

分散コンテンツ統合のための P2P 型コンテンツ管理システムの構築

稲垣 浩^{†1} 芝 公仁^{†1} 岡田 至弘^{†1}

近年、様々な種類のデータを扱う多様なコンテンツ・サーバが運用されるようになってきている。本稿では、複数のサーバに分散しているコンテンツを統合することで、コンテンツを効率良く管理することができる P2P 型コンテンツ管理システムの構成について述べる。本システムは、複数のサーバ上のコンテンツを操作し、P2P 上で扱うコンテンツ情報に変換する機構をもつ。本機構により、ネットワーク上に分散しているコンテンツの一括した検索や取得が可能になる。本システムを用いることで、多数のサーバのコンテンツを統合して管理できるため、コンテンツの検索や取得の効率化が実現される。

Development of a P2P-based Content Management System to Integrate Distributed Contents

HIROSHI INAGAKI,^{†1} MASAHITO SHIBA^{†1}
and YOSHIHIRO OKADA^{†1}

Many contents servers have been used to keep various types of data in late years. This paper describes an implementation of a P2P-based content management system that integrates distributed contents in two or more content servers. The proposed system has a mechanism of controlling the servers and creating content control data for managing them. The system manages the servers and contents with these data over P2P network. With this management method, the system provides efficient use of contents by integrating servers and contents.

^{†1} 龍谷大学
Ryukoku University

1. はじめに

コンテンツを取り扱うシステムには、データベースサーバ、ウェブサーバなど方式の異なる様々なものがある。このような各サーバに分散したコンテンツを統合することは、コンテンツを管理する上で有用である。これらの各サーバはそれぞれ独自のインターフェースやデータ管理方式を採用しており、コンテンツを統合するためにはこのようなサーバ間の差異を解消する必要がある。本稿では、異なる複数のサーバに分散したコンテンツを統合するコンテンツ管理システム (CMS: Content Management System) の構成について述べる。

現在の CMS はサーバを中心とした集中管理型が主流であるが、拡張性や管理コストの問題を解決するため、分散コンピューティングを取り入れた CMS についての研究が近年盛んに行われている。このような分散コンピューティングの手法のひとつである、Peer-to-Peer (P2P) ネットワークは、各計算機がネットワークに対し参加・離脱することを考慮し、各計算機をサーバ兼クライアントとする分散型アーキテクチャである。ノード数が可変のネットワークにおいて拡張性やロバスト性をもつ P2P は、大規模システムの新規構築だけでなく、処理能力の低い計算機の活用方法としても期待されている。

本稿では、P2P のオーバーレイネットワークを提供する通信プロトコルである JXTA¹⁾ を用いて、コンテンツのタイトル、説明、データ元のアドレスなどで構成されたコンテンツ情報を各ピアに分割して保持する P2P 型 CMS を提案する。提案システムでは、各既存サーバのコンテンツを一括して P2P 上のデータベースで管理する。この特徴により提案システムは、各計算機の参加や離脱を柔軟に行え、また複数のサーバに渡るコンテンツの検索や取得を効率的に行うことが可能である。

提案システムは CMS サーバ、CMS クライアント、CMS ドライバの 3 つの機構から構成される。CMS サーバは、P2P ネットワークを構成するピアであり、コンテンツ情報の登録、更新、保持などの制御を行う。CMS クライアントは CMS サーバ上のコンテンツ情報を操作する機構であり、CMS サーバに対して登録、検索、更新、削除などのコンテンツ情報の操作を要求し、結果を受け取る。CMS ドライバは、データベースなどの各サーバと連携しコンテンツ情報を CMS サーバへ登録するほか、CMS クライアントからの要求に応じてコンテンツの転送を行う。

以下、本稿では、2 章で構築する CMS のシステム要件と JXTA の各概念を述べ、CMS への JXTA の適用手法を述べる。3 章で提案システムの構成と処理手順を述べ、4 章で提案システムの性能評価を示す。5 章で関連研究と提案システムとを比較し、最後に 6 章でまと

めを行う。

2. コンテンツ管理システムへの JXTA の適用

本章では、提案する P2P 型 CMS のシステム要件を示し、それを実現するための JXTA の各概念を述べ、それらを CMS に適用することによる利点を述べる。

2.1 システム要件

提案システムは、データベースサーバ、ウェブサーバなどの外部サーバ上の各コンテンツを統合することを目的としている。このような環境を想定したとき、サーバの保守などのため、各計算機の CMS への参加や離脱を柔軟に行うために、ある計算機が参加・離脱したことを他の計算機が確実に知る必要がある。さらに保守などで外部サーバが CMS へのコンテンツ公開を一時停止することを想定する場合、停止時に他のクライアントによるコンテンツ参照に影響が出ないよう対策が必要である。これらの問題を解決するため、提案システムにおいては、P2P のオーバーレイネットワークを提供する通信プロトコルである JXTA を採用することで、下位のネットワーク構造を意識せずに自由な通信を行えるようにしている。

