

パーソナル・レポジトリに対する ピア・ツー・ピア型協調検索機構の提案

湯川 高志[†] 吉田 仙[†] 桑原 和宏[†]

[†] 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
〒 619-0237 京都府相楽郡精華町光台 2-4

E-mail: †{yukawa,yoshida,kuwabara}@cslab.kecl.ntt.co.jp

あらまし 組織やコミュニティにおいて、メンバ個人の持つ情報を電子的に公開し相互に交換・共有することで情報や知識の流通を円滑にし、活動を活性化させることを目指した、個人間情報交換・共有システムのフレームワークについて提案する。従来のサーバ中心型システムで問題となる情報提供のボトルネックの解消を目指し、個人的な情報を蓄積するパーソナル・レポジトリとその内容の要求・提供を司るパーソナル・エージェントから構成されるシステムが、複数連携して動作するピア・ツー・ピア型のシステムを提案している。提案システムでは、検索の際の個々のパーソナル・レポジトリ内の概念が張るベクトル空間の等化と、情報の提供範囲を制御することが課題となることを指摘し、ベクトル空間を等化しつつ検索を行う協調検索の手法を提案している。

キーワード ピア・ツー・ピア, 情報検索, パーソナル・レポジトリ, パーソナル・エージェント, 協調システム

Toward Peer-to-Peer Collaborative Information Retrieval on Personal Repositories

Takashi YUKAWA[†], Sen YOSHIDA[†], and Kazuhiro KUWABARA[†]

[†] NTT Communication Science Laboratories, NTT Corporation
Hikari-dai 2-4, Seika-cho, Soraku-gun, Kyoto, 619-0237 Japan

E-mail: †{yukawa,yoshida,kuwabara}@cslab.kecl.ntt.co.jp

Abstract A framework on peer-to-peer information exchange and sharing system is proposed. The system aims to achieve smooth knowledge and information management to activate organizations or communities. Because conventional server centric systems are suffered by information providing bottleneck, proposed framework assumes collaborative work of personal repositories, which accumulate per-user information and control requests and provisions for it. Issues on the framework are pointed out. One is equalization for vector space, which compensates differences in vector space for each personal repository. Another is automated control of provision criterion of information. Collaborative information retrieval schemes which equalize vector space of each personal repository to obtain proper information are also proposed.

Key words peer-to-peer system, information retrieval, personal repository, personal agent, collaborative system

1. ま え が き

近年、パーソナルコンピュータの普及に伴い、多くの情報が電子的に作成・保存され、これらの電子的手段による伝送も容易に可能となった。組織やコミュニティの内部において、メンバの持つ情報を電子的に公開し相互に交換・共有することで、情報や知識の流通を円滑にし、組織やコミュニティを活性化させようという取り組みも盛んとなってきている。

このような組織やコミュニティのメンバ間の情報交換・共有を行うためのシステムのベースとして、World Wide Web(WWW)システムが流用されることが多い。WWWシステムでは、情報の提供者がその情報を決った形式で記述してサーバに格納するだけで、不特定多数の利用者だれもが同じ情報を同じように閲覧することが可能である。しかしながら、WWWのようなサーバ中心型システムでは、情報を持つ者がその情報を意識的にサーバに格納する必要があること、格納した情報は原則としてそのサーバにアクセス可能な利用者すべてに公開されることが原因となり、組織やコミュニティの他のメンバにとって有用な情報を持っているにも関わらず、それらを十分に流通させられない場合も多い。すなわち、情報の提供にボトルネックが存在すると言える。この情報提供のボトルネックを解決すれば、より多くの(少なくともそれを受取る者にとって)有用な情報が、組織内、コミュニティ内において、より円滑に流通すると考えられる。

そこで本稿では、情報提供のボトルネックを解決するシステムとして、個人的に格納した情報をピア・ツー・ピア型[1]で交換・共有するシステムのフレームワークについて提案する。上述したような組織やコミュニティ内での円滑な情報流通の観点から、ピア・ツー・ピア型情報交換・共有に求められる要件を洗い出し、それらを満たすシステムのモデルを構築する。さらに、このシステムにおける重要な機能要素である協調情報検索について、その機構を提案する。検索においては、WWW同様に自然言語で記述された文書に対する検索が中心となるため、問合わせの語が含まれている文書を探す全文検索や、問合わせと意味的に類似する文が含まれる文書を探す概念検索が必然的に用いられる。本稿では、ベクトル空間モデルによる概念検索を対象とする。概念検索について考察し、これをピア・ツー・ピア型に適用する際の重要な問題は、パーソナル・レポジトリ相互で概念が張るベクトル空間が異なる点であるこ

