

## キャンパス P2P ネットワークにおけるパーソナライズド検索

杉原 健司<sup>†</sup> 志田 匡士<sup>†</sup> 吉永 努<sup>†</sup> 曾和 将容<sup>†</sup>

<sup>†</sup> 電気通信大学 〒182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: <sup>†</sup> {sugihara,shida,yosinaga,sowa}@sowa.is.uec.ac.jp

あらまし 本稿では、キャンパスレベルを想定した P2P ネットワークにおいて、そのネットワークに参加する各ユーザが分散して提供するサービスや共有ドキュメントを有効活用するための新たな情報検索方式を提案する。具体的には、ピアの評価、関連キーワードの学習、ドキュメントのスコア化とそれに伴うスコア制御機能を備えることで、各ユーザの趣向にそった情報検索を可能にする。そして、実装したシステムを利用使用した予備実験結果を示し、提案システムの有効性について議論する。

キーワード P2P, JXTA, パーソナライズド検索, 個人エージェント

## Personalized Data Retrieval for Campus P2P Networks

Kenji SUGIHARA<sup>†</sup> Sida MASASHI<sup>†</sup> Tsutomu YOSHINAGA<sup>†</sup> and Masahiro SOWA<sup>†</sup>

<sup>†</sup> The Graduate School of Information Systems, University of Electro-Communications

1-5-1 Choufugaoka, Choufu-si, Tokyo, 182-8585 Japan

E-mail: <sup>†</sup> {sugihara,shida,yosinaga,sowa}@sowa.is.uec.ac.jp

**Abstract** This paper introduces a new information retrieval method for campus P2P networks. The purpose of our proposal is effectively sharing services and documents among users who participate in the network. The system consists of three properties: evaluation of group peers, learning related keywords for retrieval, and a function of scoring services or documents. These functions are assisted by a personal agent. We show preliminary experimental results and discuss the effectiveness of our proposing P2P network retrieval.

**Key words** P2P, JXTA, Personalized data retrieval, Personal agent

### 1. 研究背景

近年 P2P (Peer-to-Peer) と呼ばれる、クライアント・サーバ方式と比べてブローカ (仲介者) を必要としないという点で違いを持ったネットワークモデルの普及により、便利で容易に情報の交換やメッセージの送受信を行うことができるようになりつつある。キャンパスネットワークやホームネットワークといったプライベート環境でのネットワークでは、P2P モデルを利用して任意のメンバ同士でリソースを共有し、様々なサービス利用や提供などが容易且つ迅速に行うことが可能となっている。P2P モデルを利用した具体的なサービスとしては、離れている人間が同時に作業を行えるコラボレーションシステムや、メンバのスケジュール共有化、各々のメンバが分散して管理するドキュメントの共有、それに伴う情報検索技術などが代表的なものとして挙げられる[9]。最近では、家電製品をインターネットに接続し、ホームネットワーク化の実現やオフィスやキャンパスといった限定された環境下で P2P ネットワークを用い、ドキュメントを有効活用することが注目されている[2]。

しかし、こうした P2P モデルでは以前からセキュリティの脆弱製や、個人情報流出、マルチキャスト通信の増加による通信負荷の拡大といった様々な問題が指摘されている。これらの種々の問題の中で、近年、P2P ネットワークシステム利用の中でセキュリティ機能の向上[8],[10]や個人のニーズに応じた新たな情報検索技術の要求が高まっている。

上記の課題に対して、P2P モデルの長所を利用しながら、問題解決に取り組もうとする様々な試みが増えてきている。特に現在のような多様な価値観、好みといったものが存在する中で、本当にユーザが必要とする情報を正確且つ迅速に見つけ出すことが重要な問題となってきている[3],[4]。つまり、ユーザに応じたパーソナライズドな検索が最も重要な課題であるということである。

本稿では、共有プラットフォームを提案する JXTA ネットワーク[5]を利用して、P2P ネットワークを構築する。そのバーチャルネットワーク上で、認証及びアクセス制御が実装されるセキュアな環境において、グループ内で各ピアが保持するドキュメントが共有され

る。その上で共有されたドキュメントに対し、各ユーザが個別に持つエージェント（個人エージェント）により、情報検索に関するユーザの好み、特性といったものが学習される。それを実際の情報検索に反映させることにより、パーソナライズドな検索を構築するためのメカニズムを提案、実装しそれに対する評価を行った。

