

ネット時代の利用者調査方法としてのログ分析

東京大学大学院 豊島雅和¹

I. はじめに

人と人の関わるグループでの「共同作業」における実態をつかもうとするのは困難な作業である。そういった中で、コンピュータの普及とともに、その存在を前提としたグループ活動の比重は大きくなってきている。

そのグループ構成員をつなぐためのネットワークとしてのキーワードは、オープンなインターネットをはじめ、インターネット技術を社内利用に用いたイントラネット、その中間に位置する関連会社とのエクストラネットといった3つのネットである。このような情報技術によるネットワークを手軽に活用できつつある現在を「ネット時代」と呼ぼう。コンピュータやネットワークなどの情報技術を活用する共同作業のグループウェア研究はComputer Supported Cooperative Work (CSCW)として進んでいる。その中で、技術中心のとらえ方のCSにおいて、支援システムを具現化する研究は活発である。一方、CWを中心とする人間、社会、仕事中心のとらえ方からの社会科学研究は、さほど進んでいないと指摘されている(文献1)。今日ではビジネス誌において個々の企業での情報技術を活用したグループでの生産性向上などの成功事例なども紹介されてきている。しかし、それらはあくまで個別の事例であり、その多くは一般化可能性を追求したアプローチにはなっていない。

本稿では情報技術の存在を前提とし、クロウズドな企業などでの利用に立脚し、また主として図1中央での電子掲示板的な情報共有を中心

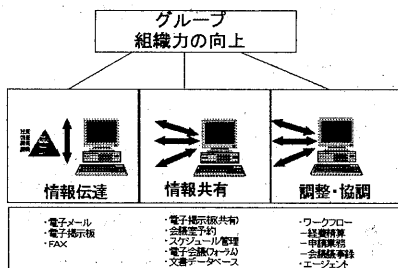


図1 グループウェアの範囲

に考察を広げる。

この範囲に議論を限定したとしても、情報発信側にて提供する情報の質やその内容の更新といった問題、利用者側からの認知、彼らのもつ価値観、利用環境などの要因が関連してくる。この複雑さの結果か、利用者の利用実態にまで踏み込んだ議論は、あまり多くはない。その中で、オープンなグループとしての特定のパソコン通信における電子コミュニティの研究は存在する(文献2、3)。それらは実態調査ということもあり、限定された環境における研究である。企業の目的に即した利用での「共同作業」としての情報共有ならば、その位置づけは鮮明である。一方、企業では、内部情報もあり、またノウハウが絡んでいると外部に発表しにくい側面も存在するだろう。この情報共有の分析、評価、また予測にまで耐えうる客観的で共通の基盤としての分析方法が整うと、この分野の研究はさらに進むことが予想される。

本稿は、共同作業における情報共有の複雑性と客観的手法の課題に答えることを目的としたネット時代の利用者分析を行うための調査手段の研究である。一方、手段だけの議論であると、具体的な活用イメージを伴わず、その効果を理解しがたい側面もあるため、イントラネットの情報発信評価の拙論文(文献6)での事例研究を、適宜題材として利用する。

II. ネット分析のアプローチ

本章では、次章以降で触れるネット時代の利用における前提とする環境、及び利用者像をつかむためのモデルを説明する。

1 ネット時代の情報環境の要件

先に触れた「情報技術の存在の前提」としてのネット時代における情報技術を活用するグループの情報環境の要件を考えてみよう。

第一には、情報機器としてのコンピュータは個人ごとに最低でも一台は存在することである。

第二は物理的なネットワークのインフラストラクチャーの整備である。全社のどここの事業所からでもこのネットワークに入れるLANやWANの環境が必要である。さらに出張先や自宅からのモバイル・オフィスのための環境が整うことも望まれる。

第三はソフトウェアの標準化である。利用者

¹ 勤務先：日本IBM株式会社

(toyosima@jp.ibm.com)

の情報を共有するためのアプリケーションに近いレベルにおいて、プラットフォームの統一も必要である。例えばグループウェアなどを利用することは、一つの解決策である。以下では、グループウェアの代表的なものの一つであるロータス・ノーツを題材とする。しかし、できるだけネットに共通する汎用的な形で述べる。

第四には、利用料金をそれほど気にせずに、快適な処理速度にて、かつ使い勝手がよく利用できる利用者としての快適さの側面である。

第五には、必要となる情報は、そのネットワーク上にて公開され、共有しうる状態になっていることである。それを、いつでもどこでも許可された誰でもが利用できることである。この五番目のレベルで語られることは、一般的には多くない。したがって次節以降は、そのあたりも含めて触れることにする。

