

ネットワーク対話の構造解析と視覚化

4ZB-3

久寿居 大

NEC ヒューマンメディア研究所

E-mail: kusui@hml.cl.nec.co.jp

1. はじめに

電子メールやネットニュース、電子掲示板などのネットワークを介した情報交換手段が発達し、膨大な量の情報が流通している。これらの情報交換手段は、ユーザ同士の情報交換などに利用されてきたが、最近ではビジネスへの応用としてマーケット調査や顧客サポート業務への利用が始まっている。

ネットワーク上で交換される情報は膨大であり、再利用は容易でない。例えば、顧客サポート業務やマーケット調査に利用する場合には、目的の情報を素早く見つけ出せることと、探し出した情報の内容を容易に把握できること、の2点が重要になる。

電子掲示板、メール、ニュースなどで交換される情報は、ネットワーク上で行われる対話であり、1) 複数の記事がまとまって一つの情報(スレッド)として意味を持つ、2) 各記事は引用とそれに対する返答という構造を持つ、という特徴がある。文脈に注目して情報の組織化を目的とした研究 [1] もあるが対話内容を素早く把握するための支援も必要と考える。今回試作したネットワーク対話構造化ツール Interaction Viewer では、メールやニュースなどを通じて蓄積された情報を、対話の流れに注目して構造化する。スレッドに注目した検索と、引用関係に基づいた視覚化機能を提供することにより、求める情報の素早い検索と対話の流れの直観的な把握を可能にし、情報の再利用を支援する。

2. Interaction Viewer 概要

2.1. 記事木構造の生成

メールやニュースなどのネットワーク上の対話では、複数の記事がスレッドを構成し、各記事の本文は引用とそれに対する返答という構造を持つ。Interaction Viewer は、各記事の構造を解析し、記事の木構造を生成する。

メールやニュースの記事は、ヘッダ部分と本文部分から構成される。記事の解析はヘッダ部分と本文部分に分けて行われ、ヘッダ部分は Subject: や

Structure Analysis and Visualization of Interaction on Network

Dai KUSUI, Human Media Res. Labs., NEC Corp.

Date:などの識別子に基づいて解析される。本文部分は、各行ごとに引用行、返答行、シグネチャ行の3種に分類される。連続した引用行や返答行は引用部分、返答部分としてまとめられ、引用部分とそれに対する返答部分をまとめてペアが作られる。ペアは0から順に番号が振られる。あいさつなどから書きはじめられた記事など、第0ペアは返答部分だけで引用部分がない場合もある。返答を先に書き引用部分を後にまとめて付ける形式の記事では、第0ペアの引用部分と第1ペアの返答部分は存在せず、第1ペアの引用部分と第0ペアの返答部分を組み合わせるとペアとみなす。生成された引用返答のペアは、対話構造の可視化に利用される。

木構造の生成は記事のヘッダ情報に基づいて行われる。ヘッダ情報に不備がある場合には、記事の引用部分を取り出し、その引用部分を返答部分として持つ記事が検索される。引用部分を持つ記事が親になるように木構造が補完される。

2.2. スレッド単位での検索

メールやニュースをキーワードで検索する場合には、質問記事と回答記事にキーワードが分散して含まれることがある。質問記事に含まれるキーワードで検索した場合には回答記事が検索できず、回答記事に含まれるキーワードで検索した場合には質問記事が検索できない。キーワードのANDをとって検索すると質問回答のどちらの記事も検索できず、ORをとって検索すると余分な記事まで検索される。

スレッド単位での検索は、記事木構造からの部分木の生成、部分木に対する検索、検索された部分木の合成という順で行われる。例えば、図1のように記事1~9で構成される木構造があり、記事1にキーワードA、記事5と9にキーワードBが含まれており、「A and B」という条件で検索した場合を考える。葉にあたる各記事から根にあたる記事までを部分木として取り出し、部分木の各記事を結合したのに対して検索を行う。図1のように条件を満足する部分木を合成し、検索結果として表示する。