章構成は以下の通りである。まず2章でP2Pネットワークの特徴と、本研究でP2Pネットワークを構築するために利用したJXTAについて述べる。3章では、本研究で提案するパーソナライズド検索のアーキテクチャを説明し、4章でその実装と実験結果に関する考察を述べる。最後に5章でまとめを行う。

## 2. P2P と JXTA

この章では、P2P 技術に関する説明と本提案で P2P ネットワークを構築するために利用する JXTA ネットワークについて簡単に述べる。

### 2.1 P2P ネットワーク

P2P 技術は様々なデバイスを対等な立場で接続し、直接情報のやりとりを行うネットワーク形態として注目されている。P2P では、それぞれのデバイスがクライアントにもサーバにも、又はサーバとクライアント機能を同時に持つという選択が可能であり、それと同時に実行することが可能である。P2P モデルがクライアント/サーバモデルと異なる部分は、ピアは全てクライアントとサーバの役割を持っている点である。例えばあるユーザは他のユーザにメールを送りたいとき、メールサーバを通らずに送信し、相手は直接にそのメールを受信することができる。またユーザはあるリソースを管理したりそれを他のユーザに提供したり、逆に他のユーザからリソースをも取得することが可能である。

P2P ネットワークに関して述べる際に重要なことは、クライアント/サーバモデルと P2P モデルは必ずしも対極的な存在ではないという点である。そして、P2P モデルはピュア P2P とハイブリッド P2P のみによる分類では不十分であり、どういったシチュエーション、環境、コミュニケーションシーンにブローカ（中央サーバ）を存在させるのか、または存在させないことを決定することが重要である[1]。本研究では、最近ハイブリッド P2P を利用したシステム開発が増加する中で、ブローカの存在を廃したピュア P2P を利用することで、実装の容易性を高めると共に、様々なシチュエーションに応じた応用性を実現する。

### 2.2 JXTA アーキテクチャ

JXTA とは P2P の共通言語の必要性により、米国 Sun Microsystems 社によって開発された P2P のプロトコルであり、現在オープンソースのコミュニティーとして開発が進められている。

JXTA は様々なプログラミング言語によって実装されている。そのため、Java の開発者のみでなく多くの開発者が JXTA を利用した P2P システムの開発が可能となる。また、JXTA 自体はプロトコルとして提供されているため、現在取り組まれていないプログラミング言語での実装も今後可能となっている。既に C 言語での実装は完了しており、今後更なる言語での実装が予定されている。

今日様々な P2P アプリケーションが存在する中で、その多くが独自の仕様により実装されているため、ピアの発見・検索やデータ共有といった P2P システムにおける共通のサービスも各アプリケーションで異なってしまう。このため、他のアプリケーションとの通信は不可能となり動作環境も限定されてしまう。

一方 JXTA では、プロトコルとして提供することにより、そうしたギャップに対する補完を可能としている[7]。

この JXTA では、情報検索に関する技術はクエリの送受信に関する簡単な定義しか成されていない。そのため、情報検索に関する新たな機能を提案・実装することで、より JXTA の利便性を向上させることが可能となる。

### 3. パーソナライズド検索

本稿で提案するパーソナライズド検索は、以下の3つの要素により構成されている。

- ・ ピア評価による問い合わせ送信先の選定
- ・ 関連キーワード群の登録
- ・ ドキュメントのスコア化とそれに伴うスコア制御

なお、ここで提案する3つの処理はできる限り個人エージェントが代行することにより、ユーザの入出力負担を軽減させる。

以下、個人エージェントとユーザの関係を述べ、その後3つの要素における設計方針を示す。

#### 3.1 個人エージェント

本節では、実際の情報検索を行う一連の処理の流れを通じて、個人エージェントがどういった作業を代行しユーザの検索作業をサポートしているかについて説明する。

本研究では P2P ネットワークを利用した新たな情報

検索技術を提案している。しかし、新たな検索機能を実装するに伴いユーザの処理負担が増加するのでは、本研究の優位性が成り立たない。そのため、上記で示した3つの機能を実装した際に生じる処理を、ユーザに代わりソフトウェア上で処理することが望まれる。それと同時に、パーソナライズな検索を実現するために各ユーザの趣向等を学習する機能が必要となる。

