

## コミュニケーション指向情報組織化手法の検討

斉藤典明[\*/\*\*]、水澤純一[\*/\*\*]、山本平一[\*]、山口英[\*]

[\*] NTT マルチメディアネットワーク研究所

[\*\*] 奈良先端科学技術大学院大学

[\*\*] 芝浦工業大先端工学研究機構

nori@nttmhs.tnl.ntt.co.jp, mizu@tnsyber.tnl.ntt.co.jp,

heiichi@is.aist-nara.ac.jp, suguru@is.aist-nara.ac.jp

あらまし

インターネットブーム以降、情報の氾濫を頻繁に経験するようになってきている。そこで、メーリングリスト上のコミュニケーションを支援するために、メーリングリストで交換された電子メールを蓄積し、蓄積された電子メールからキーワードを抽出し、このキーワードのうちコミュニケーションに深く関与しているものだけを選択することにより、メーリングリストの会話の特徴の抽出を行った。その結果、選択したキーワードの出現パターンおよび出現期間により一過性の話題と継続性の話題の区別ができた。本論文では、キーワードによる電子メールコミュニケーションの解析方法について述べる。

## Information Structuring Method based on Communication Pattern

Noriaki SAITO[\*/\*\*], Jun-ichi MIZUSAWA[\*/\*\*], Heiichi YAMAMOTO[\*], Suguru YAMAGUCHI[\*]

[\*] NTT Multimedia Networks Laboratories

[\*\*] Nara Institute of Science and Technology

nori@nttmhs.tnl.ntt.co.jp, mizu@tnsyber.tnl.ntt.co.jp,

heiichi@is.aist-nara.ac.jp, suguru@is.aist-nara.ac.jp

Abstract

We often experience overflow of information after the Internet boom. So, to support Mailing-List communications, exchanged e-mails at the Mailing-List are stored on the network, keywords are extracted from stored e-mails, the keywords are chosen based on communication, and communication pattern is clarified by using such selected keywords. As the result, we were able to distinguish transitory-topics and continued-topics by examining the appearance-pattern and the appearance-term. In this paper, an analysis method for e-mail communication by using extracted keywords is reported.

## 1.はじめに

インターネットの普及によりネットワーク上で自由にコミュニティ(ここでは機能的組織および共同体的組織の両方を差すことにする)を形成し、そのコミュニティ活動を支援することが重要になってきている。ネットワーク上での活動は、電子メールやWorld-Wide Web(Web)などのネットワークサービスによって行っている。特に非同期系のネットワークサービスであるメーリングリスト(ML)ではメンバーそれぞれが都合の良い時間に話題の進行に参加すればよい。また、最近はWebとの連携によりMLで交換された情報が蓄積されいつでも必要な時期の電子メールを再度参照できるようになっている。にも関わらずMLにおけるコミュニケーションが活発になればなるほど、メンバーが話題に付いていくための負担は増加している。

コミュニケーションが円滑に行われるには、コミュニケーションに必要となる共通知識や常識、過去の経緯などがメンバー間で共有されている必要があるが、現在のネットワークサービスによるコミュニケーションではメンバー間の共通意識を明示する仕組みがないために、メンバー間で「なんとなく」共有されているものである。そのため、新規メンバーはなかなかコミュニティに溶け込めない、電子メールそのものは非同期系のコミュニケーションでありながらメンバーでもしばらく目を離しているすきに話題の進行に参加できず発言する機会を失う、きちんと目を通せないまま整理してしまう、大量に溜まってしまった電子メールをうまく活用できず結局もう一度質問しコミュニティの雰囲気

