

計算機上におけるコミュニケーションの 履歴表示に関する研究

松浦 文崇[†] 高田 真吾[†] 中小路 久美代^{†*}
e-mail: {fumita-m, michigan, kumiyo}@is.aist-nara.ac.jp

[†] 奈良先端科学技術大学院大学 情報科学研究科
〒630-0101 奈良県 生駒市 高山町 8916-5
[‡] 科学技術振興事業団さきがけ研究21「情報と知」領域
◇ (株) SRA ソフトウェア工学研究所

概要 電子メールを用いたコミュニケーションの履歴は過去に話し合われた内容を振り返る際に有益である。本論では、メール内の話題単位間の対応関係を明示的に表示することにより会話全体の流れを可視化し、メールのコミュニケーション履歴の利用を支援する方法について提案する。今回提案する方法では、引用マークや改行といったメールの表層的な記述構造に着目してメール内の話題を抽出し、話題間の関係を対応付ける点に特徴がある。

Visualizing the History of Email Communication

Fumitaka Matsuura[†] Shingo Takada[†] Kumiyo Nakakoji^{†*}

[†] Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma-shi, Nara-ken 630-0101
[‡] JST, PRESTO
◇ Software Engineering Laboratory, Software Research Associates, Inc.

Abstract A record of e-mail communication helps us look back at what has been discussed and how decisions have been made. This paper presents our approach to support the use of a record of e-mail communication by visualizing relationships among discussion topics which has transpired over multiple email messages. The approach uses surface structures of email messages, such as quotation marks (">") and line-breaks, to parse messages and identify discussion topics without using any natural language processing techniques.

1 はじめに

電子メールを用いたグループ内のコミュニケーションでは、意思決定などの討議内容をテキストとして記録し、履歴として時系列で保存することが可能である。コミュニケーションの履歴は、

- 参加者が時間が経過し忘れられてしまったような過去に話し合われた項目について振り返る、
- 新たな参加者がそれまでの会話の内容を参照し話の進み具合を理解する、

などの際に有効である。

保存されたメールの履歴を見るためのツールでは、メールのタイトルの一覧を時系列に表示したり、ユーザやタイトルによって並べ替えたりしながら、1通ずつメールのテキストを読む方法が一般的である。しかし、1通のメールには複数の「話題」が触れられていることがしばしばあり、ある「話題」に注目してその流れの履歴を追うためには、各メールを1通ずつ閲覧するのは効率的ではない。

本論では、メール1通単位のメール間の関係ではなく、メール内の話題単位間の対応関係を明示的に表示することにより、会話全体の流れ

を可視化しメールのコミュニケーション履歴を表示させる方法について提案する。今回提案する方法では、話題の内容を自然言語処理により分析するのではなく、引用マークといったメールの表層的な記述構造に着目してメール内の話題を抽出し話題間の関係に対応付ける点に特徴がある。

2 電子メールにおけるコミュニケーションの記述構造の特徴

メイリングリストを介したコミュニケーションは、非同期な協調設計作業において情報のやりとりを通じて共通の理解を漸次的に構築させるための手段であり [1]、また、話の流れの前後関係をお互いに踏まえながら話を進め、情報を漸次的に進化させるためのものである [2]。

2.1 メイリングリストにおけるコミュニケーション

本論では、電子メールにおけるコミュニケーションとして、ある一つの作業に関してある特定の複数ユーザ間で、ある程度の期間（数週間から数ヶ月、数年の単位）にわたって意見の交換をおこなう場合を想定している。メイリングリストにおけるディスカッションはその一例であり、以下メイリングリストにおけるコミュニケーションを想定して議論を進めるが、メイリングリストとして設定していない場合でも複数ユーザ間でメールの交換をおこなう場合も同様の議論が当てはまる。

メイリングリストにおけるコミュニケーションは、

1. 参加者は 1 通のメールに 1 つ以上の「話題」を書き込み、メイリングリストに送信すると、
2. メールを受信した参加者は、そのメールまたは過去に受信したメールを部分的に引用しながら新たなコメントを付け加えメイリングリストに返信し、
3. 1,2 の繰り返しだが、参加者が納得し設定した作業のゴールが達成されるまで続く。

というプロセスとしてモデル化することができる。ここでいう「話題」とは明確に定義できるものではないが、例えば、「次回ミーティングに関して」という Subject タイトルの 1 個のメール中に現れる、「予算について」「場所について」といった大まかなトピックを指す。1 つの話題がコミュニケーションが進むにつれていくつにも分割されることもある。多くの場合、1 かたまりの引用された内容およびそれに続くコメントを 1 つの話題としてみなすことができる。