こうした複数の問題を解決するために、本研究では個人エージェントを実装することで新たに生じるユーザ処理も含め、ソフトウェア上で処理できるものまでできる限り処理を行う。これを可能にするために、本研究では個人エージェントを提案している。

### 3.1.1 検索処理

パーソナライズ検索では、ユーザは最低限の処理を行い個人エージェントが様々な処理をサポートしている。そのため、以下に検索実行時におけるユーザインタフェースと個人エージェント、そしてその他のピアとの間で行われる処理と関係性について説明する。

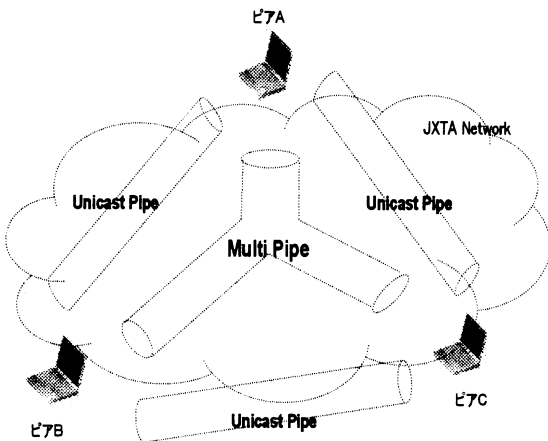


図1：ピア間のコネクション

図1は、検索処理時における各ピア間のコネクションを示している。本研究では、検索処理を行う中で複数のユーザと迅速な送受信を行うためのマルチキャスト用の Multi Pipe と、ユーザ毎に構築する1対1通信用の Unicast Pipe を持つ。こうした用途に応じた2つのパイプを持つことで、検索処理の過程で複雑に検索相手が切り替わる処理に迅速に対応することが可能であると同時に、今後新たな機能が追加される場合にも柔軟に対応することが可能である。

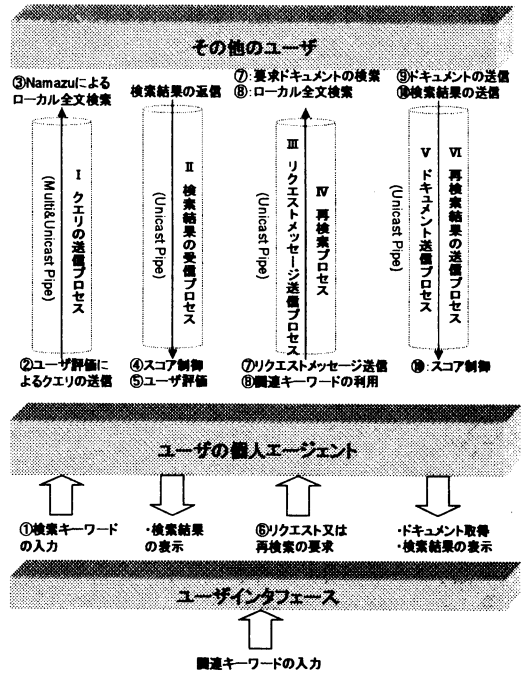


図2：検索処理の流れ順

図2では、本提案におけるアーキテクチャを示している。その中で情報検索を行う一連の処理をプロセスIからプロセスVIに分け表示すると共に、各プロセスへの移行手順やそのプロセスを通じてどのようにドキュメントを取得しているかを示す。

このアーキテクチャの特徴は、個人エージェントが検索処理の中でユーザとその他のユーザとの間の処理を代行することである。

### 3.2 ピア評価

本提案では、ドキュメントを共有するグループ内に複数のユーザが参加することを想定している。そのため、常に全てのユーザにクエリを送信することは効率的であるとは言いがたい。そのため、検索を繰り返す中で与えられる情報により、ユーザが必要とするドキュメントを持っている確率が高いピアのみにあらかじめクエリを送信することで、マルチキャストによる非効率性を排除すると共に、高い確率で欲しいドキュメントを見つけることが重要である。こうしたコンセプトを基にグループピアを評価する方法を提案する。

グループピア評価とは、ドキュメントを共有するグループ内のピアに対して、繰り返し行われる情報検索の中で返信される検索結果に応じてピアをローカルデータベースに評価・蓄積する。そして、登録された

評価データに基づきその後の情報検索を行うことで、ユーザの求める情報無作為に探すのではなく、求めるドキュメントを持つ確率の高いピアを優先的に探索することが可能となる。