どの企業、あるいは社会においても、これらの条件を満たす理想的な状態に遅かれ早かれ到来すると想定しよう。その途上の段階は、今回の考察の範囲外とする。

2 クライアント・サーバーモデル

Webサーバーやグループウェアにより、情報共有をしようとするイントラネットを基本として考えよう。大規模な企業での環境では、情報共有を目的としたサーバーを複数台、配置されることだろう。そのサーバーの所在がわかり、該当情報にアクセス権があれば、ネットワークを通しての利用が可能になる。

Webサーバーにおいてサイトを捜し、リンクなどによりディレクトリーを捜す。ロータス・ノーツであれば、ノーツサーバーと、そのサーバーの中にある文書データベース（以下DBと略）を捜す関係である（図2）。DBとはパソコン通信のフォーラム、SIGなどに相当する。

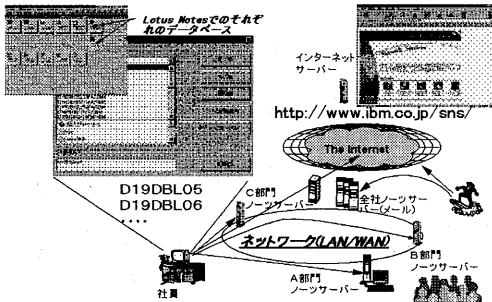


図2 企業におけるグループ

この処理形態をインターネットやイントラネ

ットのシステムのモデルとしてみよう。その構造を単純化すると、図3のようなクライアント・サーバーとして表現される。その処理の流れとしては、次のような3段階をとることが多い。

- ①クライアントは、情報アドレスを指定してサーバーへの処理を要求する。
- ②サーバーで対応した処理がなされ、処理結果はクライアントに転送される。
- ③処理が終了すると、サーバーにてログをとられる。

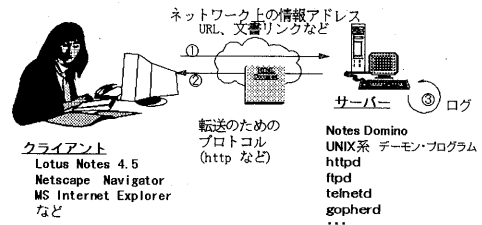


図3 クライアントサーバーシステムの構造

今回の分析の視角は、情報アドレスは既に認知していることを前提とし、上記に述べたクライアント・サーバーでログの集積される原理を使い、客観的に利用者層を分析するものである。

ログの本来の目的は、ハードウェアのリトライ回数などのエラー・ログなどにより、トラブルを解決することである。またデータベースの動作記録として、その復元などに用いられることも少なくない。今回はDBにおいて参画するときにとられるアクセス・ログ（以下単にログと略）を対象とするが、同一の仕組みである。

但し、ログのフォーマットは、それぞれのソフトウェアの実現形態によって相違する。図4は、ノーツのDBでのログの例である。ノーツのIDによる認証にてユーザーは識別され、記された日時は、そのユーザーがDBを開じた日時である。読み込みや書き込みの数は、その間に読んだり、書いたりした記事の総数を意味する。

| 日時 | ユーザー | 操作 | 記事数 |
|-------------|--------------------------|--------|-----|
| 10/01 21:30 | eiji Tanaka/SSD CS Dev | G 読み込み | 7 |
| 10/01 21:18 | Yasuko Matsui/SNS/IBM | 読み込み | 6 |
| 10/01 21:09 | Tetsu Teruya/GB SEIBU | 読み込み | 2 |
| 10/01 20:30 | Shigeko Kawahara/GB MET | 読み込み | 13 |
| 10/01 20:26 | Yukiko Suzuki/SNS/IBM | 読み込み | 12 |
| 10/01 19:23 | Naoya Yamamoto/Cross In | 読み込み | 1 |
| 10/01 19:15 | namie Matsui/SNS/IBM | 読み込み | 10 |
| 10/01 19:12 | akiko Higeta/Telecom&Mod | 読み込み | 4 |
| 10/01 19:09 | Matsushi Yamaguchi/East | 読み込み | 11 |
| 10/01 19:08 | Naomi Chikara/MO West | 読み込み | 4 |
| 10/01 19:07 | Satoko Kobayashi/Finance | 読み込み | 5 |
| 10/01 19:07 | kiyoshi Maruoka/MO West | 読み込み | 1 |
| 10/01 19:04 | Hiroshi Satoh/Open Clien | 読み込み | 4 |
| 10/01 18:33 | Hiroko Washioka/MO West | 読み込み | 7 |