とを示す。これを解決するためには一方のパーソナル・レポジトリのベクトル空間を他方のそれに等化する必要がある。このための手法として、自動関連性フィードバック法と逆行列計算法のふたつを提案する。最後に、関連する研究に対する着眼点や特性の違いについて議論し、今後の課題を述べる。

2. 背 景

2.1 動 機

WWWのようなサーバ中心型システムを、1.で述べたような個人間の情報交換・共有に応用しようという場合、次のような問題がある。

- 情報を系統的に構成してサーバに格納(アップロード)することが前提であるため、情報を公開するための精神的な障壁が高い。このため、個人が持つ多くの情報のうち、本人が公開のための意識と意欲を持った一部の情報しか流通しない。

- 原則として、公開した情報はサーバにアクセス可能な利用者全員が閲覧できるため、組織やコミュニティの中でも特定の相手にだけしか渡したくないような情報は流通しない。

個人が持つ情報の中には、提案書のドラフト、アイデアのメモ、あるテーマに沿ってクリッピングされたニュース記事など、本人にとって完成された情報とは言えないが、他のメンバにとって物考える際のヒントとなるような情報が多く含まれている。これらは、上の前者の理由により、サーバ中心型システムにおいて流通することはない。また、電子メールのメッセージや営業情報のように、提供相手と情報の内容との関係に基づいて提供すべきかどうかが決るような情報も、個人が持つ情報には多く含まれている。例えば、個人的に受取った電子メール・メッセージでも、内容がアナウンス的なものならば、良く知っている相手には提供して良いかも知れない。本来はクローズドなメイリング・リストのメッセージでも、その内容に興味を持っていてかつ信頼できる知人には、他に公開しないよう注意書きを添えて転送することも良く行われている。このようなメッセージの交換・共有は、上記後者の理由からサーバ中心型システムでは困難と言えよう。

上記のような問題に縛られることなく組織やコミュニティにおいて個人間の情報を円滑に交換・共有するためには、メンバ各々が、自分が持つあらゆる情報を格納する記憶領域(パーソナル・レポジトリ)とそれに対するアクセスや制御を提供するソフトウェア(パーソナル・エージェント)から成るシステムを

持ち、それらが協調的に動作して情報をやりとりする形態が適していると考えられる。このような形態であれば、情報をサーバにアップロードする必要がなく、個人が保有する情報のすべてを直接に交換・共有の対象とできる。また、同時に、パーソナル・エージェント同士がインタラクションを行い、相互の関係を認識するとともに、情報の内容からそれを提供すべきか否かを判断することが可能なため、その情報を必要としており、かつ、許し得る相手にだけ提供することが可能となる。このようなシステムでは、サーバや情報の仲介のみを行うブローカを介さずに、直接相互にインタラクションを行って情報を交換することになるため、これをピア・ツー・ピア型情報交換・共有システムと呼ぶことにする。

2.2 課題

上述のように、ピア・ツー・ピア型の情報交換・共有システムは、情報提供のボトルネックを解消し、より円滑な情報の流通を可能にする素地を持つシステムである。しかしながら、個人が主として自身のために蓄積した情報を扱うことから、その実現には、いくつかの課題を解決する必要がある。

まず、情報を交換・共有するためには、要求者が所望する情報を探し出さなければならない。サーバ中心型システムは、同じ要求に対しては同じ結果を返す動作が前提であるし、情報は公開を前提としているため整理された形式で文体も主として文語体で記述される。このため、検索要求の単語や文字列が文書に含まれているか否かに基づく全文検索によってある程度有用な情報を探し出すことができたし、利用者もそれで満足していた。一方、ピア・ツー・ピア型システムでは、個々のパーソナル・レポジトリには、それぞれの所有者の専門、趣味、嗜好など(これをまとめて世界観と呼ぶことにする)に基づいて情報が蓄積されている。これらの情報は、自然言語で記述された文書とはいえ、所有者の違いにより言葉遣いが異なる場合も多い。また、パーソナル・エージェント同士が連携して検索を行うのであるから、それぞれの個性に則したより有用な情報が探し出されることを利用者も期待する。このため、ピア・ツー・ピア型情報交換・共有システムでは、それぞれのパーソナル・レポジトリが構築された際の世界観を互いに擦り合わせるような協調動作を行う検索手法が必要となる。