ここでのピア評価を行うための評価基準は、検索結果の返信の有無によりそのピアに対して得点形式で加点される。検索結果が10個以上のものが返されれば2点、1個以上10個未満であれば1点、検索結果が無ければ0点が各ピアに加点される。検索結果数は、情報検索者の好み・特性等に対して各ピアを評価する1つの基準として信頼度が高い。そのため、上記で述べた加点方法を用いて各ピアを評価することで、ユーザの好みに合ったドキュメントを持つピアを優先的に探索することが可能となる。

### 3.3 関連キーワードの学習

ある情報を探そうとする時、誰もがその求める情報が見つかると考えられる検索キーワードを利用して検索を行う。しかし、その検索キーワードでは自分が求める情報を見つけられないことが多く存在する。そのためユーザは、新たな検索キーワードを考え検索を行う。求める検索結果にたどりつくまでこの作業を繰り返すことはユーザにとって負担の大きなものといえる。

この課題に対して関連キーワードの学習を本研究では提案する。この関連キーワードの学習とは、各ユーザが情報検索を通じて得た経験によりあるキーワードに対して複数のキーワードを関連させることが、より自分の求める情報を見つけ出しやすくなると考えられるものに対してインタフェースを通じて入力しローカルデータベース内に登録する。

以下に登録例を示す。この例では、3つのキーワードがお互いに関連するものとして登録されていることを示している。

[例：P2P+JXTA+吉永研究室]

関連キーワードの登録に際しては、ユーザが入力する方式を採用している。これは、ユーザが検索を繰り返す中で得た知識や経験に基づきキーワードを関連付けることが、最もユーザの趣向に合った検索を可能にすると共に、求めるドキュメントを見つけ出す確率を高められると判断したからである。

ここで登録された関連キーワードは、個人エージェントにより検索時に自動的に探索され、入力されたキーワードに関する関連キーワードが存在するならば、そのキーワードを基に再検索を実行する際などに利用する。

この関連キーワードの学習では、ユーザは複数の関連キーワード群を登録できる。多くの関連キーワード

群が構築されることは、よりユーザの趣向に沿った情報検索につながる。キーワードを関連付けて検索を行う方式は様々な形で存在するが、本提案ではユーザの経験という最も信頼たる情報を基に関連キーワードを構築することで、よりパーソナライズドな検索を可能にする。

### 3.4 ドキュメントのスコア化とスコア制御

一般的な情報検索では、検索キーワードが少しでもドキュメント内に含まれれば検索結果として返される。そのため、必要以上の膨大な検索結果がユーザの前に表示され、その膨大な数をユーザ自身が取捨選択するといった非効率な作業が続く。こうした問題の解決は、今後益々情報量が増加する中で重要な問題となってくる。そのため本研究では、検索キーワードに対するドキュメントの価値を導き出し、その価値に応じてユーザの前に表示するかを個人エージェントが取捨選択する“ドキュメントのスコア化とそれに伴うスコア制御”を提案する。

ドキュメントのスコア化とそれに伴うスコア制御では各ユーザのマシン上において、Namazu[6]を実装しローカルでの日本語全文検索を行う。Namazuの採用理由は、現在様々なシステム上で実装され信頼性の高いこと、フリーソフトであることである。そしてNamazuによる日本語全文検索では、与えられる検索キーワードに対してそのキーワードがドキュメント内に含まれる割合に対してスコア化することが可能である。ここで表されるスコアは、各ドキュメントの検索キーワードに対する1つの価値を表している。

次にスコア制御とは、上記で記述したスコアを利用して複数のグループピアから返される検索結果に対し、クエリを送信したユーザは個人エージェントにより一定のスコア以上を持つドキュメント以外はユーザインタフェース上に表示しない。つまり、ユーザに代わり検索結果の取捨選択処理を事前に行う。Namazuにより出力される信頼性の高いスコアを利用して検索結果を取捨選択することで、取り残しの無い検索が可能になる。

## 4. 実装と評価

前章で提示したパーソナライズド検索に関する実装及びパーソナライズドな検索を評価するための予備実験を行った。その結果を示し、検索効率とユーザの処理負担について考察する。

#### 4.1 実装

表 1: 実装環境

項目	環境
ネットワーク	プライベートネットワーク
言語	Java
OS	Windows XP, Linux RedHat9
JDK	Sun J2SE v1.4.2
JXTA	Sun JXTA v2.1
利用 PC 台数	9 台