図4 ロータスノーツでの画面とアクセス・ログ（一部加筆）

III. 第一ステップ: 「ログ分析により利用行動を把握できる」

最初のステップでは、「ネットワーク環境においては、ログ分析により利用者の利用動向を把握できる」と仮説をたてる。II. 1の要件を満たすコンピュータ関連製品サービスの開発、製造、販売、保守を事業内容とする従業員、約20,000人の企業でのDBの情報共有を事例研究として、その仮説の妥当性を検討する。

1 インターネット・フォーラムの位置づけ

この事例のイントラネットにて情報共有可能な部門サーバーは全国に散在し、平均すると約100人につき1台ある。多数あるサーバーの中の、そのまた一つのIフォーラムにはインターネット関連の特化した情報が集められている。インターネット・ビジネスの営業支援のスタッフ部門の管轄で、そこからの情報発信を本来の目的とするDBの内容の一部が図5である(1996年末時点)。

| | |
|----------------|-------------------------------------|
| Ownerから | |
| 96/02/05 10:10 | xxxxマーケティングガイド(Namy amuro) |
| 96/03/11 21:05 | xxxx接続サービス説明用資料(kusanagi tsuyoshi) |
| 96/04/05 10:10 | xxxx申込書(kawara tomochan) |
| 96/05/05 10:10 | xxxxご利用の手引き(Namy amuro) |
| フィールドから | |
| 96/07/05 10:10 | Real Audioを使いたい。(funaki kazuhiko) |
| 96/07/05 10:10 | Re:Real Audio. (harada masaki) |
| 96/07/01 10:10 | 技術的にはRA稼働は確認済み(okabe yutaka2, 返信文書) |
| 96/07/02 10:02 | けっこういいんですよこれが(harada masaki) |
| 96/07/03 10:10 | そんなに激安で大丈夫?(okazaki tomomi) |
| 96/07/04 10:15 | 追加料金が前提ですが、対応は検討します(tsuray usenagi) |
| 96/07/04 10:17 | とりあえず、社内でテストしてみませんか?(harada masaki) |
| 96/07/05 10:10 | 対応しました(mike toyohima) |
| 96/07/06 10:25 | 詳細がわかりましたらお知らせください(okazaki tomomi) |

図5 Iフォーラムでの情報(一部加筆)

対象はマーケティングに携わる営業あるいはSE系及び企画の職種の利用者である。なおDBアクセスのためのコントロールはせず、読込みのできる誰でもが書込むこともできる双方向性を保っている。

典型的な利用の流れを述べよう。主管部門において、有用と思われるインターネットの共有情報をDBにあらかじめ蓄積しておく。利用者は業務で必要とされる時にDBを検索、情報収集する。また企業の業務目的に即し、質問や意見交換の書込みをするといった利用である。

1:1を基本とする電子メール型と異なり、この種のDBは、その性格により一般的に1:Nの電子掲示板型と、N:Nのディスカッション主体の電子会議型とに分けられる。

Iフォーラムは、「Ownerから」の情報提供と

しての電子掲示板型を基本とする。また、「フィールドから」という質疑や意見交換をする電子会議的な側面も併せ持っている。

2 Iフォーラムを分析する

コミュニケーション研究では、内容分析や会話分析といった分析方法もある。今回の事例研究の状況を整理する手順としてそれらは参考になるので、その方式(文献4)により整理をしよう。

分析対象:

調査対象とする母集団は、マーケティング系の数千人にわたる人々である。そのうち、インターネット関連の情報を必要とする人は、その一部で、その数は明らかでない。彼らに直接的なインタビューやアンケートをするといった調査は、不可能ではないものの容易ではない。

書込みの利用者の匿名性はないのはもちろん、図4のように読込みだけの参画だけでも誰の利用かの判別はつく。したがって、本来の分析対象である利用者の代用としてログを扱える。

サンプリング:

1996年2月より11月に採集したログの全数を対象とする。この間でのログ分析によりわかる結果は確定的な結果をもたらす記述統計としての数値になる。

分析単位:

この事例でのIフォーラム全体は、一つのまとまった分析単位である。したがって、それより細かな、例えば一つ一つの記事を分析単位にすることはできない。

分析カテゴリー:

