

メモ情報に基づく情報流通基盤の提案と実現

田中基貴[†] 河合栄治[‡] 藤川和利[§] 砂原 秀樹[§]

概要: メモや作業ログの様な関心の有無により情報の価値が大きく変化する情報 (以下概要では、情報と呼ぶ) は個人の PC に蓄積されることが多い。しかしながら、それら蓄積された情報は自己を対象として作成されるため、万人にとって価値があるものではないが、関心が類似する他者にとっては関心の拡大や新たな発想のきっかけになりうる貴重な情報源である。そこで本研究では、関心の類似性を考慮した情報源のグルーピング、価値付けを行うことで、それらの蓄積された情報から各人にとって有用な情報を発見、収集するシステムを提案する。具体的には、まず、それらの PC による P2P の情報流通基盤を構築することで、情報の流通、情報流通経路の特定を可能にする。次に、各人の関心をローカルに存在する情報への参照回数をもとに情報の集合として抽出する。抽出された情報とそれらの流通経路をもとに、関心が類似する他者のカテゴリ (人により分類されたメモ情報の集合) をグルーピング、グループ内での格付けを行う。これにより選出されたカテゴリを情報源とすることで有用な情報の発見、収集を実現する。

キーワード: メモ情報、情報流通基盤、P2P、情報収集

The Proposal and Realization of an Information Exchange System based on Knowledge-Memos

Noritaka Tanaka[†] Eiji Kawai[‡] Kazutoshi Fujikawa[§] Hideki Sunahara[§]

Abstract: Knowledge-Memos is very useful for those who are interested in such memorandum. But, Memorandum created on PC is hard to be opened to the public by several reasons. So, The present condition that memorandum is not utilized is raised. In this paper, we propose information exchange system in order to solve such problem. It system consists of the information circulation mechanism using peer-to-peer technology and information gathering mechanism in consideration of concern.

Keywords: knowledge-memos, information exchange system, p2p, information gathering

1 はじめに

近年、個人が PC を保有することが一般化したことにより、メモや作業ログといった情報 (以下メモ情報と呼ぶ) を気軽に PC に蓄積する機会が増加している。ま

た、メモ情報は関心が類似する他者にとっては、関心の拡大や新たな発想のきっかけになる重要な情報であることがある。しかし、メモ情報は基本的に自己を対象とした情報であるため、発信のために情報の整形、補完がなされず、個人の PC に蓄積される傾向にある。つまり、関心の有無により情報の価値が大きく異なる情報が分散して存在しているが、それらは有効に活用されていない。そこで、本研究では、それらの PC による P2P の情報流通基盤を構築することで、情報流通を可能にする。また各人にとって有用なメモ情報のみを発見、収

[†] 奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, NAIST

[‡] 奈良先端科学技術大学院大学 附属図書館
Digital Library, NAIST

[§] 奈良先端科学技術大学院大学 情報科学センター
Information Technology Center, NAIST

集する機構を構築することでメモ情報の有効活用を目指す。

以下、本稿の構成を示す。2. 節では、メモ情報と既存の情報流通基盤について検討する。3. 節では 2. 節をもとにメモ情報に適した情報流通基盤に必要な機能について分析する。次に 4. 節で 3. 節を満たす情報流通基盤の構築に関する提案をする。5. 節では本研究に関連する研究について検討する。最後に 6. 節で本稿のまとめと、今後の課題について述べる。

2 既存の情報流通基盤

一般に情報の価値は、情報消費者の状況や関心などに左右される。また、各人の状況や関心は多種多様である。そのため、各人が価値ある情報を得るためには多種多様かつ多数の情報が必要である。しかし、情報が多種多様かつ大量になると、価値ある情報が無価値な情報に埋もれてしまうといった問題が生じる。そこで、各人にとって価値ある情報を効率よく収集する機構といったものが必要となる。メモ情報もその例外でなく、メモ情報を活用するには以下の 2 点を満たす必要がある。