また、ピア・ツー・ピア型システムでは、個人的な情報を扱うことから、情報の提供範囲を適切に制御しなければならない。すなわち、ある情報がある相

手に提供するか否かを、その情報の内容と相手との関係とから判断する手法が必要である。また、提供の形態についても制御の必要がある。すなわち、情報すべての提供、一部分の提供、部分的に隠蔽した上での提供、所在のみの提供などの提供形態があり得て、やはり内容と相手との関係とから、どの形態により提供するかが決定されることになる。このような提供範囲や形態の制御を、正確かつ効率的に行うための機能が必要となる。

3. パーソナル・レポジトリと概念ベースを用いた情報検索

3.1 パーソナル・レポジトリ

情報を蓄積し検索や操作を行うシステムの中でも、個人が持つ情報に特化したものをパーソナル・レポジトリと呼ぶ。従来は、多くの人の持つ大量の情報を集積し一様に提供することを目指した、いわゆる図書館型の情報の蓄積・検索システムが多く開発されて来た。パーソナル・レポジトリは、これとは反対に、個人が持つ情報を自身が活用することを目指したもので、書斎型のシステムと言える。その考え方の提案から実際のシステム開発まで行っているプロジェクトとして Haystack [2] が知られている。

パーソナル・レポジトリが図書館型システムと本質的に異なるのは、検索や操作に際して、それを所有する個人にシステムが適応することにある。このために、情報を蓄積する機構(レポジトリ)として半構造データベースが用いられ、情報そのものにさまざまなメタデータを付加して蓄積できるようになっている。また、検索に際しても、所有者の意図を反映した検索を行う機構が用いられる。Haystack では、それ自体では検索機能を持たず、外部の検索機構を呼出す構成となっているが、筆者らは次節に述べる概念ベースを用いた検索が好適であると考えている。

3.2 概念ベースを用いた情報検索

検索対象の文書と検索要求である検索キーとを、それぞれ多次元空間上のベクトルとして表現し、両ベクトルがなす角の余弦値を関連度とするモデルはベクトル空間モデル [3] と呼ばれている。ベクトル空間モデルでは、何らかの適切な写像法によって検索対象を同一のベクトル空間にマッピングできれば、その対象がいかなるものであろうと、対象間の関連度が統一的な方法で計算できる利点がある。主として語を検索キーとして文書を検索する用途に用いられるが、対象のベクトル空間への適切な写像法さえ定義できれば、語を検索キーとして画像や音楽を検索

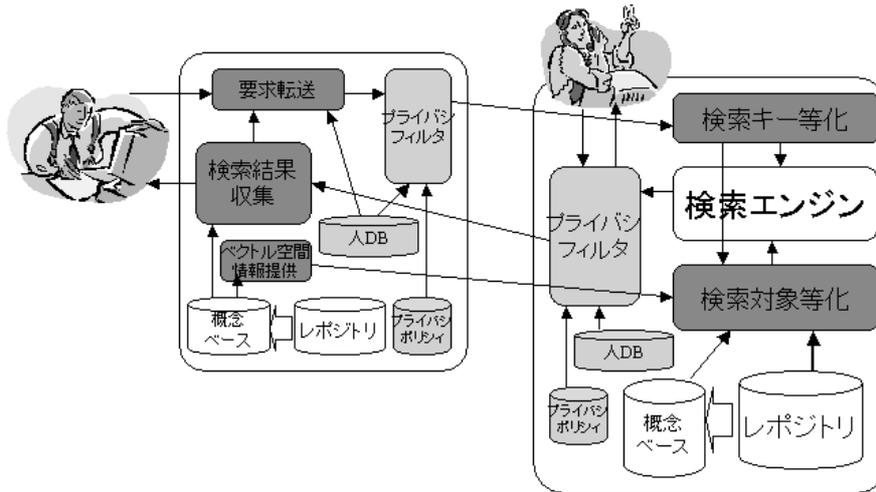


図 1 ピア・ツー・ピア型情報交換・共有システム
Fig. 1 Peer-to-Peer Information Exchange and Sharing System