また想定するネットワーク環境によって CMS の設計は異なり、たとえば、インターネット上で公開されたコンテンツに自由にアクセスできる環境を想定したコミュニケーション・システム²⁾や、施設内などの特定ネットワーク内において検索単語をもとにユーザの嗜好に合わせた検索結果を提示する手法³⁾などが提案されている。あらゆるネットワーク環境を想定した汎用的な CMS を実装する場合、処理速度、可用性、拡張性などいくつかの要件を満たす必要があり、これを解決するため、本システムにおいては、P2P のオーバーレイネットワークを提供する通信プロトコルである JXTA を用いることで、下位のネットワーク構造を意識せずに自由な通信を行えるようにしている。

2.2 JXTA の各概念と適用手法

JXTA は P2P アプリケーションを容易に開発するための通信プロトコルである。本プロトコルは、Java や C など様々なプログラミング言語での実装が可能であり、通信に互換性をもった P2P システムを開発できる。また JXTA のもつ各概念は本システムへの適用にあたって有用だと考えられる。前節のシステム要件を満たすため、本システムでは、主に以下の 3 つの概念を用いる。

JXTA ではピアやコンテンツなどあらゆる資源を、XML で記述された advertisement というメタ情報で管理している。この advertisement には、データ名、データ概要、データ元のアドレス、各ピア・各コンテンツに割り振られた JXTA 固有の ID などが記述されてい

る。各ピアは、advertisement のデータベースを分割保持しており、ID を元以後で述べる分散ハッシュ表 (DHT: Distributed Hash Table) を用いた手法⁴⁾によって、P2P ネットワーク上の資源を素早く検索することができる。

CMS において JXTA を適用する場合、コンテンツ情報を advertisement に格納し各ピアに保持させることで、JXTA による P2P ネットワーク上においてコンテンツ情報の登録・検索が可能となる。なお advertisement は文字列、数値、バイト列などで任意の構成ができるため、本システムで利用するコンテンツに応じて、コンテンツの管理方法を変更することも可能である。たとえば画像データベースと連携する場合、ファイルサイズの小さいサムネイル画像を生成し、登録するコンテンツ情報に付加することで、システム利用者は検索の段階でサムネイル画像を取得でき、実画像を取得するかどうかの判断に役立てることができる。また動画画像などファイルサイズの大きなコンテンツを扱う場合、パケット分割したコンテンツ自体を登録することで、通信負荷を抑えた分散ダウンロードが可能である。このように advertisement による資源管理は、コンテンツに応じて提供するサービスを変えることができ、分散コンテンツを統合する上で有用である。

また JXTA では、前述した advertisement の検索に、DHT と近距離ノードの探索を組み合わせた手法を用いている。DHT は、ノードのアドレスやコンテンツ情報を一意な識別子空間 (ハッシュ空間) にハッシュ値として写像し、その空間を各ノードが分散管理することで、特定ノードに負荷が集中しない高速な検索を行う手法である。DHT は、完全一致検索においては素早く、低負荷な検索が可能だが、一方、部分一致検索のような曖昧な検索は高速に行うことが難しい。この様な DHT の欠点に対し、JXTA では、DHT で目的 advertisement が見つからなかった場合、近距離にあるノードを線形探索する。これによって、部分一致検索においても検索時間を短縮し、より柔軟な検索を可能としている。

本システムにおけるコンテンツ情報の管理に advertisement を用いた場合、コンテンツ情報の探索において上記の手法が適用される結果、DHT の特徴であるノード数に依存しない高速・低負荷な検索が可能となり、また部分一致検索も実用的な速度で実現できる。また Xin らの研究⁵⁾では、JXTA のデータ探索機構の上に正規表現を処理する機構を実装することで、AND・OR 検索や日付・サイズを指定した検索などを実現している。以上のことから、JXTA を適用した本システムにおいて、高速・低負荷でかつ多機能なコンテンツ検索を提供することが可能だといえる。

また、JXTA は下位の通信プロトコルとして TCP や IP マルチキャストなどを実装しており、P2P ネットワーク上の計算機は IP アドレスなどに依存しない固有のピア ID で識別

される。このため利用者は通信相手の IP アドレス、ポート番号などのネットワーク上の位置情報を意識することなく通信が可能である。また各ピアは P2P ネットワークへの参加時や離脱時に他のピアへ ID を告知するため、各ピアが互いの状態を認識し、P2P ネットワークに支障が起きないようにしている。このため、ピア ID による計算機の発見機能を本システムに用いることで、各計算機の参加や離脱を柔軟に行える CMS を構築することが可能である。

以上のように JXTA の各概念は、計算機の参加・離脱に柔軟に対処するという要件に適用でき、これらを適用した CMS は下記に示すような利点を得ることが可能である。

- 異なるシステムのコンテンツを統合し一元管理が可能
- CMS のサーバは素早い完全一致、部分一致検索が可能
- CMS のサーバはネットワーク上の位置を意識することなく互いに通信が可能
- 計算機は互いの参加や離脱に対応し、CMS への悪影響を軽減できる