本提案における実装環境概要は、上記で示す表 1 の内容になっている。ネットワーク環境は研究室におけるプライベートネットワークを利用している。JXTA では様々なプログラミング言語で実装可能ではあるが、本システムでは、JXTA プラットフォームを利用しているシステム開発において、最も古くから利用されている Java 言語を用いて構築している。OS は Windows XP, Linux RedHat9 に対応している。

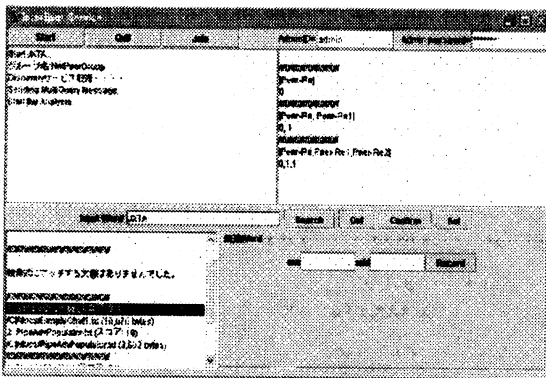


図 3: ユーザインタフェース

図 3 は、本提案におけるピアのユーザインタフェースを示す。図中左上のエリアがログを表示するためのログエリアであり、右上のエリアでは 3.2 節で述べたピア評価表が表示される。そして、中央のテキストボックスにおいて検索キーワードを入力し検索が可能となる。検索結果は、左下のテキストエリアに表示される。表示された検索結果の中からユーザが求めるドキュメントが存在すれば、ユーザはそれを指定し取得することが可能となっている。図 3 においては、ユーザは SimpleChat.txt を指定した状態を示している。ここでドキュメントが指定されることにより、個人エージェントがそのドキュメントの保有者を探索し、取得までの一連の処理をユーザに代わり行う。

#### 4.2 検索時間に関する比較

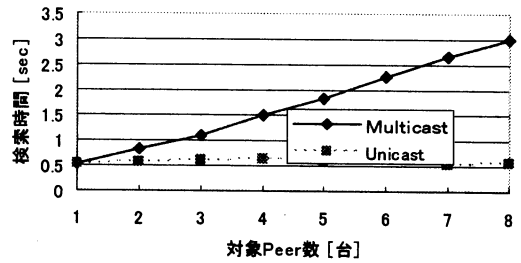


図 4: 検索時間に関する比較

図 4 は、検索対象のピアを 1 台から 8 台まで順に増やし、検索時の応答時間をマルチキャストで行った場合とユニキャストで行った場合を比較したものである。なお、ユニキャストにおける応答時間はピア評価表を用いて、優先的に探索するピアが 1 台として判断されたことを想定したものである。

一般的な P2P モデルでは、情報検索の際には主にマルチキャストによるクエリの送信を行っている。しかし、こうした無作為なマルチキャストによるクエリの送信は図 4 から分かるように、我々のプライベートネットワーク環境における 8 台のピアを検索するだけでも、その数に比例して検索時間が大きく増加していることが分かる。この原因としては、検索結果の返信時にその他のユーザと衝突した際に、返信できなかったピアは待機してから再送するオーバーヘッドが挙げられる。そのため、P2P ネットワーク規模がさらに大きくなると、マルチキャストによる情報検索は膨大な処理時間と通信負荷が生じることを予測することは難しい。

こうした結果により、ピア評価表を用いたクエリ送信先の選定により検索時間の短縮と通信負荷の減少が達成されたといえる。

#### 4.3 スコア制御による比較

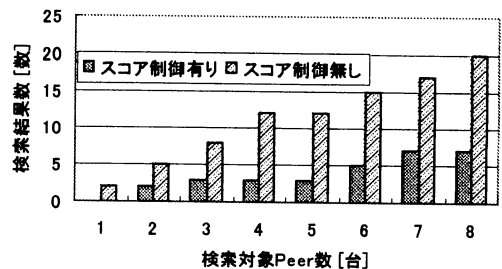


図 5: スコア制御による比較

図5は、検索対象のピアを1台から8台へと順に増やし、そこで返される返信結果数に対しスコア制御を行った場合と、そうでない場合とを比較したものである。この予備実験の前提条件として、各ピアは10個のインデックス化されたドキュメントを持ち、スコア制御を行う場合では10点以上のスコアを持つドキュメントのみ表示される条件下で比較を行っている。

