

P2P をもちいた情報共有ネットワークの開発

大塚 雅史[†] 金井 尚樹[‡] 佐藤 敏紀[†] 上原 貴夫^{††}
 東京工科大学大学院工学研究科[†] 東京工科大学工学部[‡] 東京工科大学コンピュータサイエンス学部^{††}

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 は負荷の分散, 設備コストの削減など多くの利点がある。P2P は様々な分野において, その可能性が見出されつつある。そのひとつにナレッジマネジメント (Knowledge Management, 以下 KM) がある [1][2]。

KM は 1980 年代にアメリカにおいて考案された経営手法で, 人間のもつ知識や経験を企業のもつ知的財産として集約することによる, 企業の競争力強化が目的である。KM は個人の知識を組織の知的財産とするため, KM の実現には情報の共有化が重要になるが容易ではない。また個人は職務活動において個人単位で大量の文書を作成または取得し保持している。しかし, その文書を個人が管理するだけでは企業の財産または力にならない [3]。

以上をふまえ, 我々は P2P ネットワークによる情報の共有化を実現し業務の効率化を目指している。情報を P2P ネットワーク上で共有するだけでは業務を効率化させるための情報共有とは言えない。そのため, 情報を取得しようとするユーザが効率的に取得できるように情報を管理する必要がある。

本稿ではサーバにより情報管理をおこなう, ハイブリット型 P2P ネットワークをもちいた情報共有支援システムを提案する。

2. ナレッジマネジメントにおける情報共有

KM において場における情報共有は重要な役割を果たす。場は人間が集まり知識を創造する場所であり, 新しい知識は異質な知識の建設的対立コラボレーションから生まれる。よって KM では自分のもつ情報だけではなく, 他人の持つ情報も活用することで新たな発見をもたらす作業効率向上といえる。しかし, 場は物理的なオフィス空間やブログまで様々な形式をとっている。その為, 情報共有を可能にする環境の整備が求められる。

3. 情報共有支援システム

システム全体の概念図を図 1 に示す。

本研究で開発する情報共有支援システムは, ハイブリット型 P2P ネットワークにおいて情報を共有する。そのため共有する情報は各ピアにおいて管理され, 各ピアの記憶領域において保持される。サーバにより共有する情報の所在や内容を管理することで, 情報をカテゴリごとに分類しユーザの検索効率の向上を図る。

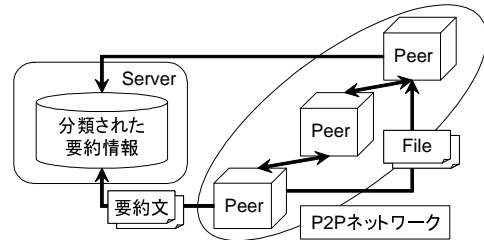


図 1: システムの概念図

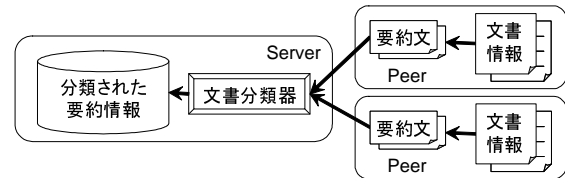


図 2: 要約情報の蓄積部概念図

各ピアはサーバとクライアント両方の機能を有するサーバントとして機能する [4]。つまり, 各ピアは検索・発見機能を有し, 常に 1 対 1 の接続をおこなう。P2P は大きくハイブリット型とピア型に分類できる。本研究ではハイブリット型の P2P ネットワークをもちいてシステムを構築する。ハイブリット型 P2P は純粋な P2P ネットワークとは違い, ネットワーク上にサーバを設置する。本研究ではサーバをインデックスサーバとして利用すると共に, 情報をサーバにおいて管理する。

3.1 情報の管理

本研究では共有する情報は各ピアにおいて保持し, ピアは情報源である文書情報から要約文を生成する (図 2)。サーバはピアが生成した要約文をもちいてカテゴリごとに文書分類をおこなうことで, ユーザの求める情報とそれに類似する情報を提示できる。これにより, 情報提供者にとっては価値の薄れた情報であっても, 他者により再利用され価値が高まる。

3.2 情報の検索

ユーザがピアから情報を検索する場合, サーバが管理する分類された要約情報を基にして検索するため, ユーザは求める情報だけでなく類似した情報についても検索できる。ユーザは自分が取得したい情報を P2P ネットワークに対して検索をおこなう。その結果ネットワーク内において情報が発見された場合にはピア同士が接続し情報を取得する。

Development of the information share network which use Peer To Peer

[†]Masafumi Ootsuka, Toshinori Satoh, Graduate School of Engineering, Tokyo University of Technology