電子メールでのコミュニケーションは記述上の約束ごとに基づくところが大きく、これらの約束ごとは一般的に確立され、多くのメールソフトにおいて支援されている。たとえば電子メールには、縦横の制約のある空間上にテキストが配置され、横の 1 行の長さは半角 65 文字未満にするといった、読み易さのための約束ごと（ネチケット）が存在する [6]。またある送られたメールに対して返答をする場合、関連する内容を引用して載せその直後に自分のコメントを書くことが多い。この場合、引用した内容と自分のコメントを区別する目的で引用した内容の行頭に引用マークと呼ばれる記号符を付けることが多く、一般に">"や"| "の記号符が用いられる。

2.2 実際のメイリングリストにおけるコミュニケーションの分析

電子メールにおけるコミュニケーションの有効な可視化のためには、メイリングリストにおいてユーザが実際にどのような記述形式を用いてコミュニケーションを進めているかの調査が必要である。そこで我々は、ソフトウェア工学授業のプロジェクトとして行なわれた仕様設計作業におけるメイリングリストからデータを収集し分析を行った [7]。

仕様設計作業において学生は、仕様要求者 3,4 名および要求分析者 1 名のグループを組み、ミーティングスケジューラシステムの仕様作成を目的として、メイリングリストを介したコミュニケーションを行った。12 件のメイリングリストにおいて計 534 通の電子メールが約 6 週間の期間に送信された。

収集されたデータの分析の結果、以下の特徴

が観察された。

- メイリングリストで交換される 1 通のメールには複数の「話題」が存在することが多くある。
- 話題は、複数のメールにわたって時には引用されながら繰り返される。したがって、複数の話題が複数のメール間に混在する。
- 分析した多くのメールにおいて、上述のようないわゆるネチケットとされるメールの約束ごとが成立している。具体的には、

- ユーザは話題を文書で書き、1 行を半角 65 文字程度に改行
- メールを受信したユーザは、送られた文書の一部を行頭に引用マークを付けて引用し、新たな文書を付加

といった約束ごとが、534 通中 491 通 (91.9%) のメールに見られた。条件に反したメール 43 通のうち代表的なものとしては、

- 発言者の名前を行頭に付けたものなど、デフォルト設定外の引用マークが使われている。(11 通)
- メール最後に引用文が残され、それに対するコメントがない。(21 通)

などがあった。

- メール間の依存関係を示すとされる “In-Reply-To:” や “Reference:” といった明示的な返信のヘッダー情報が必ずしも付加されていない。例えば、明らかに依存関係が存在すると思われる引用マークの付いているメールが 317 通あったが、そのうち、“In-Reply-To:” もしくは “Reference:” のヘッダー情報が付いていたものは計 222(70.0%=222/317) に過ぎなかった。

2.3 既存の履歴表示法の問題点

以上の調査結果より、既存の履歴表示法には多くの問題点があることがわかった。

まず第一に、既存の履歴を表示する方法の多くはメール 1 通ずつを単位としてその間の関係を表示させるものであるが [3][4]、1 通のメールに複数の話題が存在することから話題を中心と

して履歴を追う必要のある場合にはこれは不適切である。

またメールのタイトル情報やヘッダー情報をもとにしてメールのタイトルのリストをツリー構造に表示させる方法 [5] では、このようなヘッダー情報がきちんと付加されていない場合が多くあることからやはり適切でないことがわかる。

これに対して引用マークの使用など表層的な記述構造の決まり事は、強制的なルールではないにも関わらず、ユーザに比較的多く遵守されていることが観察された。

2.4 話題間の関係に基づいた履歴表示法の提案

本論では電子メールによるコミュニケーション履歴を表示する方法として、メールから話題の項目を抽出し項目間の対応関係を明示化させることでコミュニケーション全体の履歴を可視化させる方法を提案する。メールからの話題の抽出には、自然言語処理で用いられる形態素解析といった会話内容の分析はおこなわず、引用マークといったメールの表層的な構造にのみ着目する。すなわち、我々の履歴表示法は、メールでのコミュニケーションにおいて参加者が、

- 話題を文書で書き、1 行を半角 65 時程度に改行して、メイリングリストに送信する。
- 送信された文書に含まれる話題を続行する場合には、その一部を行頭に引用マークを付けて引用し、新たな文書を付け加えて、メイリングリストに返信する。

という条件に従うことを前提としている。

このような単純な方法によっておこなわれる会話履歴の可視化を支援するツールでは部分的にしか話題の関係は抽出できない。そこで我々のアプローチでは可視化はあくまでも補完的機能として位置付け、もとのメールの情報が保持されるような表示を提案する。すなわち、

- メール単位の単位性 (内容を投稿時の状態で復元可)
- メールの時系列 (内容を投稿の時系列順に従って閲覧可)