したり、逆に、画像を検索キーとして文書を探すことも可能である。

ベクトル空間への写像でもっとも基本的な手法は、ターム（一般には文書群に含まれるすべての単語から助詞等のストップワードを除いたもの）の個々を多次元空間の個々の軸に対応させ、文書のベクトルをその文書が含まれるタームごとに $TF \times IDF$ 値で表現したものである。このような、タームが直接ベクトル空間の軸に対応する手法では、語が互いに独立となってベクトル空間を張るため、ターム間の類語関係や共起関係といった意味上の特性を反映した関連度判別ができないという問題がある。

これを解決し語の意味を考慮した関連度判別が可能な検索手法として、筆者らは概念ベースを用いた検索手法を提案してきた [4]。この手法では、まず、個々のタームをそれぞれ数百～数千次元のベクトルとして表現して（これを概念ベクトルと呼ぶ）、タームとその概念ベクトルの対応を記録した辞書（これを概念ベースと呼ぶ）を作る。そして文書を、その文書に出現するタームそれぞれの概念ベクトルをすべて合成したベクトルとして表現する（これを文書ベクトルと呼ぶ）。検索キーとしては、語や語の集合さらには文章までも取り得る。検索キーからタームを抽出して対応する概念ベクトルをすべて合成し、これを検索キーベクトルとする。検索の際には、検索キーベクトルと文書ベクトルとのなす角の余弦を関連度として計算し、関連度の大きい順に一定数を検索結果とする。この手法では、語も文書も同一のベクトル空間に配置される。

この手法にとって鍵となるのは、概念ベースの構築法であるが、二つの方法が提案されている。ひと

つは、国語辞書の見出し語と語義文との対応に基づく方法である [5]。タームに対する概念ベクトルを、そのタームの辞書上の語義文に出現する語との関係から定義する。もうひとつの方法は、検索対象文書群におけるタームの共起関係に基づいて構築する方法 [6] である。検索対象の文書群を通して同じように共起するターム同士が近接して配置されるようにベクトル空間が構築される。具体的には、まずターム数だけの行と列を持つ行列を作り、文書内で、あるターム A に対して他のターム B が近傍に出現した場合に、 A 行 B 列の要素にその近さに応じてスコアを追加する。すべての文書にわたってこのスコア追加を繰り返した後、特異値分解を用いた主成分分析によって列を 100～200 程度に圧縮する。これにより、個々のタームに対して 100～200 次元のベクトルをもった概念ベースが構築される。

前者の方法による概念ベースは普遍的ではあるが、固有名詞や辞書にない新たな語には対応できない。一方、後者では、これらに対応できるが、構築に用いる文書が異なると、タームが張るベクトル空間も異なってしまう。パーソナル・レポジトリでは、当然ながら専門的な文書、時事的な文書が多く格納されることになるため、本システムでは、後者の方法で構築した概念ベースを検索に用いる。

4. ピア・ツー・ピア型情報・交換共有システムの提案

4.1 ピア・ツー・ピア型情報交換共有システムのモデル

2.2 で述べたように、ピア・ツー・ピア型の情報交換・共有システムでは、パーソナル・エージェント

同士が協調して要求者の所望する情報を探し出す検索機能が必要である。また、情報の提供範囲を、情報の内容と相手との関係とに基づいて制御する機能も必要となる。これらを実現するものとして、図1に示すようなシステムのフレームワークを提案する。

情報を実際に蓄積するレポジトリ、レポジトリ内の文書から3.2で述べた方法により構築される概念ベースと、これに用いてベクトル空間モデルにより検索を行う検索エンジンは、スタンドアロンのシステムと共通である(図中塗りつぶしなし)。協調検索(図中濃い塗りつぶし)と情報提供範囲の制御(図中薄い塗りつぶし)については、節を分けて以下に述べる。

4.2 協調検索機構

個々のパーソナル・レポジトリは、上述のように、概念ベースを用いた検索機能を持っているが、概念ベースがそれぞれ異なるため、検索キーを単に他人のパーソナル・レポジトリに入力したとしても、望むべく結果が得られるとは限らない。