気を壊す、共通意識があいまいなために誤解が生じコミュニティの雰囲気壊すなどの問題がある。

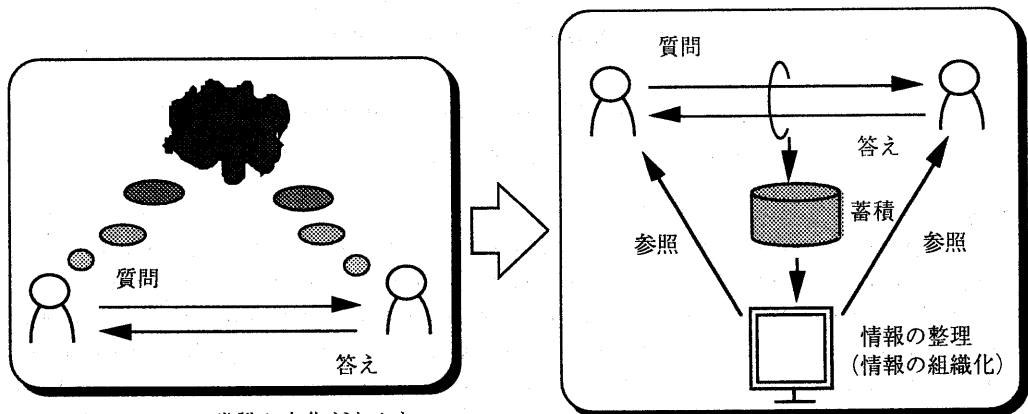
そこで、ML上で交換された電子メールを蓄積し、MLでの話題の進行を可視化する仕組みを導入する事により、これまで「なんとなく」有されてきたMLの常識や話題の変遷を抽出し、この抽出した情報により蓄積情報を組織化しメンバーに提供する事により、MLというコミュニティの運営に役立てる、あるいはこれまでの蓄積された情報を有効活用する仕組みの構築を狙う。

## 2. コミュニケーション指向の情報組織化

以上のように、ML上で交換された情報を蓄積しメンバーの共通意識を抽出し利用するには、単に情報を蓄積データとして捕らえ時系列からのブラウジングあるいはキーワード入力による検索により必要な情報を提示する仕組みを実現するのではなく、蓄積情報はコミュニケーションの過程そのものであり、そのコミュニティにおけるコミュニケーション形態にあわせて情報を整理し提供するアプローチが必要であり、このようなコミュニケーション形態を抽出し役立てるアプローチとしてコミュニケーション指向のアプローチとする。

この一つのアプローチとして、ML上での会話の中からメンバーが共通的に興味を持った話題に対するキーワードを自動的に抽出し、そのキーワードを元に一連のコミュニケーションにおける話題の特徴を提示する、あるいはML上の一連のコミュニケーションの話題の変遷を総括する情報を提示するアプローチを試みた。

蓄積情報を組織化し情報共有を支援する我々の



コミュニティの常識や文化がなんとなく共通意識として存在し閉鎖的

コミュニティの常識や文化が可視化され解放的

図1. 研究の狙い

以前のアプローチとして FISH/KINGFISHER というシステムがあった[1-3]。このシステムでは個々の情報に対してキーワードを付与しており、キーワードを入力する事により特定の情報の一覧が提示され、その一覧の中から選択することにより目的の情報を獲得するものであった。サーチエンジンも同様に入力されたキーワードに対して検索された情報の一覧が提示される。これらのアプローチでは、一つの情報が多くのキーワードにより検索されるように、一つの情報に多数のキーワードをラベルづけることが重要である。また、蓄積されている個々の情報は相互に独立した情報として扱われる。しかしながら、ML 上の情報は、個々の情報が独立して存在する場合もあるが ML 上の一連のコミュニケーション中での情報でありトピックごとにグループ分けが可能であり、どれか一つの電子メールを取り出すことはあまり適切ではない。そこで、個々の情報が独立しているわけではないので一つの情報に多数のキーワードによるラベル付けをするのではなく、一連の関係ある情報に対していくつかの主要なキーワードでラベル付けし、このラベルに基づいて蓄積情報を組織化しメンバーに情報を提供するアプローチが必要となる。

そこでここでは初めのアプローチとして、ML によるコミュニケーションを記録し、その中で興味を示された単語をキーワードとして抽出し、この抽出したキーワードに基づいて ML コミュニ