- 多種多様かつ多数の情報が取得可能である
- 有用な情報を収集する機構が存在する

そこで、この 2 点とメモ情報の特徴に留意し既存の情報流通基盤を検討する。

2.1 WWW (WorldWideWeb)

WWW は、現在最も多種多様かつ莫大な量の情報が流通している情報流通基盤である。その特徴として広く知らせたい情報や共有したい情報の伝達に適している。

ここで、メモ情報が WWW に適しているかを考える。WWW では基本的に情報を HTML に整形し、サーバーへアップロードすることから情報流通が始まる。しかし、メモ情報は自分を対象とした情報であることから、整形の手間、アップロードの手間を掛けてまで発信することはない。また、WWW では発信した情報へのアクセス制限が柔軟とはいえない。このことより、発信する情報に暗黙のうちに制限が加わることがある。

次に、メモ情報が WWW に存在している状態を想定し、有用なメモ情報を収集できるかについて考える。WWW 上での主な情報収集手段として Google[1] などの検索エンジンが挙げられる。Google は単なるキーワードだけの検索結果に満足できる結果が得られない

として、リンク構造解析 (PageRank[2]) による WEB ページの評価を併用している。これにより、基本的にキーワードを含み、かつ一般に重要と判断されるものを優先的に提供している。しかし、メモ情報の有用性は、個人の関心により大きく変化する。つまり、一般的に重要な情報が、必ずしも個人にとって有用とは限らない。このことから、WWW はメモ情報を扱うのに適しているといえない。

2.2 Mail, Instant Messenger

Mail や Instant Messenger は、相手を指定し、相手と一対一の情報流通を基本とする情報流通基盤である。メモ情報をこれらの上で発信する場合、自分のために書かれたメモ情報に相手を意識した加工が必要である。これは、メモ情報の発信を促進する上で好ましくない。また、基本的に情報流通範囲が知人に限定されるため、多種多様な情報が流通するとは言い難い。

情報収集の観点からこれらを検討する。これらは、もともと情報を相手に伝達することを目的としているため、やりとりされた情報は当事者に蓄積され、第三者がそれらの情報を活用する事は考えられていない。

2.3 ML、掲示板、チャット

ML(メーリングリスト) や掲示板、チャットは、コミュニティを形成し、その中での情報流通を基本とする情報流通基盤である。そこには関心が近い人が集まり、関心ある話題の情報が流通するため、有用な情報が飛び交うと言える。しかし、コミュニティでは関連する話題の範囲内での発信が推奨されるため、情報を発信するにはその情報に適したコミュニティを選択、もしくは発見する手間が発生する。また、発見されない場合は情報が発信されないといったことも起こりうる。

メモ情報を扱う場合においても、コミュニティにあった情報の整形、コミュニティ選択もしくは発見といった手間が発生する。これはメモ情報の発信を阻害すると考えられるので、メモ情報を扱う上で適しているとはいえない。

2.4 P2P 技術

高速、常時接続のネットワークの普及により、Gnutella や WinMX、Winny といった P2P 技術を利用した情報流通が増加している。これらは WWW と異なり、基本的に情報発信においてサーバーを必要としない。そのため、WWW でメモ情報を扱う時の問題点で

あった手間を削減できる。しかし、WWW と同様に公開した情報へのアクセス制限が十分でないことで生じる問題がある。また、これらの情報流通基盤における情報収集機構も WWW と同様にキーワード検索が一般的である。そのため、メモ情報を扱う上では必ずしも良い情報収集機構であるとは言えない。

3 要求分析

2 節で述べたように、既存の情報流通基盤をメモ情報の有効活用といった観点から捉えると満足いくものではない。そこで、本節ではメモ情報を有効活用するために満たすべき条件について検討する。

