

L-048

## P2P をもちいた情報共有支援における情報管理 Information management for the information share support required with P2P

大塚 雅史†  
Masafumi Ohtsuka

上原 貴夫‡  
Takao Uehara

### 1 はじめに

近年の飛躍的なコンピュータ技術の発展と、ネットワーク環境の整備により Peer To Peer(以下 P2P) という通信形態に対する可能性に期待が高まっている。P2P は MP3 形式の音楽ファイル交換用アプリケーションソフト Napster の登場により注目を集めた。P2P は負荷の分散、設備コストの削減など多くの利点があり、様々な分野においてその可能性が注目されている。

本研究では P2P ネットワークをもちいた情報の共有化を実現する。しかし、P2P は端末が自由に情報の送受信をおこなうために、情報漏洩といったセキュリティ上の危険がある。さらに、単に情報を P2P ネットワーク上で共有するだけでは効率の良い情報共有とはならない。そこでそれらの問題点を解決し P2P ネットワーク上で効率よく情報を共有するために、本研究ではサーバによる管理機構を導入する。それにより情報の管理とユーザの管理を実現し、効率の良い情報共有とセキュリティの向上を図る。

### 2 既存システムとの比較

情報共有における問題点として、導入しても情報の共有化がうまくいかないという問題が存在している。その事例として情報を蓄積し各ユーザが情報を取得できる環境を整えたが、似たような情報や多くの情報にまぎれてしまい目的の情報になかなかたどり着かない場合がある。この場合の問題点としては、情報を蓄積した後に分類・整理していないことが上げられる。しかし、大量に蓄積した情報を人手で分類し整理することは容易ではない。そこで本研究ではサーバにメタデータを蓄積し、メタデータに付加されている要約文を利用して自動的に分類をおこなう機能を実装する。それにより、管理者による情報管理を容易にしユーザにとっても検索しやすいシステムを目指す。また、要約文があることでそれぞれの情報の内容を把握しやすくなり、ユーザによる検索がしやすくなると考えられる。

### 3 情報共有支援システム

本研究で提案する情報共有支援システムは、Hybrid 型 P2P ネットワーク上において情報を共有する。P2P では共有する情報は各ピアの記憶領域において保持される。また、提案するシステムでは Hybrid 型 P2P ネットワークを構築しサーバにより共有する情報の管理をおこなう。サーバで情報をカテゴリごとに分類し管理することでユーザの検索効率の向上を図る。

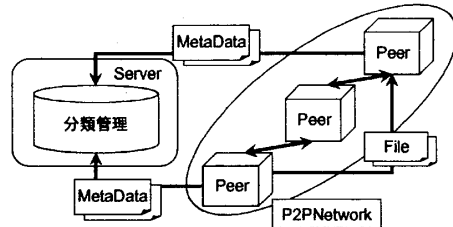


図1 システム概要概要図

各ピアが P2P ネットワーク上で共有される情報を取得する際には、必ずサーバに接続し共有している情報のリストを入手する。サーバに対して必ず接続することで、第三者による情報の取得を防ぐ。また、情報に対するアクセスが制限され共有が許可されていないユーザによる不正取得を防止する。本研究で提案する情報共有支援システムの概要図を図1に示す。

### 4 ユーザ管理

提案する情報共有システムはグループ内での情報の共有を対象としている。しかし、P2P ネットワークをもちいて情報共有をおこなう場合、第三者により P2P ネットワークに接続し情報を取得されてしまう危険性がある。そのため、本研究ではサーバによりユーザを管理することで情報漏洩の危険性を軽減する。

ユーザは管理者により事前にサーバへの登録をおこなう。登録した際、各ユーザにはユーザ ID とパスワードが割り振られる。提案する情報共有支援システムでは共有する情報に対してアクセス制御をおこなうため、ユーザをいくつかのランクに分け、ユーザ ID にランクごとの共通のコードを含ませて割り振る。サーバに対して情報を検索する際にはユーザ ID をもとにしてランクの確認をおこなう。ランクにより取得できる情報を制御することでグループ内からの情報漏洩を防ぐ。

#### 4.1 アクセス制御

本研究では共有する情報に対してアクセス制御をおこなう。共有する情報からはメタデータを生成する。そのメタデータにアクセス制御コードを属性として付加し、情報へのアクセスを制御する。アクセス制御コードは情報の作成者が情報を共有するためにメタデータを生成する際、任意に設定し付加する。

ユーザが情報を取得するためサーバにアクセスした際には、メタデータに付加されているアクセス制御コードとユーザ ID を比較しアクセスが可能なユーザである場合には、ユーザに取得可能な情報のリストを提示する。ま

† 東京工科大学大学院工学研究科

‡ 東京工科大学コンピュータサイエンス学部

た、ユーザがアクセス不可能な情報を P2P ネットワーク上から取得を試みた場合には、情報を配信しているピアが保持しているメタデータと比較して、一致しなければ情報を送信しない。