を保障することにより可読性を保持するものとする。

3 履歴表示の実現方法

話題間の関係の明示化によるコミュニケーション履歴を可視化させる方法は、

1. メールのチャンク化
2. 話題の流れの抽出
3. 話題の流れの表示

という手順からなる。チャンク化および抽出は、現在 UNIX ワークステーション上の開発環境で C 言語によってプログラムを実装中であり、MH メールシステムで保存されたメールファイル (MH では 1 通のメールに対して 1 つのファイルを生成する) を処理する。また、表示には、Smalltalk (VisualWorks 2.5) 上で実装されている ART システム [8] を利用している。以下に、コミュニケーション履歴を可視化させるまでの手順を、例を挙げて説明する。

3.1 メールのチャンク化の手順

もとのメールの文書は、図 1 のように、テキストとして文字列が羅列しており話題の対応関係が結びついていない。そこで、文字列抽出と文字列検索により、話題の対応関係を決定し、対応関係にある話題を 1 まとまりにする (チャンク化)。なお、メールの各行について、

- 行頭に引用マークが付いている文を「引用文」
- 行頭に引用マークは付いていないが、文字が表示されている文を「記述文」
- 表示された文字がない文を「空行文」

と分けて呼ぶ。チャンク化の手順をフローチャートにしたものを図 2 に示す。

チャンク化の特徴として、

1. 連続する記述文は 1 つのチャンクとする
2. 連続する引用文は 1 つのチャンクとする

メール1	メール2	メール3
AAA	> BBB	GGG
BBB	DDD	>> BBB
CCC	> CCC	> DDD
	EEE	HHH
	FFF	> EEE
		> FFF
		III
		JJJ

図 1: メールのテキスト

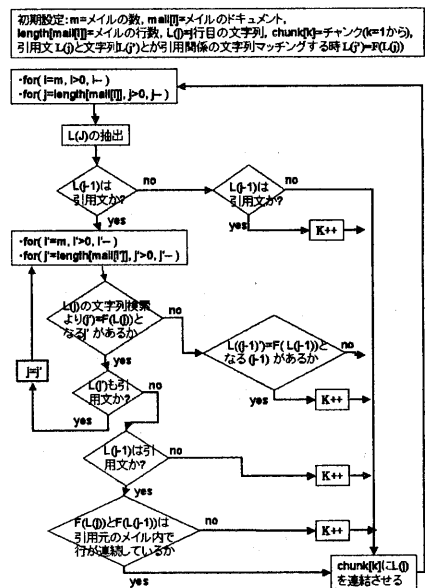


図 2: チャンクの手順のフロー

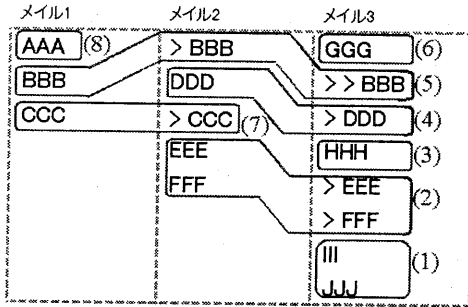


図 3: チャンク化の手順

3. 引用の内容(文頭の引用マークとスペースを除いた文字列)について、それよりも前に到着しているメール内を文字列検索し、検索された部分を1つのチャンクとして引用関係を設定する
4. 二重引用の部分は別のチャンクとする
5. メール中の空行文ならび引用マークのみの引用文(「>」)のチャンク化は行わない

などが挙げられる。

これらの規則を、履歴表示させたい全メールテキストに対して後着順に適用させることで、全メールの文書がチャンク化される。図3の例では、3通のメールは計8個のチャンクに分けられたことになる。

3.2 話題の流れ抽出の手順

話題の流れを抽出する手順は、チャンク化同様、後に着いたメールの行末より行う。

引用文のチャンクの直後に続く記述文のチャンクを1つの話題とする。また、引用文と引用関係のある記述文のチャンクも1つの話題とする。図3の例では、まず、チャンク(1)と(2)、チャンク(2)と(7)をそれぞれ合わせることで、1つの流れを抽出する。これを全メールのチャンクに対して行い、全メールの話題の流れを抽出する。

図3の例から話題の流れを抽出したものを、図4に示す。この例では、8個のチャンクから計4つの話題の流れが抽出されたことになる。

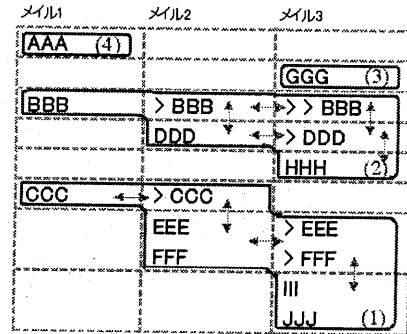


図 4: 話題の流れの抽出の手順

抽出した話題の流れ1つ1つを見ると、引用関係にあるチャンク間のつながりが横軸に、引用文とそれに対する記述文とのつながりが縦軸に、並べられている。これにより、コミュニケーションの履歴をトップダウンに追うことができる。

