

メーリングリストにおけるメール情報自動抽出： そのグループウェア的考察

山上俊彦

NTTマルチメディアネットワーク研究所

グループウェア研究会のメーリングリストを構築するために初期メーリングリストメンバを登録するための方法を検討した。最も簡易な研究会でのメールやニュースをもとにメールアドレス情報を抽出する方法を議論した。プロトタイプをPerlで作って試験したところ、さまざまな課題が提起された。この課題提起を詳細に検討し、一般のグループウェア設計にも共通であると思われるいくつかの項目を体系的に整理することを試みる。その結果、さまざまな課題を、技術的、社会的、長期挙動的、参加セマンティクスの、の4つに分類した。これらが一般のグループウェアにも適用可能なことを示唆し、これらの現実世界への適用方法について議論する。

Automatic Mail Address Extraction in Mailing List Management: A View from Groupware

Toshihiko Yamakami

NTT Multimedia Networks Laboratories

yam@acm.org

A mailing list is planned to create for Special Interest Group of Groupware in Information Processing Society of Japan. To initiate the members in the mailing list, several approaches are discussed. To easily bootstrap the mailing list, automatic mail address extraction from mail and news information is attempted. A small prototype is implemented by Perl. This prototype experiment shows a wide variety of issues for automatic mail address management. The author discusses the issues and categorized them in a systematic manner. The issues are classified by 4 groups, technical ones, social ones, long-term ones, and participatory semantics ones. From groupware design viewpoints, these four classifications are discussed to lead to implications for future groupware design.

1. はじめに

さまざまな電子コミュニティにおいては、長い間にさまざまなメッセージが電子メール、メーリングリスト、掲示板、ホームページなどで交換される。ここで、長期間の協調行動においては、メンバの出入りがあり、名簿を管理するのが困難であるという課題がある。

さらにこの名簿を管理している場合、メンバが脱退した後でも関心があるような場合、拡大メンバーリストの生成が必要である場合がある。十分、注意をして管理している場合にはよいが、管理人が何代にもわたってかわるような場合には、このような拡大リストを管理運営することにも大きなコストがかかる。

本原稿では、情報処理学会グループウェア研究会で、連絡委員の交代が多くなる設立4年目において、どのように拡大メーリングリストを作成、管理したかをケーススタディとして、このように疎に管理されたメーリングリスト運営の要求条件を整理する。

2. アドホックヒューマンネット

まず、コミュニティの生成について述べ、コミュニティの拡大に伴うアドホックヒューマンネットという概念について述べる。

コミュニティの生成については図1に示すように closed community → semi-closed community → open community の三段階に拡大するような概念的モデルが成り立つと考えられる。このような段階的進化は、例えば筆者が行ったノウハウ支援システムにおける3年間の情報伝播形態の遷移分析

[Yamakami95] のなかにも見ることができ、強いインタラクションの存在しない集団の成長としてよくみることができる。最初は小さな核集団であるが、やがて派生集団が成立し、活動が活発になると、その活動自身を参照する不特定多数のメンバが疎に結合する。これがコミュニティである。ここでは登録などが必要ない不特定多数の集団をコミュニティと呼ぶ。世の中には例えば、市役所での住民登録など暗示的に登録を背景とするコミュニティもあるが、ここでは、明示的な集団構成、登録が必要ないという側面に着目して議論を進める。

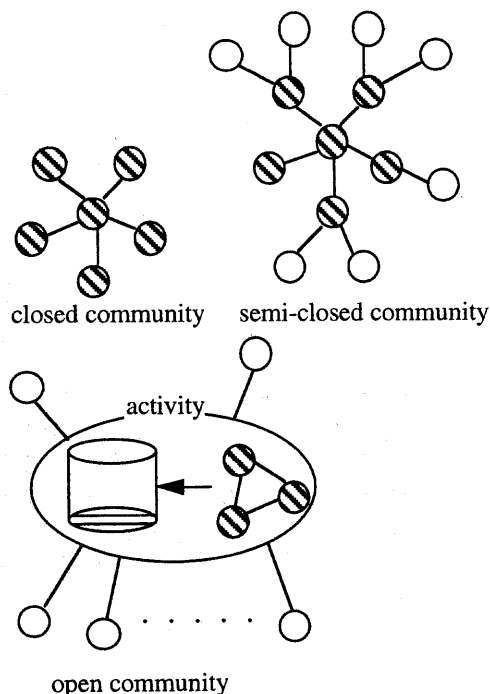


図1 コミュニティの遷移モデル

このようにして生成した特別の規約、役割分担のないヒューマンネットワークを筆者はアドホックヒューマンネットと名付けた(図2参照)