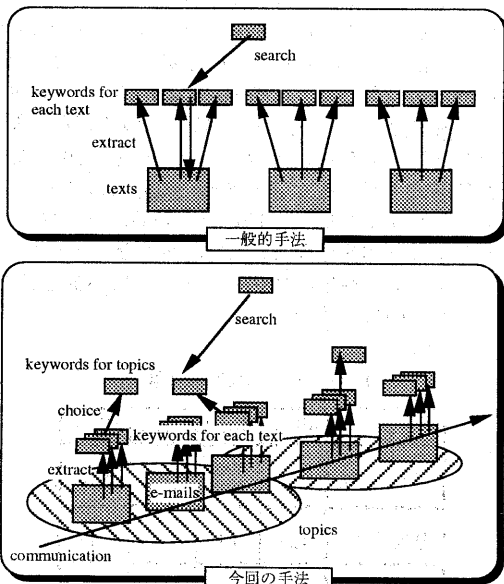


図2. コミュニケーション指向アプローチ

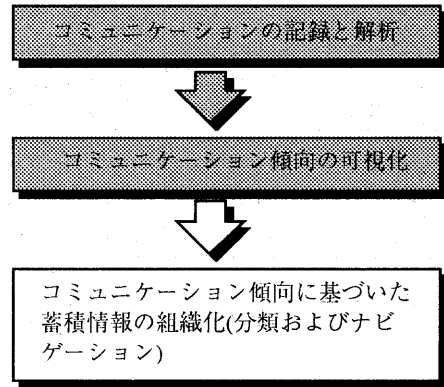


図3. アプローチ手順

ケーションの基本パターン調べた。(将来的には、このパターンに基づき蓄積情報を組織化しメンバーに提供する。)

### 3. MLコミュニケーションの観測

まず初めに、ML におけるコミュニケーションを解析するために扱った4つのMLの特徴を表1に示す。ML-1のメンバーは、元々ある程度の知人同士であり、非公式に情報を交換するMLである。ML-2はすべてのメンバーがほとんどあるいは全く面識のない人同士であり、非公式に情報を交換するMLである。ML-3は、ある程度の知人同士で構成されメンバーであり、メンバー間に上下関係があり、公式に情報を交換するMLである。ML-4は、ある程度の知人同士で構成され、公式に情報を交換するMLである。また、ML-1,3,4ではメンバーの変動は殆どなく、ML-2は自由応募のMLであるために逐次メンバーの変動がある。

これらのMLにおける電子メールの流通量の変動は表2のとおりであった(表2においてML-2のメンバー数は観測期間の最後の段階でのおおよその数である)。ML-1,3,4では一日のうち幾度にもわたって情報が交換されることは比較的少なく、単発的な電子メールが殆どである。ML-2

表1. MLの概要

	ML-1	ML-2	ML-3	ML-4
人数	約10	約60	約60	約30
期間(月)	約5	約10	約4	約12
メール数	115	2401	168	235

表2. 1日の電子メールの流通量

	ML-1	ML-2	ML-3	ML-4
11通以上	0回	85回	1回	1回
6~10通	5回	54回	4回	7回
2~5通	25回	76回	34回	43回
1通	13回	37回	30回	48回

ではML上に投稿された電子メールを元に非常によく会話が進展しているMLである事がわかる。

#### 4. キーワードの抽出と分類

先に述べたコミュニケーション傾向に基づき情報を組織化しメンバーに提供するシステムの実現にあたり、図4のようなシステム構成を考え一部運用している[4-5]。このシステム構成において電子メールの解析を行うにあたり、電子メールにはヘッダがあるものの必ずしも必要な情報がとれるわけではない事と、コミュニケーションの内部に検討を言及するために電子メールの本文に対して解析を加える事とし、MLへの適用を意識し短時間でキーワードの抽出処理が可能な方法ということで、(形態要素法などではなく)文章中から機械的にキーワードを抽出した。具体的には、「3文字以上16文字以下の漢字とカタカナからなる2回以上出現した文字列」

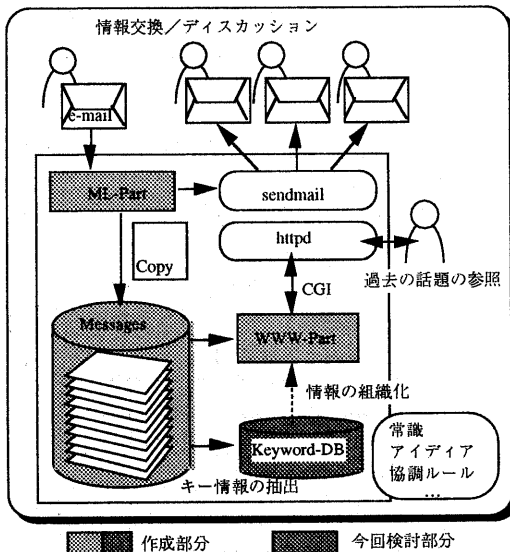


図4 システム構成

をキーワードとした。この方法では短い名詞やひらがなキーワード、英文字キーワードの抽出はできないが、今回は厳密なキーワード抽出が目的ではなく、コミュニケーションの特徴抽出が目的であるために、この方法により得られたキーワードを元にコミュニケーション中のキーワードの出現パターンを調べる事とした。この方法により抽出できたキーワードの例を表3に示す。この方法により抽出されたキーワードは表3の上段のようなものが中心であるが、文章表現において下段のようなノイズもいくつか抽出された。