3.1 メモ情報の発信について

メモ情報は自己を対象とした情報であり、手軽に作成できるといった利点がある。この利点が崩れると情報の生成、発信が阻害される。言い換えると、流通するメモ情報の種類や量の低下を招き、メモ情報の有効活用にとって好ましくない状態となる。つまり、メモ情報を扱う情報流通基盤は情報発信に関する手間がかからないことが必要となる。

また、メモ情報はプライベートな内容を扱うことがある。プライベートな内容を含むメモ情報を発信する時、ある人になら見れても良いが、他の人には見られたくないといった要求が挙げられる。この要求を解決するため情報流通範囲を柔軟に設定できる必要がある。

3.2 メモ情報の収集について

WEB ページなどの公開を前提としている情報は、不足した情報の補完、情報の体系化がなされていることが多い。そのため、それらの情報は関連する知識を持っていない人に対しても有用なことがある。一方、メモ情報は、対象とする物事の核となる情報だけが記されていることが多い。そのため、メモ情報はそれらを補完する知識と結び付くことで始めて価値を生み出す。この違いは各人にとって有用な情報を収集する場面においても違いを生み出す。それは、有用なメモ情報を収集する場合は、情報収集者の知識範囲を強く意識する必要がある点である。ここで、知識範囲は各人の関心と密接に関係していることを考慮すると、有用なメモ情報の収集には関心を考慮した情報収集機構が必要である。

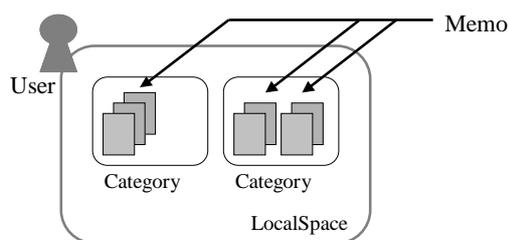


図 1 メモ情報とカテゴリーのイメージ図

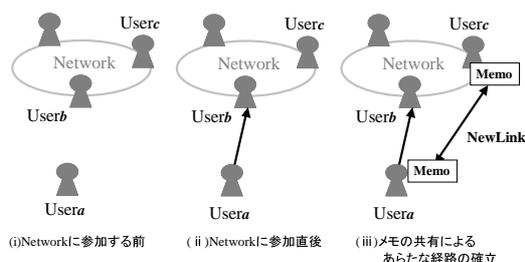


図 2 メモ流通による情報流通基盤の構築

3.3 要求分析のまとめ

これまでの要求をまとめると、メモ情報の有効活用を促進する情報流通基盤には、以下の三点を満たすことが要求される。

- 発信に対するコストが少ない。
- 発信した情報の流通範囲の設定が柔軟である。
- 個人の関心を考慮した情報収集機構

4 メモ情報に適した情報流通基盤の設計

本節では、3 節で述べた要求を満たす情報流通基盤を提案する。基本的な方向性とし、メモ情報発信に関する条件を満たすために、メモ情報が生成、蓄積される PC をノードとする P2P な情報流通基盤を構築し、その上で個人の関心を考慮した情報収集機構を構築する。

4.1 P2P な情報流通基盤

メモ情報の流通には発信に関する負担を極力減らす必要がある。そこで、メモ情報が作成された PC がそのまま発信環境となるような、P2P な情報流通基盤を構築する。これにより、WWW のようなサーバが不要となりアップロードなどの手間も削減できる。

ここでは、具体的な P2P の情報流通基盤の構築について述べる。通常、個人は複数の事柄に関心を持ち、その関心ごとにメモ情報を分類している事が一般である。そこで関心ごとの分類をカテゴリーと呼ぶ(図 1)。

表1 属性と流通範囲

属性	流通範囲	二次流通
<i>public</i>	全ユーザ	可
<i>private</i>	自分	不可
<i>private</i> + [流通先]	自分 + 流通先	不可