- ・ 特定の管理機関がない
- ・ 利用者からの登録が必要ない
- ・ 柔軟な構造
- ・ 特定の連絡系統を持たない
- ・ インフォーマルな関係を共用する

図2 アドホックヒューマンネットの定義

このようなヒューマンネットワークは自然発生的にできた大規模ネットワークにおいて観察することができる。このようなヒューマンネットワークは通常は特別の活動をせず場面場面において情報交換や交流をすることにとどまっている。しかし、例えば学会のヒューマンネットワークにおいて、個々のメンバに対して特定の働きかけ(イベント)などを通知しようとする、非常にコストがかかる。

3. アドホックネット電子化の要求条件

このような条件において、電子通信技術

を利用してアドホックヒューマンネットを電子化し、働きかけのコストを小さく、アドホックネットの運用を効率的に行いたい、というニーズがある（表1参照）。

表1 メーリングリストの相互のユーザ要求

発信者	・タイムリーな情報発信 ・特定属性の受信者への到達可能性 ・低コスト
受信者	・タイムリーな情報入手 ・活動への気付き ・情報洪水の弊害

特に日本のインターネットユーザが300万人を超えるような状態においては、インターネットを利用した電子化のニーズは強い。

しかし、このようなメーリングリストの管理については、そもそもそのようなメーリングリストを前提としないアドホックヒューマンネットにおいては、さまざまな課題が存在する（図3参照）。

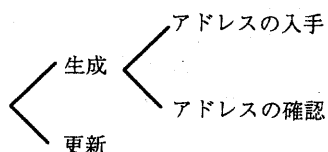


図3 アドホックヒューマンネットの電子化の課題

グループウェア研究会のメーリングリストを作るにあたって次のような方法があった。（1）学会の情報に基き作成する

多くの研究会が錯綜し、事務局に過度の負担をかけられないという制約がある。これは従来のグループウェアにおける事務局、総務と設計者との問題と同じである。

（2）研究会独自データに基き作成する

独自データは十分メンテされていない。また、初期の会員名簿が異なるプラットフォームで作成されており確認する。

（3）電子メールから自動生成する

多くの広報情報が作成されており、電子的な交換が行われている。このような電子メールアーカイブがある段階では、自動生成することも可能である。

例えば学会においては、登録会員の管理は行われているものの、メンバリストの管

理は学会で行っており、連絡委員会の名簿管理とは独立に行われている。また、学会におけるさまざまな活動の広がりやを考えると学会にメーリングリストの管理を委託することは困難な状況にある。

メーリングリスト管理の問題点は、メンバの変更が頻繁である、メンバが異動によってメールアドレスが変更になることが多い、さまざまな一時的異常により不達になる場合メーリングリスト管理に負担がかかる、組織によってドメインネームの変更などがある、などである。

メーリングリストにおけるメールアドレスの更新について、例えば、グループウェア研究会連絡委員（約25名）における更新履歴は次の通りである。1995年4月～1997年4月までに26回。同じマシン上で管理されているISO SC18のメーリングリストが19回である。必ずしも一般性は高くないが、ほぼこの規模でメールアドレスの更新が起こっている。比較的電子メールに親和性の高い集団でもメールアドレスの更新は比較的高い頻度で起こっている。このような状況では、直接アドホックヒューマンネットにかかわっていない人々に多くの作業が予想されるそのような作業を委託することは難しい。

このため、グループウェア研究会では、研究会で交換されたメールなどの中から関係する利用者を抽出するプロトタイプの実験をボランティアで開始している。

4. メーリングリスト化へのアプローチ

実験のために、まず、グループウェア研究会の連絡委員、研究会の準備のためのメーリングリストからメールを抽出し、そのメールアドレスの解析を通して、メーリングリストの初期状態を自動生成することを検討した。

初期生成後は、登録された利用者に対して、メーリングリストへの登録の可否を問い合わせ、それに基づいて最初のメーリングリストを作成する。

以後は、研究会への参加者を更新する形でメーリングリストのメンテナンスを行うことが考えられる。本原稿では、最初のメールアドレスの抽出実験からさまざまな考察を行った結果を報告する。

5. 自動抽出のプロトタイプ実験

メールアドレスの自動抽出実験を行った。

プログラムは Perl (version 5.003) で記述した。図4に全体の構成を簡単に示す。プロトタイプは約200行のPerlプログラムである。

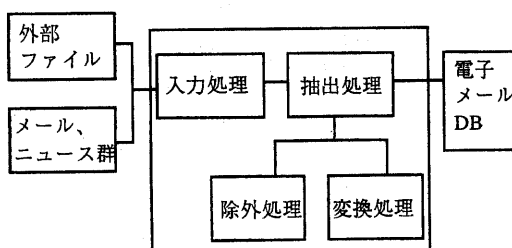


図4 アドレス自動抽出の処理の流れ

電子メールアーカイブから自動的に電子メールを抽出した。電子メールアーカイブとしては、グループウェア研究会の連絡委員のメーリングリストなどグループウェア研究会関連のメーリングリストの過去15ヶ月の約1000通のメールを対象とした。