例えば、あるユーザ(ユーザA)が携帯型の情報機器についての情報を探しているとしよう。ここで、ユーザAは、ノートパソコンとPDAは類似のもののみなしているとする。すなわち、ユーザAのレポジトリには、“ノートパソコン”というタームと“PDA”というタームが近接して共起する文書が多く蓄積されているとする。このレポジトリから生成される概念ベースは、このふたつのタームの関連度が高くなるようなベクトル空間を張る。一方、他のユーザ(ユーザB)はPDAにはあまり興味がなく、その概念ベースは“ノートパソコン”と“デスクトップパソコン”が関連するようなベクトル空間を張っているとする。ただし、興味がないとはいえ、PDAに関する(ユーザAにとって有用な)情報もある程度は持っているものとする。

このような状況において、ユーザAが、自分自身のパーソナル・レポジトリから携帯型の情報機器に関する情報を検索するには、“ノートパソコン”を検索キーとすれば十分である。ところが、ユーザBのパーソナル・レポジトリに対して“ノートパソコン”を検索キーとして検索を行わせても、ユーザAが本来望んでいるPDAに関する情報は結果に含まれず、むしろ望んでいないデスクトップパソコンに関する情報が多く含まれることになる。

このような問題を解決するために、本稿で提案する情報交換・共有システムは、ベクトル空間を等化する機能を持つ。図1に描かれたベクトル空間情報

提供、検索キー等化、検索対象等化がこれを実現するための要素である。ベクトル空間を等化しつつ検索するための方式は、パーソナル・レポジトリ間の通信量、自己のベクトル空間の相手への開示の程度、等化の精度などにより、種々考えられる。本稿では、次章において、いくつかの方式を提案する。

4.3 情報提供範囲の制御機構

ユーザBのパーソナル・レポジトリ内部においてある情報が検索された場合、次には、その情報を情報要求者であるユーザAに提供して良いかどうかを判断する必要がある。前述のように、その判断は、情報の内容と、要求者との関係とに基づいて成される。また、どのような形態で提供するかも同時に判断する必要がある。

もっとも原始的な方法としては、要求のあった情報についてパーソナル・エージェントがユーザB(すなわち人間)に確認を行い、本人の判断に従って情報を提供することであろう。このような方法ではユーザの負担が大き過ぎるため、判断を自動化あるいは半自動化することも必然的に求められよう。

どのような方法を採用にしろ、情報提供範囲の制御を実現するためには、システムは次のような要素を持たなければならない。

- 相手に関する情報を格納する人データベース
- どのような相手に対してどのような情報を提供して良いか(あるいは悪い)かを規定するプライバシー・ポリシー
- プライバシ・ポリシー、情報の内容、および相手の情報に基づいて、情報提供の可否を判定するプライバシー・フィルタ

プライバシー・フィルタが、提供可否だけではなく自動判断不可という結果を出せるようになっていれば、判断が微妙な情報についてのみ人間が判断するといったことも可能となる。

5. 等化検索手法の提案

5.1 等化検索

4.2で述べたように、パーソナル・レポジトリごとに概念ベースが張るベクトル空間が異なるため、単に検索キーを転送しただけでは、情報を要求する利用者の意図を反映した検索結果を得られない。

これを模式的に表すと、図2のようになる。検索エンジンは、検索キーベクトル周辺で一定半径の超球内にある情報を探し出す動作をする。ユーザAのベクトル空間における超球は、ユーザBのパーソナル・レポジトリでは図に示すような超楕円体に写像

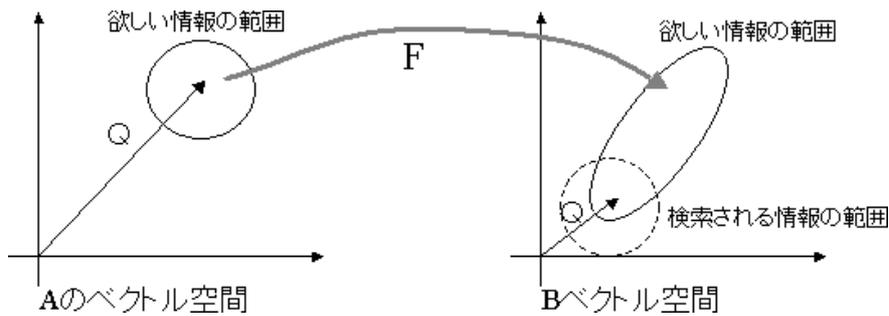


図 2 ベクトル空間の写像
Fig.2 Vector Space Mapping

され、検索キーベクトルも図の位置に写像されるとする。ユーザ B のシステムの検索エンジンは、自己のベクトル空間上で一定半径の超球内にある情報を探すため、ユーザ A が意図した領域 (すなわち超楕円体) とは異なった情報が結果として得られることになる。

