であると判定できる明らかな特徴として、「どうでしょう？」が上げられる。この対応関係を、スレッド内の全ての返信 405 個について調べた。その結果を表1に示す。

明らかな質問文	
でしょうか(?)	102
でしょう?	8
ますか	8
教え	4
かね?	3
かもしれません(ね;が)	3
のでは?	3
ご教示ください	2
よね?	2
その他	6
	141
返信の付きやすい文	
欲しい	4
お願い(いた)します	3
がわからない	3
	10
QA構造の数/全応答関係数	171/405

表1：QA 構造の Q 部の特徴

次に、スレッド内で議論が進む過程で随時作成される「まとめ」と、QA 構造の関係について調べた。今回対象としたスレッドでは、合計 51 個のまとめが作成されていた。

51 個のまとめについて、前に選び出した 171 個の QA 構造を利用しているものと、そうでないものについて調べた。結果を表2に示す。

QA構造が直接利用されているまとめ	16
複数の構造を統合して作成されたまとめ	32
QA構造とは無関係に作成されたまとめ	3
まとめの合計数	51

表2：まとめと QA 構造との関係

4 評価

表1について、この結果から、QA 構造の Q 部に明らかな質問文を含む率は 82.4%、返信の付きやすい文も含めた場合は 88.3%であった。

残りについては、内容としては QA 構造になっているものの、自然言語文としての特徴を見いだせなかったものである。

QA 構造と認められる投稿は、特定の語句に多く対応している結果となった。この結果は、特定語を辞書化することで、QA 構造の自動抽出に有効であると解釈される。

次に表2の結果から、QA 構造を元にして作成されたまとめが多いことが分かった。QA 構造と関係なくまとめが作成されているものもあったが、これは今回注目したスレッドのまとめは、全てこの掲示板の作成者らによって投稿されたものであり、自分で気づいた事をそのまままとめにしているものがあつたためと考えられる。また、まとめに利用されなかった QA 構造も多くあつた。

5 まとめ

以上の結果から、システムの作成者とユーザが討議するスレッドを自動要約する場合、質問文の特徴を捉え、QA 構造を抜き出す作業が有効であると解釈できる。

6 おわりに

今後は、QA 構造化、並びに、それを用いたまとめを自動作成させる為に、さらに検出条件を詳細化し、評価実験を行う予定である。

(文献)

- 井佐原 小作 内元「討論型ニュースグループを対象とする知的ニュースリーダーの開発」情報処理学会研究報告 vol97 No53
- 山見 村越 島津 落水「電子メールを利用したコミュニケーションにおける討議スレッド自動抽出法の実装と評価」情処・自然言語 137-10 (2000・6)