アクセス者名は分析対象の外的基準(文献5)として特徴を表わす質的変数である。その外的基準を特徴づける量的変数は、アクセス日時、読込み量、書込み量であり、カテゴリーに相当する。これら計4つは、当事例独自なものではなく、ノーツのDBにおいて汎用的に得られる情報である。

なお、このログでわかる組み合わせ以上の情報を得られないことは、カテゴリーの制約でもある。ネットワーク上で原理的に情報収集可能なものが(例えばWebサーバーにおけるホスト名)ログに含まれるならば、カテゴリーは異なる。但し、Webの場合では利用者を確定でき

ない場合もある。

データ記録：

ログは処理をされるたびに追加されていく性格をもつ。サーバーのディスクサイズには物理的な制約もあるため、一般的には古いログは消される。したがって、その消される前にログデータを記録しておかなければ分析はできない。ログを採集する分析の時間的単位としては、月次とし、時系列分析をするためのファイル化をする。なおログは、主管部門の管理者が利用者に見せない設定をしていない限り、誰でも見ることには可能である。

3 分析結果の整理

ログは連続した行のテキスト形式であった。したがって最初のステップは、それを各カテゴリに分離する必要がある。利用者名を変数として、表1に示す例のような、ソート及び集計を実施し、月あたりの総読込み量／書込み量、また実読込み／書込み人数を計算する。

| 人→ | 甲 | | 乙 | | 丙 | | 丁 | | 合計 |
|----------|----|----|----|---|----|----|----|---|----|
| アクセス日(時) | 1 | 3 | 5 | 3 | 20 | 25 | 15 | 7 | 25 |
| Read数 | 10 | 13 | 22 | 3 | 6 | 14 | 10 | 1 | 3 |
| Write数 | 0 | 0 | 0 | 0 | 3 | 1 | 0 | 0 | 0 |
| アクセス回数 | 3 | | 4 | | 1 | | 2 | | |

表1

読込み書込みの時系列での変化をマクロに把握するのは第一の分析であり、図6はそれをグラフ化したものである。この集計により、例えば9月において447人が、延べ約18,000記事の読込み、また32人により270の書込みを実施されたことがわかる(総アクセス回数は約2,500)。このような、どれだけの人にどれだけ記事が読まれたかといった関係は、従来研究では明らかにならなかったものである。

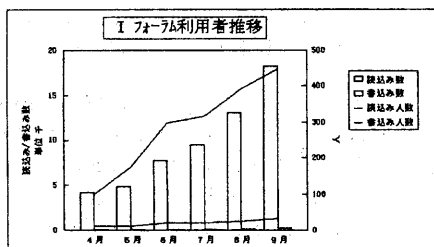


図6 読込み/書込み量、実利用人数による評価

第二の分析は個人の利用をマイクロに扱うもので、DBのアクセス回数分析である。利用者ごとの月あたりアクセス(接続)回数を新たな量的変数とする。これはカテゴリ変数には存在しなかったものであり、評価基準によっては読込み量より重要な意味をもつ代表値である。ある月における、各自のアクセス回数のデータを多い順に並べた順序統計量が図7である。

月あたりアクセス回数20以上のクラスA、10以上のクラスB、10未満のクラスCと三つに分け、当月と前月のアクセス回数の推移を追跡するのが第二の分析である。この評価指標、及び分析の観点は、本稿の目的とは多少相違するので、文献6、7を参照とし、詳細は割愛する。

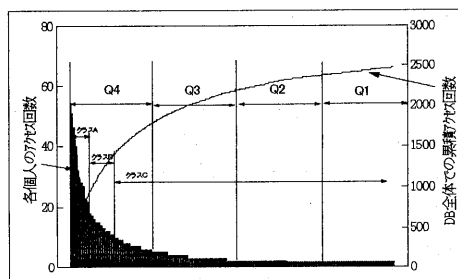


図7 利用者別アクセス回数の分布

4 ログ分析は有効であるか

「利用者離れ」は、利用者に対象とするDBがアクセスに値しない情報源と評価を下され、利用がなくなることとしよう。経営資源の投資評価という観点から、DB管理者にとり、この度合いを認知することは重要である。

ネット時代に留意すべきことは、ウェブ・イヤーやドッグ・イヤーなどの言葉もあるように、情報発信側のコンテンツ及び利用者による支持の具合は、例えば2、3ヶ月といった短期間で変化することである。