## 5 情報管理

P2P ネットワーク上で情報を共有する場合に、情報は各ピアの記憶領域において保持されるため、共有されている情報について一元管理することは難しい。しかし、情報共有という観点から考えると P2P ネットワーク上でどのような情報が共有されているのかを管理者が管理できる機能が必要である。本研究ではその機能を実現するために共有する情報から要約文を含むメタデータを生成し、サーバにおいて利用し情報を管理する。

情報をサーバにおいて管理することで、管理者により違法性の高いファイルや共有するには有益じゃない情報を排除することができる。本研究で提案する情報の蓄積・管理部の概要図を図 2 に示す。

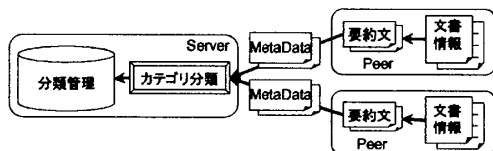


図 2 共有情報の蓄積・管理部の概要図

### 5.1 メタデータの利用

本研究では、メタデータを利用することで P2P ネットワーク上で共有する情報を管理する。メタデータにはファイル名、情報の作成者、更新日時、要約文、要約文生成日時、発信者 ID、アクセス制御コードなどを属性として付加する。

メタデータはサーバにおいて情報を管理するためだけでなく、ユーザがサーバにおいて取得できる情報のリストを取得する際には、ファイル名だけでなく要約文も提示し、ユーザが取得する情報を選択するための材料としても活用する。

### 5.2 要約文生成手法の検討

本研究では共有する情報の対象を文書情報としているため、サーバでの情報管理をおこなうのに、共有する情報から要約文を生成しもちいる。

要約文はユーザがネットワークに対して情報の共有を許可した際に生成される。生成した要約文はメタデータに付加されサーバに登録される。

要約文は以下の過程を経ることにより生成される。

1. 文書の内容を解析する
2. 解析結果から重要箇所を抽出する
3. 重要箇所を連結し要約文を生成する

本研究では各文書を単語単位に分割するために Namazu, KAKASI というツールをもちいる。単語分割に

は KAKASI をもちいる。KAKASI は動作が軽く、単語の追加をおこなう場合にも容易におこなうことができる。

重要箇所の抽出は主に単語の重みを指標としてもちいる。本研究では単語に対する重み付けには TF-IDF 法をもちいる。TF-IDF 法は単語の重み付けをおこなう場合には一般的にもちいられる方法である。しかし、出現頻度をもちいて計算しているので文書の長さによって精度に差が出る。そのため、出現頻度を正規化し文書長の長さによる影響を少なくする必要がある。また、TF-IDF 法は高い精度で計算するためには大量の文書が必要になる。これらの問題点をふまえると TF-IDF 法の結果のみを利用して要約文を生成するには、精度の点において問題がある。よって本研究では TF-IDF 法のほかに情報利得比 [1] の考え方ももちいる。情報利得比による単語の重み付けの手法は、複数の検索文書の間に存在する類似性構造を階層的クラスタリングにより抽出し、その構造に即している語に高い重みをつける。これら二つの手法により要約文生成をおこない精度を比較し、計算速度なども加味し要約文生成手法を決定する。

### 5.3 情報の分類

本研究ではサーバにおいて情報をカテゴリごとに分類し情報を管理する。情報を分類するために各ピアにおいて生成されたメタデータに含まれる要約文を利用する。

カテゴリごとに情報を分類することで、ユーザの検索効率の向上が見込める。また、情報がカテゴリごとに分類されているため、ユーザが求める情報だけでなくカテゴリ内の他の情報も同時に閲覧することができる。これは求める情報だけでなく、カテゴリ内の情報も付加情報として利用することができる。これにより共有する情報が過去に利用され情報提供者にとっては価値の薄れた情報であったとしても、他者に再利用されることで情報としての価値が高まる可能性が生まれる。

## 6 おわりに

本稿では P2P ネットワーク上での情報共有において、サーバでメタデータを利用し要約文にもとづいた情報分類する方法について述べた。また、ユーザをランクにわけ情報へのアクセスを制御する方法についても述べた。発表時には要約文生成部分を実装し述べる予定である。

本研究では Hybrid 型 P2P ネットワークをもちいた情報共有を提案している。情報共有をおこなう場合に情報管理は重要である。情報管理は共有されている情報を分類・整理し、ユーザが利用しやすくすることが目的である。そのために本研究ではサーバをもちいて情報の整理・分類をおこなっているが、ユーザビリティという点において課題が残る。今後は課題としてユーザビリティを考慮し、ユーザの利用しやすいインターフェースも考える必要がある。

## 参考文献

- [1] 佐々木拓郎, 野澤正憲, 森辰則著「情報利得比に基づく語の重要度と MMR の統合による複数文書要約」自然言語処理研究会報告, 情報処理学会 2002