またシステム開発者は、JXTA の API の利用方法さえ守れば、コンテンツやメタ情報の管理方法や計算機への役割分担などを自由に決めことができ、運用環境に適した CMS を構築しやすくなるという利点もある。このように、JXTA を適用した本システムは分散コンテンツの統合を行う上で有用であるといえる。

3. コンテンツ管理システムの設計

本章では、前章で述べた JXTA の適用手法をもとに設計した P2P 型 CMS のシステム構成、コンテンツ登録・検索・転送の処理手順を述べる。

3.1 システム構成

図 1、図 2 が提案システムの主構成、および内部構成である。提案システムは CMS サーバ、CMS クライアント、CMS ドライバから構成される。

CMS サーバは、他の CMS サーバ群と JXTA のプロトコルにより P2P ネットワークを構築し、CMS クライアントや CMS ドライバからの命令に従い、公開、更新、検索などのコンテンツ制御を行う機構である。コンテンツ制御部は CMS クライアントや CMS ドライバから送られてきたコンテンツ登録や検索などの命令に基づき、コンテンツ情報を advertisement に変換したり、advertisement の公開、探索を行う機構である。advertisement 生成部は、登録命令の際に CMS クライアントや CMS ドライバから提供されたコンテンツ情報を元に、advertisement を生成する機構である。また、advertisement データベースは、生成されたコンテンツの advertisement を管理している。コンテンツ制御部は advertisement データ

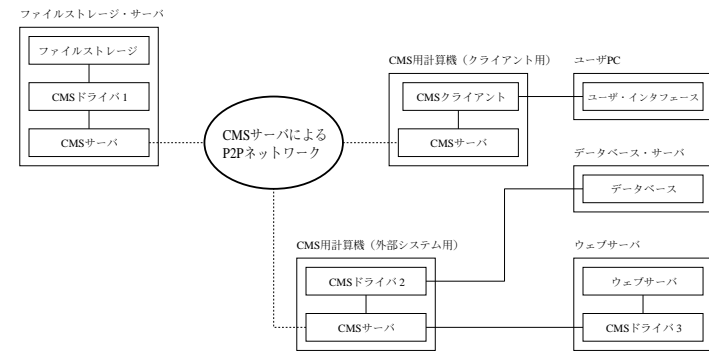


図 1 提案システムの主構成

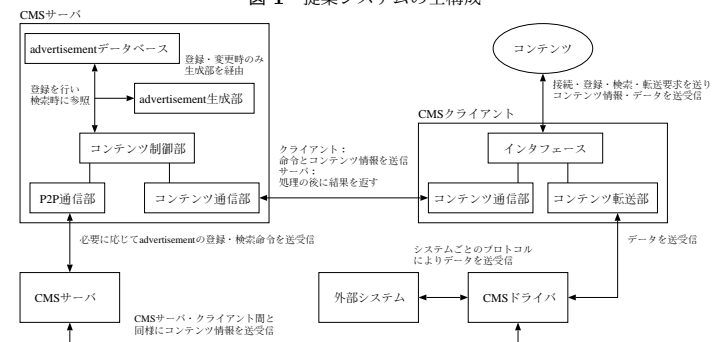


図 2 提案システムの内部構成

表 1 CMS クライアント・CMS ドライバがもつ命令

命令名	機能
put	ローカルのコンテンツ情報を CMS へ登録
search	指定した条件と一致したコンテンツ情報の取得
get	search の結果をもとにコンテンツを取得
update	CMS へ登録済みのコンテンツ情報の更新
remove	CMS へ登録済みのコンテンツ情報の削除

ベースを参照し、advertisementの登録・探索・削除を行う。P2P通信部は、CMSサーバ同士の通信を行う部分であり、IPアドレスが未知のCMSサーバであっても発見し、通信することが可能である。一方、コンテンツ通信部は、CMSクライアント・CMSドライバとの間でコンテンツ情報や処理命令を送受信する機構である。本機構は、P2P通信部とは異なり、JXTAのネットワークとは独立した通信を行っている。CMSサーバ同士はIPアドレス、ポート番号などのネットワーク上の位置情報を事前に告知せずとも、JXTAの機構により互いを発見することが可能である。また各CMSサーバはP2Pネットワークから離脱時に、その旨を他サーバに告知することにより、コンテンツ管理に支障が起きないようにしている。

CMSクライアントは、ユーザや外部のシステムなどがCMSを利用するための機構である。この機構はP2Pネットワークに直接参加せず、CMSサーバと通信を行い、コンテンツ登録や検索などの命令を送信することで間接的に外部のコンテンツサーバを利用することが可能である。また、内部の機構であるコンテンツ通信部はサーバと通信を行い、登録・検索・削除など表1のような命令を送信し結果を受信する。また、コンテンツ転送部では検索で得られたコンテンツ情報を元に、他のクライアントやドライバとコンテンツの送受信を行う。インタフェースは、ユーザや外部サーバからCMSの操作命令を受け取り、コンテンツ通信部・転送部へ命令を送り、結果を受け取る。