この時系列での変化を、第一のマクロ分析と第二のマイクロ分析の両面から把握できた。すると、DBはどのような利用状態にあり、どう評価されているかのDBの活性度合いといった質的な関係を推測できるだろう。前節で触れた分析などから、「アクセス・ログによる分析(以下、ログ分析と略)により利用者の利用行動を把握できる」との仮説は妥当性があると認められるだろう。

IV. 第二ステップ：「比較分析」

利用者の利用実態を数量的に把握できるとなれば、比較のための分析を可能にする。図6の時系列比較は同一のDB内でのそれであったが、本章は分析対象を変えて比較検討をする。

1 ネット時代における少数派

どのグループにおいても、初級、中級のユーザーが多数いる一方、ごく少数の上級パワーユーザーも存在する。必ずしも彼らと同一人物であるとは限らないものの、極めて少数の人により、多くのDBの書込みや読込みを占めることが多い。Iフォーラムの例においても例外ではなく、クラスAにて総読込み量の約30%、クラスBまでで約60%の読込み量である。

それを伝統的な統計による平均や分散で推測すると、例えば図7のような分布においてクラスAは人数的に5%未満であり、統計的に特殊な人と排除されてしまう可能性がある。

ネット時代の個人の分布は「複雑なまま」に眺め、本質を見誤らないような留意が必要になる。さらにネットワークの中心性の側面も忘れてはならない。瀬尾はインターネットのニュース・グループの投稿行動のモデルを研究しており(文献8)、その成果との共通点もある。

また川上による電子会議を支えるROMとRAMの研究(文献2)においても、平均すると書込みと読込みの人数比率は、ほぼ1:10だとしている。その書込み者の中で、何度も書込みをする人の占める割合はさらに少なくなる。

2 フォーラムの特性を考える軸

Iフォーラムの事例でも、読込みと書込みの人数比率は約5-8%であり、パソコン通信での結果と少なくとも傾向は同一である。分析対象の視点としての一つの軸は、異なった結果をもたらす可能性のあるネットワークのオープン、クローズ性をおこう。オープンなネットワークとは、パソコン通信やインターネットのような利用者が、ある条件を満たせば参加を拒むことのできないものをいい、クローズとは、その組織員になることが必要で、利用者からの選択肢は基本的にはない企業内などの閉じたネットワークとする。

さらに、Ⅲ. 1で触れた掲示板型、電子会議型の相違としてのインタラクティブ性をもう一つの軸として分類したものを図8に示す。

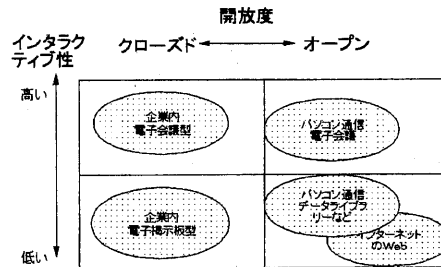


図8 フォーラム特性の4つのタイプ

この四象限でとらえられるとしたら、それぞれの分析対象において相違はあるかは、興味深い検討課題となる。そのクローズ間での利用者分布の比較を次節で示す。あわせて、少数派を顕在化させる試論でもある。

3 アクセス回数の分布の近似関数

ここで再度図7の各個人のアクセス回数分布の形状に着目しよう。この減少関数は、利用者の参加の広がり意味する傾きと、最高頻度のアクセスにより表わせるだろう。そこで「どのDBにおいても、ネット時代における利用者のアクセス回数分布は、規則性を持つ減少関数となる形状になるのではないかと第二の仮説を設定する。暫定的にその関数を、 $Y = A \cdot X \log e(X) + B$ としよう²。

ここで第二象限の代表として、比較的書込みの多い電子会議型のDB1と、第三象限のそれとして読込みのみを許す掲示板型のDB2を考える。その各DB間の利用者の上位40人の累積アクセス回数を、相対度数として作成する。その比較したグラフが図9である。

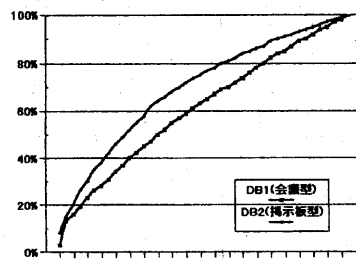


図9 フォーラムタイプによる利用者分布形状の相違

少なくともクローズドなネットワークでは、

² 図7は $Y = -14.3 \log e(X) + 68.2$ に近似

数量的な関数として利用者の分布形状を表現し、比較することができそうである。

なお本章で示した論点の内容に関しては、パラメータの意味や、統計的な裏付けをさらに加え、別の機会に論じたい。あくまで本稿では、ログ分析で利用状況が数量化されたことにより、客観的な尺度にて比較を可能にする方法の側面について強調をするものである。