ここで、 $User_a$ を例にとり (図 2-(i))、情報流通基盤の構築について述べる。まず $User_a$ は既にメモ情報流通網上に存在する $User_b$ を指定することで情報流通基盤に参加する (図 2-(ii))。次に $User_a$ は 4.3 節で説明する情報源選出手法により、有用な情報源を選出する。選出された情報源よりアクセスが許可されているメモ情報を取得する。 $User_a$ は取得したメモ情報を有用だと判断すると、自分のローカルに存在するカテゴリーに分類し保存する。この時、 $User_a$ のシステムは保存したメモ情報がどこから入手したメモ情報であるかを記録する。また $User_c$ のシステムもメモ情報が $User_a$ により保存されたことを記録する。これにより、 $User_a$ と $User_c$ 間に新たな流通経路が確立したとする (図 2-(iii))。ここで確立された流通経路は次回の情報源の収集に利用される。このようにメモ情報の流通により新たな流通経路を確立していくことで情報流通基盤を構築する。

4.2 流通範囲の制御

上記のような P2P な情報流通を考えると、メモ情報は拡散していく。しかし、メモ情報にはプライベートな内容を含んだメモ情報も存在するので、それらの拡散を防止する仕組みが必要である。そこで、メモ情報の公開に対する属性を明確に区別することで、二次的な流通が可能であるかどうかを決定し、メモ情報の拡散を制御する。ここでは、属性として“*public*” と“*private*” を定義する。メモ情報に“*public*” の属性が与えられた場合、そのメモ情報は流通した先で属性の変更が可能とする。つまり、流通したメモ情報の流通範囲の設定は流通した先に任せる。“*private*” の属性が与えられた場合は、その他に流通先も設定する。この場合は、流通先に設定された人にだけに流通が可能であり、流通した先での流通範囲の再設定は不可能とする。流通先が設定されなかった場合は、メモ情報は全く流通しないとする。これら属性の設定により、公開可能なメモ情報やプライベートな内容を含むメモ情報が統一的に扱え発信に対するコストが低減する。

表2 関心が類似するカテゴリーリストの例

カテゴリー	カテゴリー内のメモ情報	類似度
$Category_a$	$Memo_1, Memo_2, Memo_3$	45
$Category_b$	$Memo_1,$	20
$Category_c$	$Memo_2, Memo_3$	25
$Category_d$	$Memo_2$	15
$Category_e$	$Memo_3$	10

4.3 関心を考慮した情報収集機構

有用なメモ情報を収集するには、関心を考慮した情報収集機構が必要である。そこで本研究では、関心が類似する他者が保有するメモ情報は有用であるといった経験則的な方法を用いる。具体的には、情報源として人を対象とすると、人が複数の関心を持つことから、提供するメモ情報に関心のないメモ情報が含まれることが考えられる。そこで、最も関心が類似するカテゴリーを選出し、そのカテゴリーに属するメモ情報を提供する手法を提案する。以下、カテゴリーの選出手法について述べる。

4.3.1 関心の表現方法

関心が類似する他者を発見するには、自己の関心を適切に表現する必要がある。関心の表現が正しくなければ、関心が異なる他者からのメモが増加し、不要な情報の提供が大量に提示されてしまう。

関心を表現する最も単純な方法は、キーワードを基にプロフィールを作成することで関心を表現する方法である。しかし、この方法ではキーワードは表現力に乏しい上、同一キーワードでも異なった意味で解釈される事がある。また、関心は時間の経過により変化するため、その変化へ対応することが難しい。このような観点から、キーワードによる関心の表現は適切であるとはいえない。

そこで本研究では、保有するメモ情報により関心を表現する。メモ情報はキーワード以上の情報を含んでいるため関心をより適切に表現できる。また、一定期間毎にメモ情報の参照回数を記録し、参照回数の多いものを対象とすることで現在の関心を表現できる。