CMSドライバは、データベースサーバやウェブサーバなどからコンテンツ情報を取得しCMSサーバへ送信するための機構である。この機構は、CMSクライアントと同様に登録、検索、削除などの命令機能（表1参照）をもつほか、外部サーバのコンテンツを読み出すための機能をもつ。たとえば、ウェブサーバであればリンクを抽出・解析し適切なメタ情報を設定する機能をもち、またデータベースであればコンテンツを参照するためのSQLのようなデータベース操作機能をもつ。これにより読み出されたタイトル、説明文、アドレスなどのメタ情報は、advertisementに格納できるよう値を整理しCMSサーバへ登録することで、サーバ間のコンテンツ情報に汎用性を持たせることができる。このような機能を用いることで、CMSサーバはCMSドライバを介して外部サーバに接続し、コンテンツの登録、検索などを行うことが可能となる。また、サーバごとのコンテンツ情報をadvertisement向けに正規化させる機能を実装することで、コンテンツを提供する各サーバのコンテンツ管理手法が異なる場合にもコンテンツの登録や検索に一貫性をもたせることを可能にしている。このため、検索条件に一致するコンテンツ群が複数のサーバに存在している場合でも、本システムではコンテンツがCMSサーバで一様に扱われ、通信負荷や計算負荷をかけない並列

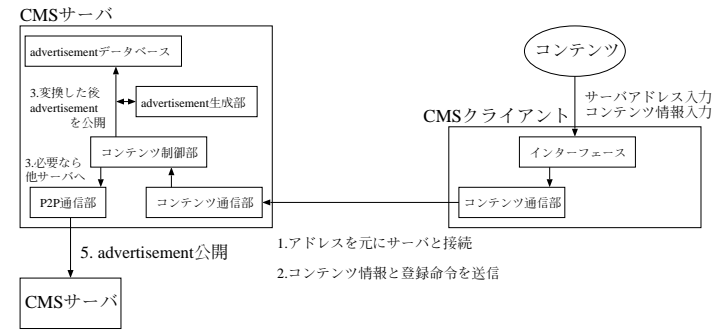


図3 コンテンツ登録の手順

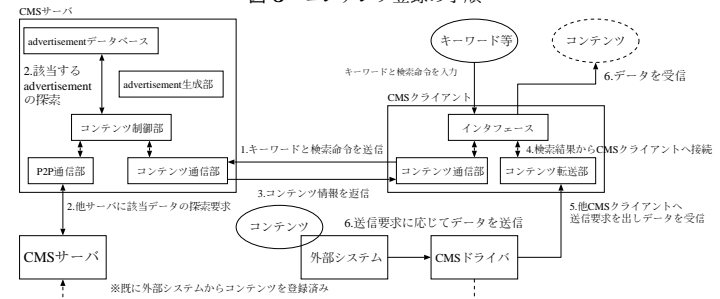


図4 コンテンツ検索・転送の手順

での検索を可能としている。このため本システムを介したコンテンツ検索は、それぞれの外部サーバで個別に検索や取得を行う手法に比べて、処理時間の短縮が期待できる。

3.2 システムの処理手順

本システムにおいてCMSクライアントとCMSドライバで実装される命令であるput（コンテンツ登録）、search（コンテンツ検索）、get（コンテンツ転送）の命令を行ったときの内部機構の処理手順をそれぞれ述べる。

本システムでは、P2Pネットワークを構築するCMSサーバが稼働している必要がある。またCMSサーバの稼働後は、CMSクライアントやCMSドライバからの接続を受けるために、自身のIPアドレスなどをCMSクライアントやCMSドライバへ告知する必要がある。CMSクライアントとCMSドライバの参加手順は似ており、別途告知されているアドレスを元にCMSサーバに接続するだけでよく、この通信はJXTAに依存しないものであ

る。また外部サーバが CMS へ参加する時は、既に CMS へ参加済みの CMS ドライバから外部サーバに接続してもらう必要がある。

CMS クライアントや CMS ドライバが put する時の処理手順を以下に示す。(図 3 参照)

- (1) CMS クライアント・CMS ドライバが、ユーザや外部サーバから入力された CMS サーバのアドレスをもとに、CMS サーバに参加要求を送信し、参加要求を受けた CMS サーバが要求元の CMS クライアント・CMS ドライバと接続を行う。
- (2) CMS クライアント・CMS ドライバが、put 命令によりコンテンツ情報を CMS サーバへ送信する。
- (3) CMS サーバはコンテンツ情報を advertisement に変換し、自分や他の CMS サーバのデータベースに登録する。

CMS クライアントは、公開リストなどにより、接続に必要な CMS サーバのアドレス入手する。またコンテンツ情報は、タイトル、説明文、関連タグ、登録期間、コンテンツを所持する CMS クライアントのアドレスなど、システムであらかじめ定義された形式に従い作成する。CMS サーバは、送られてきたコンテンツ情報を XML 形式である advertisement に変換し、advertisement データベースへ登録することでコンテンツの登録が完了する。なお advertisement の登録場所は DHT により各 advertisement に付与された ID に基づいており、場合によっては他の CMS サーバのデータベースに送られることもある。