[‡]Naoki Kanai, School of Engineering, Tokyo University of Technology

^{††}Takao Uehara, School of Computer Science, Tokyo University of Technology

city of Technology

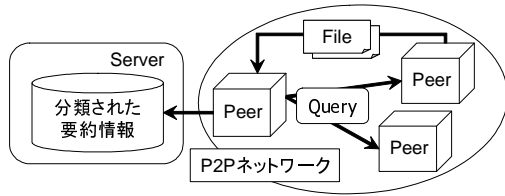


図 3: 情報検索部の概念図

3.3 要約文の生成

ピアは共有する情報である文書情報から要約文を生成する。要約文はサーバにおいて共有する情報の管理にもちいられる。要約文は以下の過程を経て生成される。

1. 文書を解析する
2. 解析結果から文書中の重要箇所を抽出する
3. 重要箇所を含む文章を連結し要約文を生成する

重要箇所の抽出は単語の重みと共起度を基にする。単語の重み付けはTF-IDF法などをもちておこなう。また、共起度は単語の位置情報から単語間の距離を算出し共起関係を決定することにより求められる。文書を単語単位に切り分けた後に、未知語が拾えるように適正な処理をおこなう。要約文はサーバにおいてカテゴリ分類において利用される。

3.4 情報のカテゴリ分類

サーバは要約文を基にしてピアが保持する情報に関するメタデータを生成する。メタデータには、情報の作成者、情報の更新日時、要約文の作成日時、要約文などを記述する。生成したメタデータをもとにサーバで情報のカテゴリ分類をおこない共有する情報の管理を実現する。

カテゴリ分類ではクラスタリングをもちいる。クラスタリングはデータ解析手法の1つで、データの集合を共通の特徴をもつクラスタに切り分ける手法である。この特徴はおもに類似性や距離尺度に基づき求められる。

ユーザが情報検索する際に情報がカテゴリに分類されているため、目的の情報に類似した情報を付加情報として同時に取得できる。付加情報をメタデータに記述することにより、情報検索の効率を向上させることができる。

3.5 キャッシュ制御

我々はP2Pネットワーク上の各ピアに、他のピアに公開する情報のキャッシュを生成させ、それを他のピアに対して配置させることで検索効率の安定を図る。キャッシュの制御方法は田中氏らの研究により提案されている[5]。キャッシュ制御方法として、情報を取得したピアと発信源であるピアを繋ぐ経路上にあるピアにキャッシュを配布する方法(図4,方法1)、ピアに対してランダムにキャッシュを配布する方法(図4,方法2)などがある。

キャッシュ制御では効率性と信頼性のバランスを取る必要がある。そのためキャッシュの生成数維持と配布先の選定が課題となる。

そこで本研究では規模の異なる環境でのキャッシュ制御の有効的な方式を検討する。そのために本研究では複

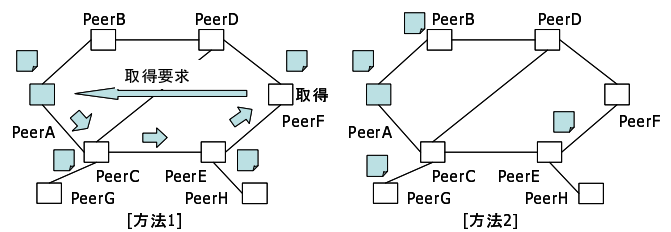


図 4: キャッシュ制御方法

数のモデルケースを作成し、さまざまなキャッシュ制御方法を実験する。実験により導かれた傾向から、各モデルケースにおいての最適な制御方法を検討し実装する。

4. おわりに

本稿では提案する情報共有支援システムについての全体像を述べ、今後実装すべき機能について述べた。

本研究ではハイブリッド型P2Pネットワークをもちいて情報共有支援システムを構築する。サーバをもちいることでネットワーク内において共有する情報を管理できる。企業内の情報共有システムについて考えた場合には、サーバにより情報を管理する方が管理者が情報を把握できるため都合が良い。なぜなら、管理者に有益ではない情報、法に抵触するような情報などを排除できるからである。

今後の課題としてはピアに対するセキュリティ機能の実装がある。P2Pネットワークでは一度ネットワーク上に流れてしまった情報はシステムの管理者でも、完全にコントロールすることが出来ない。情報漏えいを引き起こす原因のほとんどは、人為的なミスである。そこで利便性を下げずに、ピアがローカルに保持している情報の機密性を高める必要がある。安全で低コストに機密を保持する方法としてローカルに保持する情報を分割して保存する、秘密分散法[6]の適用などを検討している。

参考文献

- [1] 野中郁次郎, 梅本勝博著「知識管理から知識経営へ-ナレッジマネジメントの最新動向-」人工知能学会誌,16巻,1号,pp.4-14,2001
- [2] 國藤進, 川口高平著「ナレッジマネジメントとIT」人工知能学会誌,16巻,1号,pp.42-47,2001
- [3] 国領二郎, 野中郁次郎, 片岡雅憲著「ネットワーク社会の知識経営」NTT出版,2003
- [4] Brendon J. Wilson 著「JXTAのすべて P2P Javaプログラミング」日経BP社,2003
- [5] 田中剛, 盛田保文, 上野和彦, 愛川和宏「企業向けP2P情報共有システムにおけるキャッシュ制御方式の提案」DICO2004,2004
- [6] 鹿島隆行, 宇田隆哉, 伊藤雅仁, 市村哲, 多胡和也, 星徹, 松下温「PC共有による安全で低コストなP2P分散ファイルシステムの提案」DICO2004,2004