

## 送信管理簿を用いた送受信メールの関連抽出の試み

3E-2

清末悌之(kiyosue@nttvd.t.hil.ntt.jp)、犬童拓也(indo@nttvd.t.hil.ntt.jp)  
NTTヒューマンインタフェース研究所画像通信研究部

## 1. 背景

コミュニケーションの手段として電子メールシステムが業務の中で多く使われるようになってきた。次第に交わされる電子メールの数が多くなり、単位時間当たりに読まなければならない数が増えている。その時点でホットなスレッドの数が増加し、複数のコンテキストに基づく議論がコンカレントに行なわれるようになってきている。

このような問題点に対し、メールを使用する環境を予め制限することで、一連の流れを把握することができるようにする構造化メールのアプローチが検討がされている[1]。代表的なものに、The CoordinatorやgIBISなどがある。しかし、これらは、予め想定しているコミュニケーションのモデルを逸脱する場合には対処ができない。即ち、予め、コンテキストの大枠を決めてしまっているところに課題があると思われる。

一方、メールの構造自体は通常のままとし、メーラが予め設定しておいたフィルタを介して、自動的にカテゴリ毎のディレクトリに振り分けるという手法がある[2]。しかし、これも予め想定しているカテゴリへ振り分けるのであり、時々刻々発生する新しいカテゴリや、細分化されるカテゴリへはフィルタリングできない。また、一通のメールを一つのカテゴリに分類するだけでは、複雑な内容を持つ送受信メール間の関連に対処できない。即ち、複数のカテゴリに関与する場合は、コピーを複数のカテゴリに存在させることや、ラベルを複数付与するしかない。

これらの現象を引き起こす原因は以下のようなことであると考えられる。

- ・電子メール間の関連は木構造のような単純な構造ではなく、ハイパーなネットワーク構造になっている。
- ・事前にカテゴリ付けできる場合は少なく、カテゴリは常に流動（発生／消滅、派生／統合）している。
- ・メールのやり取りは、トップダウンの中央集権的な枠組みの中で交わされているのではなく、ユーザ各自の意図により自主的に行われている。

本発表は、このような現象に着目し、新しく電子メールを受信した時に、過去に受信もしくは送信した電子メールで関連したものがあるかどうかをナビゲートするシステムについて述べたものである。

## 2. アプローチ

上記自動関連抽出機構を実現するために、我々は以下のスタンスで検討を行った。

- ・内容の自動解析ではなく、人間の内容理解を阻害する要因の低減をめざす。内容理解そのものは人間の作業ととらえる。
- ・コミュニケーション形態を規定せず、ナビゲートをめざす。予めコンテキストの大枠を決めない。
- ・メールのヘッダに付与されている情報を使って関連情報を抽出する。メールの内容は用いない。
- ・通常のメールの送受信のための手続きに以外に、特別な操作をしないようにする。新規にヘッダを付与することもしない。通常のヘッダを用いる。

また、我々はここで、送信したメールを常に「保管」という手法を用いることにした。

1章で述べた問題点に対処するために、現在、電子メールシステムのユーザは、自分宛にCcを付ける等により送信メールを保管することを行っている。リプライを返す側も、リプライの元になった受信メールの内容やヘッダの一部を、送信するメールの本文中に引用することが多い。メールを受け取る側は、送られてきたメールとそれ以前のメールの関連を知るために、内容に頼って関連を把握しているわけである。

## 3. 構築するシステムの内容

我々がここで目標とするのは、主にディスカッション目的でメールの送受信が頻繁に行われたとき、もしくは行われているとき、その経緯を過去にさかのぼって把握することができるようにすることである。

メール管理のイメージを図1に示す。ここでの主要な特徴は以下のとおりである。

---

"An extraction method of relations among mails using a table of sended mails"