次に、CMS クライアントや CMS ドライバが search や get をする時の処理手順を以下に示す。(図 4 参照)

- (1) CMS クライアントは search 命令により検索要求を CMS サーバへ送信する。
- (2) CMS サーバがデータベースから該当する advertisement を探索する。
- (3) 該当する advertisement があれば、格納されているコンテンツ情報を CMS クライアント・CMS ドライバへ返信する。
- (4) 検索結果を受信した CMS クライアントは、目的コンテンツを所持する CMS ドライバのアドレスを検索結果から読み込み、そのアドレスに接続する。
- (5) CMS クライアントが、get 命令によりコンテンツの転送要求を CMS ドライバへ送る。
- (6) 転送要求を受けた CMS ドライバが、要求元の CMS クライアントへコンテンツを送信する。

search において、CMS クライアントは、タイトルの一部やタグなどの検索項目を指定して検索要求を CMS サーバへ送信する。CMS サーバでは送られてきたキーワードに該当する advertisement を探索し、該当する advertisement をコンテンツ情報に逆変換した後、CMS

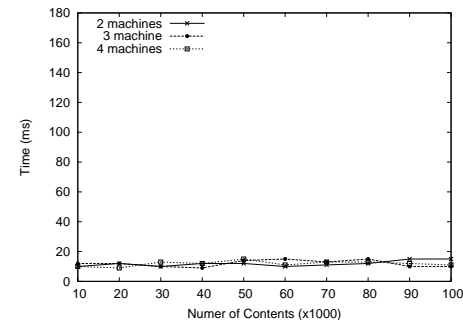


図 5 完全一致での検索時間

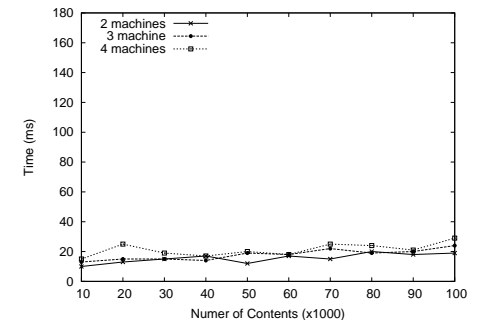


図 6 前方一致での検索時間

クライアントへ返信する。これにより CMS クライアントは検索結果を得ることができる。また get の処理は CMS サーバを介さず、CMS クライアント・ドライバ間で直接行う。まず得られたコンテンツ情報に記述された IP アドレスを元に、コンテンツを所持する CMS ドライバに接続し、コンテンツの転送要求を送信する。そして他の CMS ドライバが、該当するコンテンツを命令元へ送信することで、コンテンツの転送が完了する。

4. 評価

提案システムによるコンテンツの管理効率の改善を確認するため、以下の評価実験を行った。

- コンテンツ数と計算機数が検索時間へ与える影響
- 外部サーバと連携時のコンテンツ取得時間
- 実運用サーバ統合時のコンテンツ取得時間

なお評価環境として、Pentium4 2.4GHz プロセッサとメモリ 1024MB を搭載した計算機を複数台用意した。これらを 1Gbps のイーサネットに接続した同一セグメント上で実験を行った。計算機の OS は Linux 2.6.27 であり、本システムの実行系として Java SE 6 を用い、JXTA は Java 2 SE 用のバージョン 2.5 を用いた。

4.1 コンテンツ数と計算機数が検索時間へ与える影響

まず、本システムに参加する計算機の数や登録しているコンテンツ数が増加したときの検索時間への影響を評価する。本システムのようにスケーラビリティを考慮した CMS においては、多数の計算機が本システムを利用し、多数のコンテンツが登録された場合でも、応答時間が利用に耐えうるものであることが求められる。

本評価では、CMS サーバが一定数のコンテンツを保持している状態において、CMS クライアントがコンテンツの検索要求を送信してから、CMS サーバから結果が返ってくるまでの時間を測定する。登録するコンテンツ情報は文字列であり、検索対象とするコンテンツ名は、アルファベットの連番と IP アドレスの結合による最大 18 文字の文字列である。また検索方法は、キーの完全一致検索、前方 6 文字の前方一致検索、中央 6 文字の中央一致検索の 3 種類をそれぞれ行う。CMS サーバの台数が 2 台から 4 台までのそれぞれの場合において、各 CMS サーバが保持するコンテンツ情報の数を 10,000 から 100,000 まで 10,000 単位で設定し、3 つの検索処理の処理時間をそれぞれ各 10 回測定し、平均値をそれぞれの検索時間とする。