4.3.2 関心が類似するカテゴリーの選出手法

関心が類似するカテゴリーを発見するには、カテゴリー間の類似度を測定する必要がある。そこで、本研究

では同一のメモ情報を多く含むカテゴリーほど関心が類似すると考えた。

具体例をもとに関心が類似するカテゴリーの選出方法を説明する。はじめに 4.2.1 節の関心の表現で選ばれたメモ情報を $Memo_1, Memo_2, Memo_3$ 、そのメモ情報への参照回数が 20,15,10 回であるとする。次に、 $Memo_1$ を含むカテゴリーをネットワーク上から選出しリストに格納する、残る $Memo_2, Memo_3$ についても同様にカテゴリーを選出しリストに格納する。これにより得られたリストに含まれるカテゴリーを $Category_a, Category_b, \dots, Category_e$ とする。次に、リストに含まれるカテゴリー毎に含まれる $Memo$ の参照回数を足し合わせ、そのカテゴリーの類似度とする (表 2)。類似度の高い N 個に関心が類似するカテゴリーとして選出する。表 2 の場合、 $Category_a, Category_c, Category_b, \dots$ の順で類似度が高いことがわかる。

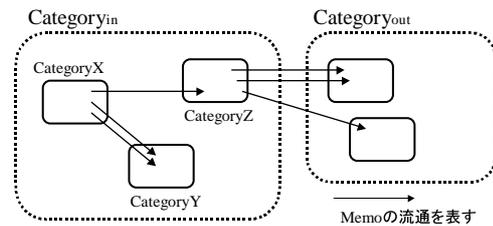
以上の方法では、適度にメモ情報が共有されている必要がある。あるカテゴリーでメモ情報が他のカテゴリーと共有されていない場合、初回のみキーワードによる検索などを用いて情報を収集し、共有する必要がある。

4.3.3 流通経路を用いたカテゴリーの選出方法

4.3.2 節で選出したカテゴリーは関心が類似するメモ情報を含んでいる可能性は高い。しかし、一つのカテゴリーに多種多様なメモ情報を含んだカテゴリーなど、雑多な情報を多く含むカテゴリーを選出する可能性がある。そこで、選出されたカテゴリーが有用な情報源としての信頼性があるかどうかを決定する必要がある。ここでは、その手法について述べる。基本的な方針として、あるカテゴリーに含まれるメモ情報がどのような流通をしたかを統計的に扱うことで信頼性を決定する。

ここで、 $Category_{local}$ での情報収集について考える。まず 4.3.2 節の手法で選出されるカテゴリーの集合に関心の範囲にあるカテゴリーといった意味で $Category_{in}$ 、それ以外を $Category_{out}$ とする。次に $Category_{in}$ からカテゴリーを一つ取り出し $Category$ とする。 $Category$ には $Category_{local}$ と共有されていないメモ情報が N_{all} 個存在するとする。 $Category_{local}$ の有用な情報源は、関心が類似する $Category_{in}$ も同様に有用な情報源と考えられる。このことより、 $Category$ に含まれるメモ情報が $Category_{out}$ より、

$Category_{in}$ に流通しているほど $Category$ の信頼性は高いと考えられる。そこで、 $Category$ に含まれるメモ情報毎に流通経路を調べ $Category_{in}$ へ流通している割合を求める。その結果が α 以上であった場合、そのメモ情報を有用である可能性が高いと判断する。この計算を $Category$ の他のメモ情報についても計算し、 α 以上であったメモ情報の個数を N_{good} 個とする。この時、 $Category$ は有用である可能性が高いメモ情報を N_{good}/N_{all} の割合で保有している。この割合をカテゴリーの信頼性の指標とし、割合が大きいカテゴリーを情報源として選出する。



※上図では、Category ZよりCategory Xの方が信頼性がある情報源である。
(Categoryには同数のMemoが存在すると仮定)