予備実験により検索結果数は、スコア制御がある場合とスコア制御が無い場合とでは大きな差が生じた。これは図5からも分かるように、スコア制御が無い場合では検索対象台数がわずかに8台でさえ20以上の検索結果が返信されている。この結果から、検索対象ピア数の増加に比例して検索結果数も増大することが確認できる。このことは、ユーザが検索結果を取捨選択する作業も又検索対象ピア数に比例して増加すると考えられる。

スコア制御有りで行った場合での検索結果の数は、スコア制御無しで検索した場合での検索結果の数と比べて最大30%弱にまで抑えられている。これは単純に検索結果の数を減らしたのではなく、本当に必要な検索結果だけを残した結果である。このことから分かることは、いかにユーザにとって必要のない検索結果が多く存在し、それが検索結果として返されているかという現実である。そして、ユーザの検索結果に対する処理の軽減については、検索結果の数の大幅な減少と同時に視認性を高めていることから改善されたと言える。こうした結果から、スコア制御をより厳しく設定したり、より緩やかな設定に変更したりと柔軟に使い分けることでキャンパスネットワークのようなより広範な環境でも、各ユーザに即したパーソナライズドな検索を実現することが期待できる。

## 5. まとめと今後の課題

本稿では、キャンパスネットワークレベルでのP2Pネットワーク環境において、共有されるドキュメントをユーザに即した形で見つけ出すための新たな情報検索に関する手法を提案した。パーソナライズド検索では、ピア評価を用いることにより送信先の選定を行うことで求めるドキュメント保有者との迅速なアクセスを可能にする。そして、ドキュメントのスコア化とスコア制御によりユーザの求めるドキュメントを正確に見つけ出すための過程を短縮している。また、個人エージェントが処理を代行することでユーザの負担を軽減している。

今後は、キーワード別に登録されるピア評価表の構築とより柔軟なスコア制御機能の構築と、ホームネットワーク環境に向けたストーリーミング利用したファイルの送受信を可能にすることを予定している。さら

に、今回は予備実験における評価のみであったが、想定するキャンパスネットワークといった今回の実験環境よりも大きなネットワーク環境において実証実験を行うことにより、今回提案した方式による検索の効率性という面でより大きな効果を生み出すことの証明を予定している。

## 謝辞

本研究は一部科学研究費補助金 基盤研究(C)(2) 課題番号15500033の援助による。

## 文献

- [1] 星合隆成, “P2P総論—最新動向と将来展望,” 電子情報通信学会誌, Vol.88, No.1, pp.46-53, January 2005.
- [2] E. Turcan, and R. L. Graham, “Peering the Smart Homes,” Proceeding of the First International Conference on Peer-to-Peer Computing, August.2002.
- [3] A. D. Stefano, G. Pappalardo, C. Santoro, and E. Tramontana, “SHARK, a Multi-Agent System to Support Document Sharing and Promote Collaboration,” International Workshop on Hot Topics in Peer-to-Peer Systems, pp.86-93, October.2004.
- [4] H. Mei, and S. Chang, “PP-COSE: A P2P Community Search Scheme,” The Fourth International Conference on Computer and Information Technology, pp.416-423, September.2004.
- [5] Sun Microsystems.Project JXTA.<http://www.jxta.org>.
- [6] Namazuプロジェクト.<http://www.namazu.org/>
- [7] D. Brookshier, D. Govoni, N. Krishnan, and J. C. Soto, JXTA: Java P2P Programming, Sams Publishing, 2001.
- [8] TRAN XUANHOA, 杉原健司, 吉永努, 曾和将容 “JXTAネットワークに対するユーザ認証およびアクセス制御の導入,” 情報処理学会研究報告, Vol.2003, No.126, pp.65-70, 2003.
- [9] 杉原健司, TRAN XUANHOA, 吉永努, “プレゼンス情報によるP2Pネットワーク支援,” 第66回情報処理学会全国大会, pp.331-332, March.2004.
- [10] 志田, 杉原健司, 吉永努, “JXTAを用いたP2Pネットワークにおける認証機能委託と知的検索,” 第67回情報処理学会全国大会, March.2005 (発表予定)。