より適切な検索結果を得るためには、本来意図した領域の情報がより多く入るように、ベクトル空間を等化しなければならない。以降の節では、これを実現する 2 つの方法を提案する。

5.2 関連性フィードバックによる検索キーの等化

図 3 に示したように、検索キーベクトルを超楕円体領域の重心に近づければ、検索領域が依然として超球のままであっても、適切な情報が結果に含まれる比率は大きくなる。これはすなわち検索キーベクトルの等化であり、関連性フィードバックとして知られている手法の応用とも言える。関連性フィードバックは、検索機構の提示する結果に対して、それが望ましい情報であるか否かを人間が判定して示す、人間から機械へのフィードバックであった。一方、本提案は、あるシステムでの検索結果が、検索を要求したユーザにとって望ましいか否かを、そのユーザの概念ベースを利用して自動的に判定しフィードバックをかけるものである。すなわち、パーソナル・レポジトリ間でフィードバックをかけることになる。

本手法では次のような手順で検索を行う。ここで、要求元のユーザが持つシステムを P_A 、情報を検索して提供する側のシステムを P_B と表す。

(1) ユーザが検索要求を発行すると、 P_A は P_B に検索キーを転送する。

(2) 検索要求を受けた P_B では、自己の概念ベースを用いて検索キーに対応する検索キーベクトルを計算し、それに基づいて検索を行い、上位の一定件数を P_A に返信する。

(3) P_A は検索結果を受取った後、自己の概念

ベースに基づいて検索キーと検索結果の個々の情報との関連性を計算し、それぞれの情報が適合しているか否かの判定を P_B に通知する。

(4) P_B は、判定結果に基づいた関連性フィードバックにより検索キーベクトルを調整し、再び検索を行う。

(5) 3 と 4 を適当な回数繰り返す、検索結果を得る。

本手法では、要求元は判定結果のみを通知するため P_B が P_A の概念ベースの全容を知ることは困難である。一方、 P_B は、適否を判定をさせるために、検索結果に含まれる多くの情報を P_A に提供することになる。つまり P_B は、 P_A が本来知らなくとも良い (要求したものではない) 情報まで提供してしまうことになる。また、図 3 からわかるように、検索結果領域となる超球の中心を調整するだけなので、 P_A における超球が P_B において扁平な超楕円体に写像される場合には、適切な情報が含まれる比率はあまり高くない。

5.3 逆行列計算による概念空間等化法

P_A のベクトル空間と P_B のベクトル空間とは何らかの写像関係にある。この写像を線形変換として近似することにする。概念ベース生成の元となる文書群が、自然言語という大きな枠として共通のルールで記述されていることと、その生成過程より、この近似によっても致命的な誤差は生じないだろう。

図 4 に示したように、 P_A から P_B が線形変換 F で表せるとすると、逆変換 F^{-1} がわかれば、 P_B 上の情報に対する P_A のベクトル空間での位置を推定することができる。