表3. キーワードの抽出例

適切なキーワード	フラットスキャナー、電子マネー、エクストラネット、相互接続、クレジットカード決済、
不適切なキーワード	全然話、無沙汰、今月末、-----、今回

この方法により蓄積された電子メールから単純にキーワードを自動抽出した際のキーワード数は表4の「全文から」の欄である。電子メールにおけるコミュニケーションでは関心のあった部分は引用し、その部分について返事を書くのが一般的であることに着目し、電子メール中の引用された部分をMLにおいて関心の示された部分と仮定し、この部分からのみキーワードを抽出し検討する事とした。この場合のキーワードの抽出数は表4の「引用から」の欄である。

次にこれら抽出されたキーワードをその出現パターンから分類した。まず、そのキーワードが始めてML中で引用された日時と最後に引用された日時の差分をとり、この差分をキーワードのML中の寿命とし、仮定として30日未満の寿命を持つキーワードを短い寿命のキーワード、30日以上以上の寿命を持つキーワードを長い寿命のキーワードとした。(ここで、30日という期間はキーワードの寿命の全体的傾向と、MLにおけるコミュニケーションの話題の変化の期間から暫定的な値として直感的に決めた。)

表4. キーワードの抽出量

	ML-1	ML-2	ML-3	ML-4
メール数	115	2401	168	235
全文から	435	7010	1523	1113
引用から	74	2778	825	368

表5. キーワードの寿命と出現頻度の関係

	低頻度	高頻度
寿命の短い キーワード	偶然抽出されたもの	流行性の話題
寿命の長い キーワード	偶然抽出されたもの	常識化している話題

さらにキーワードの出現回数を出現頻度とし、短い寿命を持ち出現頻度の高いキーワードは一過性だが非常に興味の示された話題とし、長い寿命を持ち出現頻度の高いキーワードはML中で定期的に興味をもたれている話題と仮定し、解析の対象とした。

ここでは、調査可能な範囲という事で出現頻度の上位10%を頻度の高いキーワードとし、各MLにおける頻度の高いキーワードの出現パターンを次のように分類した。分類にあたっては今回は傾向をつかむ事が目的であるために定性的に分類し、その結果は表6の通りであった。

- ・I型:非常に短い期間(2日以内程度)に非常に多く引用された。
- ・II型:特定の時期に特に目立って多数回引用されたわけではないがトータルとして多数回引用された(話題のピークが存在しないキーワード)。
- ・III型:基本的には非常に短い期間に非常に多く引用されたが、その他の時期にも何回か引用されている。
- ・IV型:何回も特定の時期に目立って引用されている。
- ・NG:明らかにシグネチャ部分を引用したために抽出されたキーワード、しゃべり言葉から偶発的に抽出されたキーワード、プログラムのバグにより抽出されたキーワード。

表6からわかる通り、寿命の短いキーワードではI型の出現パターンが最も多い。このパターンはMLコミュニケーションにおけるあるトピックの話題の盛り上がりのピークを示している。この

「話題のピーク」という点において、III型はI型の同類と考えると寿命の短いキーワードの約7割が「ある一定期間だけ出現しそのあと2度と出現されなかった話題」を示すキーワードであり、このキーワードによりML中で刻々と変化している話題の概要を示す事ができる。寿命の長いキーワードではおよそ6割がIV型であり、「ML中で何回にもわたって議論が展開された話題」を示すキーワードであり、このキーワードによりMLにおいて常識化した話題や重要な話題を知ることができると考えられる。

一方、このIV型を「話題のピーク」という点から考察すると、寿命の長いキーワードは複数回話題のピークを有することによって成り立っていると考えることができ、先に仮定した30日を区切りに寿命の長さを計るよりは妥当な分類と考えられる。このことからキーワードの分類は