図 3 流通経路によるカテゴリ選出の例

5 関連研究

本研究が目的とするメモ情報の活用を対象とした研究は見当たらない。そこで、本研究における提案手法に関連する研究について比較、考察する。

メモ情報を扱う研究として [3] がある。[3] はメモを情報と情報の関連付けに用いることで、メモを通じた有用な情報の発見を支援している。本研究はメモそのものの活用を目的としている点で異なる。しかし、メモ情報が新たな発想や関心の拡大といった有用性を有することに着目した点において類似している。

発信されにくい情報の流通促進を目的とする研究として [4][5][6] がある。これらの研究はメモ情報の発信を促進することに注目している。また、[5][6] においては発信された情報に対するキーワード検索機能を提供している。本研究では、メモ情報の有効活用においてはキーワード検索は不向きであると考え、関心の類似、情報源の選出といった手法を採用した点で異なる。

本研究で採用した P2P での情報流通モデルに関係する研究として [7][8][9] が挙げられる。それぞれ有用な情報の流通を目的とし、流通時に情報受信したユーザによるフィルタリングを行っている。そのため、多くの人

が興味を持つ情報や、信頼性が高い情報を流通させるには適しているが、メモ情報など小人数が興味を持つ情報を流通させるのは困難である。本研究では、関心に基づく情報源の選出を行い、さらに選出された情報源の性質を調べ、各人にとって有用な情報源を厳選する。これにより、有用なメモ情報の流通を可能としている。

興味や関心を考慮した情報提供手法として、協調フィルタリングによる情報推薦が存在する。しかし、これらの手法では興味、関心が異なるユーザの評価も情報推薦に影響する問題が存在する。本研究では、関心が類似するユーザの情報のみを用いるのでこのような問題を回避している。

6 おわりに

本稿では、個人の PC に蓄積されたメモ情報の有効活用のため、蓄積されたメモ情報発信の活性化、メモ情報の関心を考慮した情報収集方法を提案した。

現在、提案手法を実証するためプロトタイプを用いた研究室の学生を対象とした評価実験を行っている。今後、実験参加者にアンケートを実施し、研究室で活用されている情報流通基盤と提案システムの比較、各人が受信したメモ情報の有用性といったものを調査することで提案手法の有効性の検証を行っていきたい。

参考文献

- [1] <http://www.google.com/>
- [2] Page,L., Brin,S., Motwani,R., Winograd,T., “The PageRank Citation Ranking: Bringing Order to the Web” 1998
- [3] 梅田恭子, 安田孝美, 横井茂樹, “知識メモを活用した研究情報共有方式の提案”, 情報処理学会論文誌 Vol.42 No.11 p2562-p2571(2001)
- [4] 佐藤彩子, 河合栄治, 藤川和利, 砂原秀樹, “個人の知識情報を対象とした情報発信・流通モデルの提案”, 電子情報通信学会 第 14 回データ工学ワークショップ DEWS2003
- [5] 石井恵, 中渡瀬秀一, 牛島浩一, 佐藤哲司, “発信協調型ナレッジ共有システムの提案”, 信学技報 AI2001-48(2001)
- [6] 湯川高志, 大黒毅, 桑原和宏, “パーソナル・レポトリに対するピア・ツー・ピア型協調検索機構の提案”, 信学技報 AI2001-48(2001)
- [7] 竹内亨, 鎌原淳三, 下條真司, 宮原秀夫, “ユー

ザの関連性に基づいた情報伝播モデル”, 電子情報通信学会 第 12 回データ工学ワークショップ DEWS2001

- [8] 吉田匡志, 伊藤雄介, 沼尾正行, “口コミによる分散型情報収集システム-WAVE を起こそう”, 電子情報通信学会 第 10 回マルチエージェントと協調計算ワークショップ MACC2001
- [9] 猿渡俊介, “ピアツーピアネットワークによる情報伝播を用いたコンテンツ共有システム”, 信学技報 NS2001-275,IN2001-231(2002)