具体的には、次のようになる。いま、 P_A 上のあるタームの概念ベクトルを v とし、同じタームの P_B 上の概念ベクトルを u 、変換行列 F とすると、

$$u = vF$$

と書ける。ここで、ベクトルの次元数を M とする。

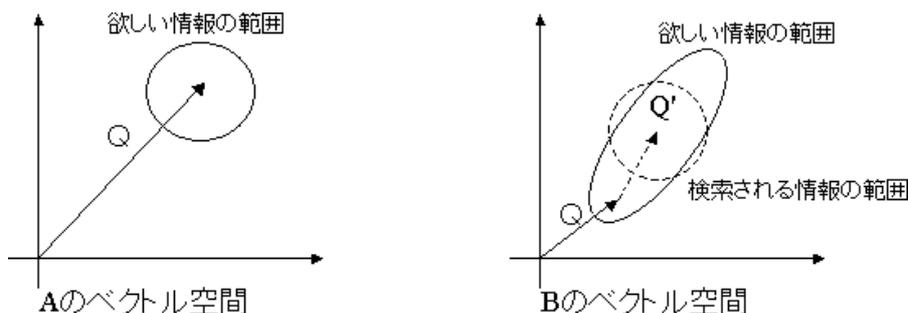


図3 関連性フィードバックによるベクトル空間の等化
Fig.3 Automated Relevance Feedback Method

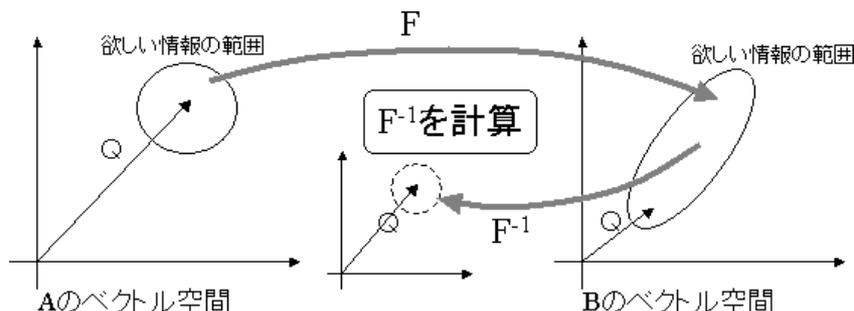


図4 逆行列計算によるベクトル空間の等化
Fig.4 Inverse Matrix Method

当然ながら F は $M \times M$ の正方行列となる。 P_A から M 個のタームを選択し、それらの概念ベクトルを v_1, v_2, \dots, v_M とする。ただし、タームを選択する際は、これら概念ベクトルが一次独立となるようにする。個々のタームの概念ベクトルを1行として、すべてを並べて作った $M \times M$ の行列を V と書こう。一方、それぞれのタームに対応する P_B での概念ベクトルを u_1, u_2, \dots, u_M とし、これらを行として並べて作った $M \times M$ の行列を U とする。 V と U との関係は、

$$U = VF$$

となる。

ここで、 V の逆行列 V^{-1} を計算できたとすると、

$$V^{-1}U = V^{-1}VF = IF = F$$

となる。ただし I は $M \times M$ の単位行列である。

このことから、 P_B は、以下のような手順で、変換行列 F を推定することができる。

- (1) P_A から概念ベクトルが一次独立であるような M 個のタームとその概念ベクトルを得る。
- (2) 概念ベクトルを行として並べて作った行列 V の逆行列 V^{-1} を計算する。
- (3) 自身の持つ概念ベースでの対応するタームの概念ベクトルを行としてならべて行列 U を作る。
- (4) $F = V^{-1}U$ として変換行列 F を推定できる。

変換行列 F が推定できれば、これの逆行列 F^{-1} を用いて、図4のように、 P_B における概念ベクトルや文書ベクトルが P_A においてはどのような値になるかを計算できる。

本手法では、もしも P_A と P_B のベクトル空間の関係が真に線形変換となっていれば、 P_B に蓄積された情報を P_A の概念ベースを用いて評価したと等価になる。しかし、これは、 P_B に P_A のベクトル空間の全容が知られることを意味する。すなわち、検索結果は要求者の意図により適合するものとなるが、その代償として、要求者の概念ベースが張るベクトル空間、すなわち自分の世界観を相手にさらけ出してしまふ。

6. 議 論

ピア・ツー・ピア型の情報検索が、従来の検索とは何が本質的に異なるかを考えてみる。図5は、システムが検索を行うにあたって検索対象についての先見的な知識をどれだけ持っているかにより、情報検索を分類したものである。サーバ中心型のシステムに対する従来型の情報検索では、検索対象に対する先見的な知識が検索開始前にすべてわかっている(図右側)。しがたって、検索結果の品質を高くしやすいと言える。一方、図中央は、どこからか発生し流れて来る情報ストリームの中からユーザの所望する情報のみを抽出する、情報フィルタリングである。