この評価方法による結果を、検索方法別に図 5、図 6、図 7 に示す。図 5 が完全一致、図 6 が前方一致、図 7 が中央一致での検索を用いた結果である。完全一致の検索では、コンテンツ数や計算機の台数が増加した場合の検索時間の変化がほぼない。一方、前方一致の検索では、完全一致の検索に比べ、コンテンツ数の増加と計算機の台数の増加のそれぞれの場合とも、検索時間が若干増加している。しかし、その増加時間は小さいため、コンテンツ検索を行う上では問題なく運用できる範囲だと考えられる。また、中央一致の検索では、計算機の台数の増加により検索時間が若干増加している一方、コンテンツ数の増加では、前方一致の場合より検索時間が大きく増加している。

以上の結果から、完全一致検索において、コンテンツ数や計算機の台数の増加が検索時間に与える影響は低く、また中央一致検索においては、検索時間がコンテンツ数と計算機の台数に若干影響されることが分かった。一方、中央一致検索においては、検索時間がコンテンツ数に大きく影響されることが分かった。しかし、中央一致での検索時間への計算機台数の増加による影響は小さいため、1 台あたりで保持するコンテンツ数を抑えることで検索時間は短縮することも可能であり、コンテンツ検索において中央一致での検索が可能となることの利点は大きい。

4.2 外部サーバと連携時のコンテンツ取得時間

複数の外部サーバから本システムにコンテンツを登録済みのとき、他の計算機がコンテンツを検索し取得したときの処理時間を測定し、外部サーバと連携時の性能を評価する。提案システムのコンテンツ検索では、各外部サーバへ検索要求を送信せずとも、CMS サーバへの一度の検索要求で目的のコンテンツを探し当てることができる。このため、各外部サーバでコンテンツを直接検索し取得する場合に比べ、処理時間が短くなることが期待できる。本項目の評価により、本システムを介した分散コンテンツの検索や取得の実用性を示す。

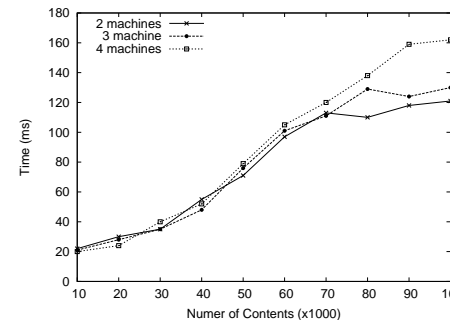


図 7 中央一致での検索時間

表 2 画像に対応するコンテンツ情報

項目	値
文書番号	1027
文書名	官庁文書
断・片	断片
横サイズ	26.7cm
縦サイズ	18.4cm
行数	6
文字種	1
コメント	唐開元26年(738)。 紙背絵画片付着、前裁せし断片。
URL	img/img0027.jpg

具体的には、以下の 2 つのコンテンツの取得方法におけるコンテンツ取得時間を測定し比較する。なおコンテンツ取得時間は検索開始からコンテンツの受信完了までの合計処理時間とする。

(a) 本システムを介して検索し外部サーバのコンテンツを取得

(b) 本システムを介さず外部サーバを直接検索し、外部サーバのコンテンツを取得

(a) の取得方法では、外部サーバが本システムへコンテンツ情報を登録済みの状態において、CMS クライアントから CMS サーバへ検索要求を行い、CMS ドライバを介して外部サーバのコンテンツを受信する。一方、(b) の取得方法では、各外部サーバへ別々にコンテンツ検索を直接行い、各外部サーバから直接コンテンツを受信する。

これらの実験において、現在我々が構築している古典籍デジタルアーカイブで取り扱っている「大谷文書」を評価用のコンテンツとした。このコンテンツは標本画像と対応する表 2 のようなコンテンツ情報で構成される。なお画像のファイルサイズは 7~14KB と比較的小さく、コンテンツ情報の値は数字、アルファベット、日本語のマルチバイト文字を用いている。また検索方法は、キーの完全一致検索、中央 4 バイト（全角で 2 文字）の中央一致検索の 2 種類をそれぞれ行う。この方法で、合計 100 のコンテンツを各外部サーバに所持させ、2 種類の検索を行い、外部サーバとなる計算機が 1 台から 4 台までのそれぞれの場合において、検索開始から画像の受信完了までに用いた時間を各 10 回測定し、平均値を処理時間とする。

(a) では、CMS サーバ用の計算機 1 台、CMS クライアント用の計算機 1 台を用い、このほか評価用の画像・コンテンツ情報と、コンテンツ情報を読み込み CMS サーバへ登録する

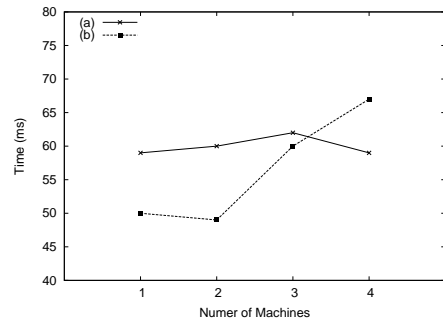


図 8 完全一致検索でのコンテンツ取得時間の比較

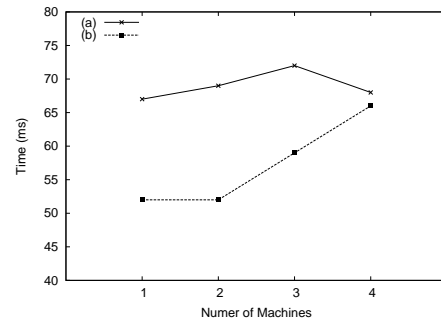


図 9 中央一致検索でのコンテンツ取得時間の比較

表 3 実運用サーバ統合時のコンテンツ取得時間 (ms)