さまざまな課題がプロトタイププログラム実行中に発見され、それらはプロトタイプ機能の中に順次盛り込まれた。体系的に整理すると以下になる。

(1) 抽出精度の問題

メールアドレスの抽出として不十分である技術的問題を分類する。

(1-1) 電子メール以外の抽出

・メッセージIDなどヘッダ情報からの抽出
AA00565@CLXS-10.ipsj.or.jp、0@0、等。

注：XXXnnnnn@niftyserve.or.jp などをは有効である。

・IPアドレスを含むものhoshi@10.33.21.10

(1-2) 語分割の問題

・Bjm@ics.kagoshima-u.ac.jp、など先頭の大文字が挿入される。日本語処理の問題ともからむと思われる。

・などで分割されて一部が失われる

(1-3) !や%による中継ドメイン指定の問題

・foo%bar@organization.company.com など

(1-4) 1行から複数の名前を抽出する

・1行に複数の名前があることを仮定するべきかどうかの問題

(2) メールアドレス不適合問題

抽出としてはメールアドレスとして抽出されているが、メールアドレスが無効になったり、アドレス自体が記述された段階で誤っていたり、などの意味的にメールアドレスとして成り立たないものを分類する。

(2-1) 古いアドレスの問題

・無効になった情報 foo.ntt.jp や bar.nttdata.jp など、企業や団体におけるDNS体系の更新、など。

(2-2) 利用者が記述した誤りメールアドレスをそのまま抽出

・記述自体が誤り
・メール送信時に誤って送ったメールの宛先アドレスから抽出

(3) 社会的問題

社会的問題としては、メールアドレスの抽出としてはOKであり、該当するメールアドレスは存在しているが、目的にそぐわないものを分類する。

(3-1) 海外アドレス

・招待講演者など、uk, edu, it, com アドレスが含まれる

注：foo@acm.org など国際機関アドレスを使ったり、海外にcomアドレスを登録して利用する日本利用者をはねないようにする。

(3-2) ユーザ以外のメールアドレス

・root、MAIL-DAEMON など

注：例外的には、root が加入者ということも可能性としてはあり得る。

(3-3) メーリングリスト

・ipsj-gw@nttmhs.ntt.co.jp などメーリングリストが登録される。

(3-4) 複数アドレス

・利用者の発信アドレスから抽出した場合、さまざまな発信ドメインがそのまま収集される。このため、同一の利用者に対して複数の名前が抽出されることになる。

(4) 参画的問題

参画的問題としては、メールアドレスの抽出としてはOKであるが、利用者自身がメーリングリストのアドレスとして利ある場合を分類する。

以下のような課題に対することが要求される。

・このメーリングリストについて参加した

- くないという組織構造的課題への対応
- ・メーリングリストの脱退の管理
 - ・一時的不達などのエラーへの対応

これらの課題について、その抽出状況の量的分布について図5に示す。

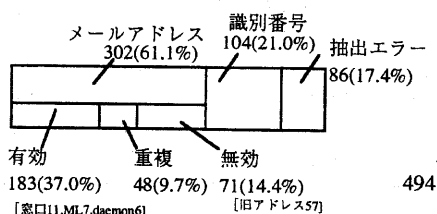


図5 抽出メールアドレスの内訳

抽出されたメールアドレスは、グループウェア研究会の登録会員数480に比べてほぼ1/3であり、今後、研究会発表者などによって年100程度ずつ拡大していくことによって当初の目的を達成できるレベルにあると考えられる。

ただし、メーリングリストなど直接自動利用する上では社会的な課題があり得るのが5%程度あり、実際の運用にあたってはさらに検討が必要であるという結果を得た。

6. グループウェアの観点からの考察

メールアドレスの抽出については、技術的な課題とともに、メールアドレス自体が使われるコンテキストと目的とする自動抽出との差異、また、ヒューマンエラーという避け難い問題、また、抽出自体は成功しても、抽出された情報自体が時間の経過とともに協調行動の目的と合わなくなっていくこと、など協調行動の設計、とくに長期的設計に関して、検討すべき課題を多く含んでいる。

これらを整理して、図6に課題の概念体系を示す。これより導出するデザインモデルの概念図を図7に示す。

このような概念モデルは、技術と社会、また社会の中でも個人と組織、また全体が、共時的な構造および時間とともに変化する経時的構造の両方を持つグループウェアの性質とよく対応していると考えられる。