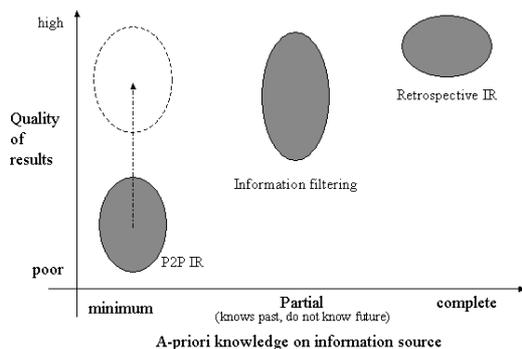


図 5 情報検索システムの分類
Fig. 5 Categories of IR Systems

過去に流れてきた情報に関する知識は持っているが、これから到来する情報について知ることはできない。このため、情報が来ごとに漸次的に情報源の特性を推定するような手法 [7] を用いて検索結果の品質を高めることになる。図の左側が、本稿で述べたピア・ツー・ピア型の情報検索であり、検索を開始するまで他のパーソナル・レポジトリに蓄積された情報に関する知識はまったく得られない。このために従来の検索手法をそのまま適用しただけでは、良好な検索の品質を得ることができない。本稿で提案したように、パーソナル・レポジトリ間で、それぞれが蓄積する情報に関する知識を交換し、それを利用して検索を行うことによってのみ、検索結果の品質を高めることができるだろう。また、情報フィルタリングと同様に、過去の検索を通じて相手のパーソナル・レポジトリに関する知識を漸次的に推定するような手法も検討の必要がある。

本システムでは、分散検索あるいはメタ検索 [8] と共通する課題も多い。しかし、これら検索手法では、個々の検索システムがかなり広範な情報を含むことを前提にしている点や、検索結果を集積する際の手法が中心であり検索システム間のインタラクションはあまり考慮されない点が、本システムとは異なっていると言える。異分野の検索システム間でのオントロジ変換に関する手法等は、本システムへの応用を検討する必要がある。

7. まとめと今後の課題

個人が持つ情報を適切に流通できることを目指したピア・ツー・ピア型の情報交換・共有システムのフレームワークについて提案した。提案した情報交換・共有システムは、情報を蓄積するパーソナル・レポジトリと、その要求や提供を制御するパーソナル・エージェントから構成され、パーソナル・エージェント同士が相互にインタラクションを行って、利用

者にとって有用な情報を交換する。このようなシステムの課題として、パーソナル・レポジトリごとにベクトル空間が異なることによりそのままでは適切な検索ができない点、情報の提供可否や形態を判断する手法が必要な点を抽出した。本提案のフレームワークでは、これらの課題を解決するため、ベクトル空間を等化しつつ検索を行う協調検索機構と、情報の内容と相手との関係とから情報を提供するか否かや提供の形態を制御する提供範囲制御機構を備えている。さらに、協調検索機構においてパーソナル・レポジトリ間でベクトル空間を等化する手法として、自動的に関連性フィードバックをかける方法と、ベクトル空間相互が線形変換の関係にあると近似して変換逆行列を推定する方法とを提案した。

筆者らは、サーバ中心型の概念ベースを用いた検索システムを開発済みである [4]。これを利用し、提案した協調検索手法と合わせて、ピア・ツー・ピア型情報共有・交換システムを実現することが今後の課題である。また、提供範囲制御機構についても、今後具体的な手法について検討して行く予定である。

文 献

- [1] A. Oram ed. Peer-to-Peer, O'Reilly, 2001.
- [2] E. Adar, D. Karger and L. Stein. Haystack: Peer-User Information Environment, Proc. 1999 Conference on Information and Knowledge Management, pp. 413-422, 1999.
- [3] G. Salton and C. Buckley. Team Weighting Approaches in Automatic Text Retrieval, K. S. Jones and P. Willet, Readings in Information Retrieval, Morgan Kaufmann Publishers, pp. 323-328, 1998.
- [4] T. Yukawa, K. Kasahara, T. Kato and T. Kita. An Expert Recommendation System using Concept-based Relevance Discernment, Proc. ICTAI2001 (to be published).
- [5] 笠原, 松澤, 石川. 国語辞書を利用した日常語の類似性判別, 情処論, 38(7):1272-1284, 1997.
- [6] H. Schütze and J. O. Pedersen. Information Retrieval Based on Word Sense, Proc. 4th Annual Symposium on Document Analysis and Information Retrieval, pp. 161-176, 1995.
- [7] J. Callan. Document filtering with inference networks, Proc. 19th ACM SIGIR, pp. 21-28, 1995.
- [8] L. Gravano, C. Chang, and H. Garcia-Molina. STARTS: Stanford proposal for Internet meta-searching, Proc. ACM SIGMOD, pp. 126-137, 1994.