検索方法	最小値	平均値	中央値	最大値
完全一致検索	17	56	61	99
中央一致検索	35	70	70	111

検索と取得の処理時間を評価する。外部サーバとして、古典籍情報を扱う 2 つの古典籍データベース A, B を用いる。データベース A, B はそれぞれ異なる古典籍画像とその文字情報を扱っており, A, B に分散しているコンテンツを本システムで統合する。

評価環境として, データベース A 用の CMS ドライバと B 用の CMS ドライバを用いる。これらの CMS ドライバは, データベースの文字列から図 2 のようなコンテンツ情報を抽出し CMS サーバへ登録する機能をもつ。また CMS サーバ 1 台と CMS クライアント 1 台を用いる。手順として, まず CMS サーバ 2 台を起動し, それぞれの CMS ドライバがデータベース A, B のすべてのコンテンツを CMS サーバへ登録する。そして, CMS クライアントが CMS サーバと接続し, コンテンツを検索後, CMS ドライバへ該当するコンテンツの取得命令を送信しコンテンツの送受信を行う。このとき, CMS クライアントの検索からコンテンツ取得完了までの時間を測定する。

外部サーバのコンテンツは画像とコンテンツ情報から構成され, コンテンツ数はデータベース A が約 1,000 個, データベース B が約 400 個である。画像のファイルサイズは 3~7KB であり, コンテンツ情報の値は数字, アルファベット, 日本語のマルチバイト文字を用いている。検索方法は, 文書番号や文書名の完全一致検索, 中央 4 バイト (全角で 2 文字) をキーとしたコメントの中央一致検索の 2 種類をそれぞれ行う。これらの 2 種類の検索について, 本システムへコンテンツ情報を登録済みの状態において, CMS クライアントが検索開始から 1 つの画像の受信完了までに用いた時間を各 20 回測定し, 最小値, 平均値, 中央値, 最大値を示す。

この方法による評価結果を表 3 に示す。結果では, 完全一致での検索, 中央一致での検索とともに, 最小値から最大値までのばらつきがある。これは画像サイズの違いによる受信時間の変化によるものだと考えられる。また全体として中央一致での検索の取得時間の方が完全一致での検索の取得時間よりも大きい傾向が見られる。これは, 完全一致での検索が文書番号や文書名など, 文字列が短く, ヒット数が少ないのに対し, 中央一致での検索では長いコメント中の一部分を検索するため, 多くの検索結果が得られるためであると考えられる。ただし, コンテンツの検索方法の違いによる処理時間の差よりも, コンテンツの転送のための処理時間の方が, 全体の処理時間としては大きく影響する。このため, 実際のコンテンツ

CMS ドライバを実装した計算機を 1~4 台用いた。手順としては, まず CMS ドライバが CMS サーバへコンテンツ情報を登録する。そして CMS クライアントが, CMS サーバへ検索要求を送り, 該当するコンテンツを CMS ドライバから受信する。

一方, (b) では, 評価用の画像・コンテンツ情報と所有するコンテンツのインデックスをもつ計算機を 1~4 台用いた。このインデックスはコンテンツ情報と画像のアドレスとを対応させリスト化したものである。検索を行う計算機は, これらの計算機からインデックスを取得し, 指定した項目とキーで文字列検索を行う。検索条件に合致するコンテンツがあった場合は, そのインデックスの所有者である計算機に対し, コンテンツの送信を要求することで検索したコンテンツを受信することが可能となる。

この方法による評価結果を, 検索方法別に図 8, 図 9 に示す。図 8 が完全一致, 図 9 が中央一致での検索を用いた結果である。これらと比較したとき, 外部サーバとなる計算機の台数が少ない場合, 直接のコンテンツ取得の方が, 本システムを介した場合に比べ取得時間が短い。しかし, 計算機台数の増加にともない, 直接のコンテンツ取得時間は大きく増加している。このため, 計算機が 4 台のときでは, 完全一致検索において本システムの方が取得時間が短く, 中央一致検索においては同じ程度の時間がかかっている。この結果は, 直接のコンテンツ取得では, 外部サーバが増加することでインデックスの取得や検索の回数が増え, 本システムにおける取得時間を超えるためだと考えられる。この結果より, 外部サーバの数が多環境では本システムの利用が有効であるといえる。