3.3 話題の流れの表示の支援

現在メールから抽出した話題の流れは、ARTシステム[8]を利用して表示の支援を行っている。

ARTシステムは、文章を書いてまとめる作業を支援するものであり、Elementと呼ばれるドキュメントの部分部分を基本単位として、2次元空間上にElementをマッピングさせることで、全体のドキュメントを構築することができる。

このARTシステムの2次元空間に図4のような縦横の格子を想定し、メールから抽出した話題の流れを配置させる。実際に話題の流れを配置させたものを図5に示す。

この支援システムでは、左側の2次元空間に、各メールから抽出したチャンクを1つのElementとして配置させる。図4同様、引用関係にあるチャンクが横軸に、引用文とそれに対する記述文とが縦軸に並べられており、コミュニケーションの履歴を上下左右の各Elementで追える。

そして、システムの右上の空間には、抽出した話題の流れがドキュメントとして表示される。これにより、注目し選択したElementの話題の履歴(図5の文字を反転させたElement群)をテキストで見ることが可能となる。

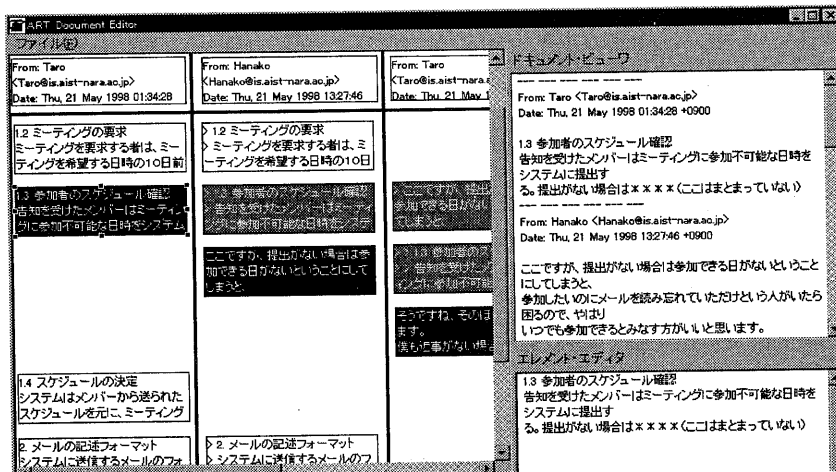


図 5: 履歴表示支援システム

4 おわりに

本論では、メールのやり取りの約束ごとに基づきメールから話題の項目を抽出し項目間の対応関係を可視化させることで電子メールによるコミュニケーション履歴を可視化させる方法について述べた。今後は、約束ごとの条件に反するメールへの対応および実際にユーザに利用してもらい有効性と問題点の調査を予定している。

参考文献

- [1] G. Fischer, J. Grudin, A.C. Lemke, R. McCall, J. Ostwald, B.N. Reeves, F. Shipman, Supporting Indirect, Collaborative Design with Integrated Knowledge-Based Design Environments, Human Computer Interaction, Special Issue on Computer Supported Cooperative Work, Vol.7, No.3, pp.281-314(1992).
- [2] G. Fischer, K. Nakakoji, J. Ostwald, Supporting the Evolution of Design Artifacts with Representations of Context and Intent, Proceedings of Symposium on Designing Interactive Systems, ACM, pp.7-15(1995).
- [3] 佐藤, 佐藤: ネットニュース記事群の自動パッケージ化, 情報処理学会論文誌, Vol.38 No.6 pp.1225-1234(Jun 1997)
- [4] 小作, 内元, 井佐原, 知的ニュースリーダー-HISHO-の開発, インタラクシオン'98 論文集, pp.153-156(1998)
- [5] David H. Crocker, Standard for the Format of Arpa Internet Text Messages, RFC822(Aug 1982)
- [6] S. Hambridge, Netiquette Guidelines, RFC1855(Oct 1995)
- [7] 松浦, 高田, 中小路: 要求獲得の過程における電子メールを用いたコミュニケーションに関する研究, ソフトウェアシンポジウム'98 論文集 pp.35-43(Jun 1998)
- [8] 山本, 高田, 中小路: "Representational Talkback" の増幅による「書いてまとめる」プロセスの支援へ向けて, 人工知能学会誌 Vol.14 No1 pp.82-92(Jan 1999)