個々の検索結果の記事がどのような文脈で発言されたのかは重要である。部分木への分解と検索結果の合成を行うことにより、記事木構造から必要な文脈

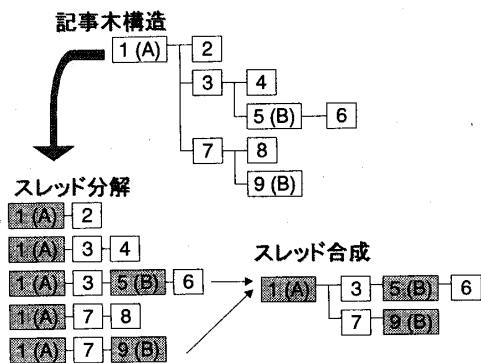


図1 スレッド単位での検索

の部分だけを残し余分な枝を削除できる。

2.3. 引用に基づく対話の構造化

メールやニュースを用いた対話では発言数が増えると、スレッドが巨大になり話の流れを追うのが困難になる。また、求める情報を探す場合にも複数の記事の内容がどういふ文脈で関連しているかを素早く把握できることが重要である。Interaction Viewerは、各記事に含まれる引用部分に基づき、会話の流れを表示するストーリーボードと呼ぶ対話構造の可視化機能を提供する。

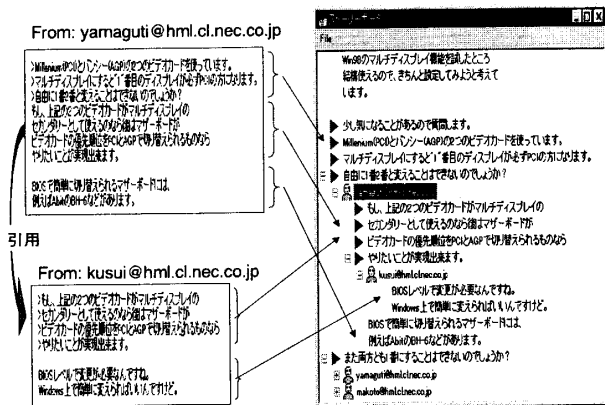


図2 ストーリーボード

ストーリーボードでは、記事本文を解析して作成した引用返答のペアを利用する。記事解析時の引用返答のペアは、記事木構造の親子関係を利用して更新される。他の記事のフォローであり、引用部分を持たない記事の場合には、本文全体を返答部分、フォロー対象記事の本文全体を引用部分とみなしてペアとする。図2のように返答部分の発言者をノードとし、引用部分と返答部分のペアから引用部分が発言者ノードの親、返答部分が発言者ノードの子にな

るように木構造を生成する。記事中のどの部分に誰がどんな返答をしたかを一覧表示でき、対話の流れが容易に把握できる。

3. 利用イメージ

図3に示すようにメインWindowに木構造表示されるタイトル (Subject) や、検索 Window を利用して目的の情報に含まれるスレッドを探す。検索条件には、キーワード、送信時刻、ヘッダ情報を組み合わせて指定でき、検索結果は記事木構造の部分木として表示される。

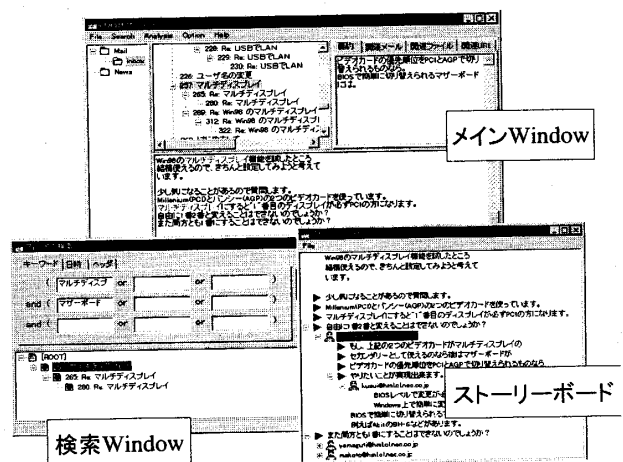


図3 UIイメージ

目的の情報に含まれるスレッドが見つければ、話題の発端を含んだメールを選択し、ストーリーボードを表示する。ストーリーボードによって、スレッド内での会話の流れを木構造で表示し、一つの記事に複数の話題が含まれているような場合でも、メール中の誰がどの部分にどんな返答をしたか、どのように会話がつながるかを素早く把握できる。

4. おわりに

Interaction Viewerは、ネットニュース記事や受信した電子メールに含まれる情報の再利用支援を目的とする。対話の流れに注目し、記事を木構造で表示する。記事に含まれる複数の文脈をストーリーボードと呼ぶ木構造で表示することにより、求める情報の検索と対話の流れの直観的な把握を支援する。

参考文献

[1] 斎藤:MLにおける情報組織化モデルの検討、情報処研報 98-GW-29-9, pp. 49-54, 1998