- ・話題のピークを一回だけ持つキーワード (=寿命の短いキーワード)
  - ・話題のピークを複数回持つキーワード (=寿命の長いキーワード)
  - ・話題のピークを持たないキーワード
- とする事が可能であり、III型は実は寿命の短いキーワードであり、IV型は全て寿命の長いキーワードであると分類できる。

しかしながら話題のピークを持たないII型がコミュニケーションの中でどのような意味を持つかは今後の検討課題である。

最後に実例ということでML-4において実際に抽出されたキーワードの上位10%を分類したものを表7に示す。なお、このMLは社内におけるネットワークの運用について情報交換およびディスカッションするためのMLであり、実際に抽出できたキーワードを見るとMLのおおよその雰囲気が伝わると思う。実際の抽出量を見てわかる通り、寿命の長いキーワードは寿命の短いキーワードと比較して絶対数が少ない。比率に関してはここでは述べないが、各MLにおいて

表6. キーワードの出現パターン(各MLの上位10%の合計の内訳)





	I型	II型	III型	IV型	NG
寿命の短い(30日未満)キーワード	56.5%	19.2%	12.1%	7.1%	5.0%
寿命の長い(30日以上)キーワード	—	19.0%	13.3%	61.0%	6.7%

別々であり、実際のコミュニケーションの活発さと関係あると思われる。

## 5. おわりに

今回の検討では、蓄積した電子メール本文中の引用部分だけに着目しキーワードを抽出し、その出現パターンから一回だけML中で関心を示されたトピックを表わすキーワード、何回も関心が示されているトピックを示すキーワードが抽出された。今後の課題としてこの話題のピークを利用してキーワード相互の関係づけを行い、相互に関係づけられたキーワードによる実際の電子メールへマッピングおよびグルーピングによる蓄積情報の組織化、この組織化された情報空間によるMLコミュニケーションの具体的な支援手法の実現が残されている。

表7. ML-4におけるキーワードの分類結果

	寿命の短い キーワード	寿命の長い キーワード
I型 	一本化 リモートアクセス 58回線 構成員 ハッカー対策 プロセス マ管G 停電中 情報通信環境 所長名 全館停電 武蔵野ロケ Mネ研全体 サブドメイン	
II型 	侵入犯 スケジュール アカウント 水別室	アドレス
III型 	技術的 ハッカー 変更作業	マシン
IV型 	横須賀 電話番号 パスワード HI研 モデム 連休中	メール Mネ研 ホスト名 ホスト
NG	斉藤典明 富永** 岡田**	
再分類		
I型 III型	50.0%	
IV型		27.8%
II型	13.9%	
NG	8.3%	

## 文 献

- [1]Y.Seki, T.Yamakami, A.Shimizu, "Flexible Information Sharing and Handling System -Towards Knowledge Propagation-", IEICE Trans. Commun., Vol.E77-B, No.3 Mar., 1994
- [2]斉藤, 爰川, 山上, 水澤, "Internet技術によるノウハウ共有システム", 信学技報, OFS95-7, May, 1995
- [3]N.Saito, J.Mizusawa, H.Yamamoto, S.Yamaguchi, "A Study of Knowledge-Sharing Method for Cooperative Work", 1st International conference IMAGINING USES, Bordeaux, France, pp.476-487, May, 1997
- [4]N.Saito, J.Mizusawa, H.Yamamoto, S.Yamaguchi, "A Study of Knowledge-Sharing Service based on WWW and Groupware", 1997 Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT '97), Hanoi, Vietnam, No 11-1, pp11.1.1-11.1.5, Mar., 1997
- [5]斉藤, 水澤, 山本, 山口, "MLとWWWによる情報の組織化手法の検討", 信学技報 OFS97-20, pp.49-54, Jul., 1997
- [6]佐藤, 佐藤, 篠田, "電子ニュースのダイジェスト自動生成", 情報処理学会論文誌 Vol.36 No.10, pp.2371-2379, 1995
- [7]前田, こうじ谷, 西田, "連想構造を用いた情報整理システム", 情報処理学会論文誌 Vol.38 No.3, pp.616-625, 1997
- [8]爰川, 高杉, 杉田, 国藤, "組織情報共有におけるキーワード抽出法の検討", 信学技報 OFS96-23, pp.51-56, 1996