4.3 実運用サーバ統合時のコンテンツ取得時間

提案システムを用いて, 複数のデータベースサーバ上の分散コンテンツを統合したときの

操作に要する時間は、コンテンツのサイズに依存する所が大きい。

5. 関連研究

本章では分散コンテンツの統合に関するシステムを提案している関連研究を述べ、それらと比較したときの本システムの利点・欠点や本システムの改善案を述べる。

杉原らの研究³⁾では、P2Pにおいて、過去の検索履歴をもとにユーザの嗜好に沿った検索結果を提供する手法を提案している。具体的には、文書の検索において、あらかじめ文書ごとに各単語が含まれる割合を求め、単語の出現頻度のインデックスを文書ごとに作成する。検索するときは、各インデックスをもとに、キーワードの出現頻度がある閾値以上の文書のみを検索結果として返すことで、本当に必要な文書のみを通信負荷を抑えて提供することができる。提案システムでは、コンテンツ数が多い場合における中央一致検索の処理時間の増加が欠点となっている。この欠点に対し、杉原らの手法を適用して、中央一致検索をインデックス上の単語の完全一致検索に置き換えることにより、本システムの検索時間を改善できる可能性がある。したがって、事前の構文解析の処理時間を考慮した場合の検索時間の短縮効果を試算した上で、この手法を適用することも考えられる。

飯村らの研究⁶⁾では、BlogやWikiなど個人が情報発信できるメディア・プラットフォームを、P2Pで連携する情報共有環境を提案している。本システムと同様に通信プロトコルとしてJXTAを採用しているが、その利用法は本システムと大きく異なる。本システムはP2Pを構成する各サーバがコンテンツ情報を保持・探索し、サーバが互いの参加・離脱を認識することで柔軟な接続が可能なCMSを実現している。一方、飯村らの研究ではJXTAのグループ管理の機能を活用している。具体的には、各ピアが属するピアグループを定義し、グループ作成や各ピアのグループ参加・脱退などの機能を1つのスーパーピアに集約させることで、グループごとにピアやコンテンツへのアクセス権を設定することが可能としている。これにより特定ピアに対する情報の隠蔽ができ、例えば小グループでの強調作業時に、大グループでは共有すべきでないコンテンツが出てきた場合、大グループにはコンテンツを閲覧できないよう設定することができる。

6. おわりに

本稿では、複数のサーバに分散したコンテンツを一元管理するCMSについて述べた。本CMSは、JXTAを適用することで外部サーバやユーザの参加や離脱を柔軟に行え、コンテンツの検索や取得を効率良く行うことができる。

評価実験により提案システムでは、参加する計算機の台数や登録するコンテンツ数が増加した場合においても、コンテンツの検索時間の増加量は小さく、拡張性を保っていることを確認できた。また、本システムに参加する外部サーバが多い場合でも、各外部サーバから直接コンテンツを取得する場合と比べ、コンテンツ取得時間の増加量は小さいことを確認した。また、実際に運用している外部サーバ上のコンテンツを統合したときのコンテンツ取得時間は、サーバからの直接の取得に比べて差が小さく、ユーザが利用する際には支障ない程度である。したがって、コンテンツ数や計算機台数が多い環境において、本システムを利用することにより、コンテンツの取得時間の改善が期待できると考えられる。

今後は、本システムに登録するコンテンツ情報として、文字列以外に縮小画像の付加や、パケット分割したコンテンツ自体の登録など異なる登録手法の実装も試し、外部サーバの検索や取得の機能を維持した上で評価する予定である。また本稿の登録手法と比べた遅延時間を評価する予定である。

参考文献

- 1) Halepovic, E. and Deters, R.: Building a P2P Forum System with JXTA, *in The Second IEEE International Conference on Peer-to-Peer Computing*, pp.41-48 (2002).
- 2) Ismail, A., Merabti, M., Llewellyn-Jones, D. and Sudirman, S.: Sharing and Storing Serendipity Moments in Human Life Memory, *9th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting (PGNet 2008)*, UK, Liverpool, pp.23-34 (2008).
- 3) 杉原健司, 志田匡士, 吉永 努, 曾和将容: キャンパス P2P ネットワークにおけるパーソナライズ検索, 情報処理学会研究報告, 2005-SLDM-119, Vol.2005, No.27, pp.55-60 (2005).
- 4) Traversat, B., Abdelaziz, M. and Pouyol, E.: Project JXTA: A Loosely-Consistent DHT Rendezvous Walker, <http://www.jxta.org/docs/jxta-dht.pdf> Edited by: Dana Petcu Received: May 29, 2007 Accepted (2003).
- 5) Xiang, X., Shi, Y. and Guo, L.: Rich Metadata Searches Using the JXTA Content Manager Service, *AINA '04: Proceedings of the 18th International Conference on Advanced Information Networking and Applications*, Washington, DC, USA, IEEE Computer Society, pp.624-629 (2004).
- 6) 飯村卓也, 吉見圭司, 竹井菜奈子, 張 国珍, 金 群: P2P ネットワークを用いた異なるメディア・プラットフォーム間における情報共有の提案, 情報処理学会研究報告, 2006-GN-60, Vol.2006, No.60, pp.25-30 (2006).