例えば、このようなモデルに基づいて、従来のグループウェア研究を振り替えてみると、すべての現象を網羅的に検討した

研究は非常に少ないことがわかる。初期のグループウェア研究は非常に技術に片寄っており、さまざまな機能の実現が追求され、その結果が報告されていた。次に、社会的なインタラクションが注目され、アイコンタクトのような個人間インタラクションの問題、セキュリティやワークフローのような組織的な問題、spaceとplaceなどのような社会空間の問題については議論されている。大規模な組織技術移転などについては組織的な問題、長期的な問題が議論されている。

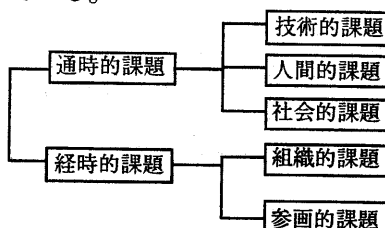


図6 課題の概念体系

		共時的	経時的
技術的		抽出精度	システム 更改
		役割アドレス	
社会的	人間的	人間的エラー	転出
		混合 複数アドレス	変化
	組織的	メーリング リスト	組織変化 異動

図7 デザインモデルの概念図

デザインモデルはこのようなさまざまな動きを統合するものであり、電子メールアドレス抽出のような単純な問題がこのような広範なデザインモデルに直接マッピングできるということは非常に有効である。この理由として、電子メールアドレスというものが非常に社会的なものであるということもある。逆に言えば、電子メールが1990年代の後半にいたってそこまで社会性を帯

びてきていることにもよると思われる。

従来のデザインモデルは、機能的なもの、プロセス的なものなどから考えられているものがあつたが、今回は、システム構築と同時にコンカレントエンジニアリング的に問題発掘を行い、障害から出発して本質を捕えようとするブレークダウン解析のような手法によって検討したものであり、よりユーザ中心的なデザインモデルであると考えられる。

7. 今後の課題

今後の課題を以下に示す。

- (1) グループウェアデザインの方法論
 - ・適用の一般性
 - ・グループウェアに関する設計結果と方法論との対応

- (2) メーリングリストの管理

- ・メーリングリストの管理（不達時、利用者の脱会）
- ・利用者の意向をどのように反映するか（利用者からどのように脱退などの意向を受け取り、その方法をどう周知共有するか）
- ・メーリングリストの利用方法の共有、認知
- ・意図せざる使い方による問題への対応（メーリングリストのループや不在転送プログラムの誤動作など）
- ・誰が管理するのか（アドホックヒューマンネットワークの管理という問題）

- (3) メールアドレス抽出プログラムの技術的問題

- ・例外処理を埋め込むかどうか
- ・特定アーカイブに依存しないような作り
- ・特定ドメインの排除（ipsj.or.jp はいらない、など）
- ・ルールの記述、メンテ

8. むすび

グループウェアをどのようにグループウェア研究会の運営、効率化、利便性向上に利用していくか、ということは、グループウェア研究会のプラットフォームや参加者の地域的、学際的な広がりを見ると優れてグループウェア的な課題であると考えられる。

本稿の段階では、まだ、メールアドレス

の抽出における問題把握にとどまっており、具体的なメーリングリスト運営においてはまたさまざまな課題を抽出、解決していかなければならないと考えられる。

しかし、コミュニティの成長からアドホックヒューマンネットワークという概念を提案し、そのための情報通信的支援としてアドホックヒューマンネットワークの電子化支援という課題をとりあげた。今後この範疇における研究は、アドホックヒューマンネットワークがさまざまな層に拡大するとともに注目されていくと考えられる。

メールアドレスの抽出という一見単純な課題がさまざまなグループウェア共通の目標である「複数の意味を持ち動的に変化する社会に対してどのように情報通信システムを適用していくのか」という課題に対応した、技術的、社会的、組織的、参画的課題を持っていることを示し、体型的デザインモデルについて考察することができた。

謝辞

本研究にあたって有益な助言をいただいたNTTマルチメディアビジネス開発部大町部長に感謝いたします。

参考文献

- [Grudin91] Grudin, J.: The convergence of two development contexts, *ACM CHI'91*, pp.¥ 91-98 (1991).
- [Nosek95] Nosek, J, and Shephard G.: Making groupware payoff: the British model, *Commun. ACM*38, 7 (July 1995), pages 11-13.
- [Urquijo93] Urquijo, S., Scrivener, S., and Palmen, H.: The Use of Breakdown Analysis in Synchronous CSCW System Design, *Proc. of ECSCW'93* (Eds. by Michaelis, G., Simone, C. and Shmidt K.), pp. 281-293, Kluwer Academic Publishers (1993).
- [Wall90] Wall, L., Perl, 1990.
- [Yamakami95] Yamakami, T.: Information Flow Analysis: An Approach to Evaluate Groupware Adoption Patterns, *Trans. IPSJ*, Vol.36, No.10, pp. 2511-2519 (October 1995).