Yasuyuki KIYOSUE, Takuya INDO

NTT Human Interface Laboratories Visual Communication Laboratory

- ・アカウント毎に、一通の電子メールを一通のファイルとして、管理する方法で行う。
- ・電子メールは受信したメールだけでなく、送信したメールもファイル化の対象とする。  
このシステムを用いて、電子メールを登録するときの処理と、登録した電子メールを基に、関連を抽出するときの処理を以下に示す。

1) 登録時

ファイル名については以下の命名規則に従うようにした。

- ・ファイル名には、アカウント毎にユニークな番号 a を付与する。
- ・ファイル名には、また、関連のあるメールファイルの番号 c を付与する。
- ・さらに、番号 a と番号 c で表される二通の電子メールの関連形態を、簡単な記号 b で表しファイル名に付与する。形態は、現在主要に行われている機能形態にとどめ、送信、リプライ送信、フォワード送信、受信、リプライ受信の5つとする。
- ・上記3つのファイル名の要素 a、b、c を、(ピリオド) で区切り一つのファイル名とする。

要素 a は順番に付与することとし、要素 b、c は、送信時は送信時のユーザのコマンドから、受信時は、In-Reply-To フィールドを利用することで抽出することとする。送信時には送信メールのコピーを、この命名規則に基づいてファイルを作成する。受信時には、いったんテンポラリファイルとして格納し、その後上記命名規則に基づいてファイル名を決定する。

2) 利用時

利用時には、登録時に作成されたファイル群のファイル名を対象にして、簡単なパーザを用いて関連を抽出する。すなわち、ファイル名を基にして関連するメールを抽出し、ここからさらにイモズル式に関連するメールを抽出することで一連の経緯を得る。

- ここでの処理は以下の手続きを再帰的に進める。
- ・現在、注目しているファイル名を、a . b . c とし、c が null のとき、単なる送信か単なる受信のとき、他のメールのとの関連はない。
  - ・c が null でないとき、a' . b' . c' で a' = c なるファイルを探し、このファイルとの関係が b である。

例えば、図1中のアカウントAでメール(ファイル) 5 . r . 4 を見ているとき、このメールに関する一連の経緯は、図2で示されるようになる。

4. まとめ

どのようなコンテキストにも対応できる、電子メールの関連抽出のための手法を検討した。今後は検討内容のインプリメントを進めるとともに、まったく異なる経路に到着した複数の受信メールの互いの関係を抽出する手法について検討したい。

いつもご指導いただく島村部長、西村グループリーダー、および熱心に討論いただいたグループの方々に感謝します。

参考文献

- [1]石井「CSCWとグループウェア」8章、1994、オーム社
- [2]「MH&xmh」11章、1995、ASCII

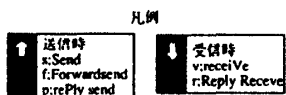
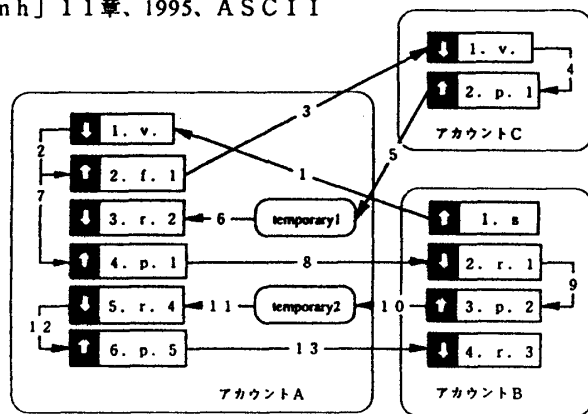


図1 動作概念図

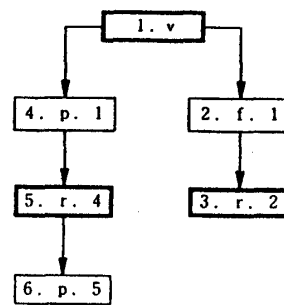


図2 抽出された関連情報