V. 調査方法におけるログ分析

ここで、ログ分析の方法は、従来からある調査技法とどのように異なるかを整理してみよう。E. F. ストーンは、組織行動の調査に関するいくつかの調査戦略に関する評価の次元を整理している(文献9)。それをもとに、記録型の間接的測定方法であるログ分析での調査方法の特性をつけ加えたものが表2である。実験的な方法ではないため、独立変数としての操作可能性に欠ける。しかし、情報発信側にとられた活動の結果を、時系列変化を通じた時間遅れにて知ることができる。また一回の事例研究やログ分析のみでは、一般化できるものではない点を、比較分析によりある程度補うこともできる。

ログ分析による調査方法は、今後のネット環境の整備に伴い、事例研究とフィールド調査の両方の特徴をもつ方法との認識が強まると思われる。

| 評価次元 | 実験室実験 | シミュレーション | フィールド実験 | フィールド調査 | サンプリング調査 | 事例研究 | ログ分析による調査方法 |
|-----------------|-------|----------|---------|---------|----------|------|-------------|
| 最初の準備費用 | M | L-H | M-H | M-H | H | L-H | H |
| 対象一人あたりの限界費用 | L | L-H | M | M | L-M | L-H | H |
| 独立変数の強度 | L | L-M | M | H | H | H | H |
| 変数の範囲 | L | M | M | H | H | L-H | H |
| 独立変数の操作可能性 | H | M-H | M | N | N | N | N |
| 因果仮説の検証可能性 | H | M-H | M | L | L | N | H |
| 研究が調査者を変化させる可能性 | L | L | M | M | L | H | H |
| 混同変数のコントロール可能性 | H | M-H | L-M | L | L | N | H |
| 実験者期待効果の可能性 | H | M | M-H | M | L | H | H |
| 要求特性の可能性 | H | M | M-H | M | L | H | H |
| 評価意識の可能性 | H | M-H | M-H | M | L | H | H |

| | | | | | | | |
|----------|-----|-----|---|---|---|---|---|
| 状況の自然性 | L | M-H | H | H | H | H | H |
| 行動の状況依存度 | H | L-M | H | H | L | H | H |
| 一般化可能性 | L-H | L-H | L | L | H | N | N |

N:なし, H:高い, M:中, L:低い

表2 調査方法

VI. おわりに

最後に今回の事例と絡め、比較分析以外の応用例に触れよう。

第一は、競合あるいは補完するDBとの関係である。本稿は、情報発信側のDBに立脚した分析フレームであった。他のDBも選択可能である利用者の立場に視点を移した分析をできれば、より広義な利用者像を把握できる。すると「利用者離れ」などの現象を検証できる可能性も高まる。

第二には、利用者のアクセス日時から、サーバーのキャパシティ・プランニングや、利用者予測のマクロなモデルを立案できる。それに対応した効果的な情報発信についての知見も得られるであろう。

ログ分析の考え方は、具体的なカテゴリーは別としても、汎用的で、かつインターネットやイントラネット、そしてエクストラネットまで広い範囲で適用が可能なものである。本稿で見えてきたような時系列や他の分析対象と比較検証を可能にするログ分析は、今後のネット時代の組織行動を知る上で貢献できる有効な分析・調査方法の一つと結論付けられるだろう。

最後に、本稿作成に当たってコメントを頂いた諸氏各位に深く感謝いたします。

参考文献

- [1]西垣通, 組織とグループウェア, NTT出版, 1992
- [2]川上善郎他, 電子ネットワークの社会心理, 誠信出版, 1993
- [3]池田謙一, ネットワーキング・コミュニティ, 東大出版会, 1997
- [4]橋元良明, コミュニケーション学への招待, 大修館, 1997
- [5]新村秀一, パソコン楽々統計学, 講談社, 1997
- [6]豊島雅和, イントラネットによる情報発信評価のための一提案, 研究計画学会, 1996
- [7]豊島雅和, インターネット情報発信の分析評価方法, 電子情報通信学会, 1997
- [8]瀬尾雄三, ネットニュースにおける投稿行動の解析, 筑波大学修士論文, 1996
- [9]E. F. ストーン(野中郁次郎他訳), 組織行動の調査方法, 白桃書房, (1978)1980