

## 分散 WWW サーバを使った検索エンジンの隠蔽

木村聡宏 梶原清彦 飯尾和彦

NTT ソフトウェア研究所

〒180 東京都武蔵野市緑町 3-9-11

{kimura, kajihara, iio}@slab.ntt.co.jp

### 概要

インターネット / イン트라ネットの普及とネットワークセキュリティ技術の発展を背景として、エクストラネットが普及しつつある。Virtual Enterprise の実現形態の一つであるエクストラネットでは、各社の利用者は、従来通り既存の社内情報システムを利用し、情報の分散を意識することなく、社間にまたがって必要な情報を共有できることが重要である。しかしながら、現状では、各社の情報システムが各々異なり、異種情報システム間の相互接続が困難であるため、社間に分散した情報の共有も困難である。

本稿ではこれらの問題を分析し、我々が新たに設計したエクストラネット用の分散型情報共有システム CoOrbiter の情報共有機構と情報検索方式を提案する。CoOrbiter では、アダプタにより拡張された各社の WWW サーバを連携させることによって、利用者から他社の検索エンジンを隠蔽し、また、全体の制御機構を持たない分散型の情報共有環境を実現し、各社の情報システムの多様性と分散に対応している。

## Hiding of Search Engines Using Distributed WWW Servers

Akihiro KIMURA Kiyohiko KAJIHARA Kazuhiko IIO

NTT Software Laboratories

3-9-11 Midori-Cho Musashino-Shi Tokyo 180 Japan

{kimura, kajihara, iio}@slab.ntt.co.jp

### Abstract

Internet/intranet have been coming into wide use and network security technology is being developed in many companies. Extranet, private network on the Internet, uses these technologies for organizing virtual enterprise. Users on an extranet require other companies' information on the extranet to do their work. It means that information sharing is essential function on an extranet. However, it is difficult to establish information sharing on an extranet, because many companies on the extranet want to use their own information systems for information sharing.

In this paper, we present an overview of new WWW-based mediation system for information sharing for an extranet and information search mechanism using it. CoOrbiter, the name of the mediation system, realizes information sharing with no central control, flexible connection to existing information systems, and information search mechanism hiding many kinds of search engines.

## 1 はじめに

インターネット/イントラネットの普及とネットワークセキュリティ技術の発展を背景として、エクストラネットが普及しつつある。

エクストラネットは、VE(Virtual Enterprise)を実現するためのネットワーク形態の一つであるため、情報共有のための環境は必要不可欠である。エクストラネットでは、各社の利用者は、従来通り既存の社内情報システムを利用しながら、情報の分散を意識することなく、社間にまたがって必要な情報を共有できることが重要である。

しかしながら、現状では、各社の情報システムが各々異なり、異種情報システム間の相互接続が困難であるため、社間に分散した情報の共有も困難である。

本稿では、エクストラネット用に我々が設計した新しい分散型情報共有システム CoOrbiter における情報共有機構と情報検索方式を提案する。第2節では、エクストラネットの情報共有環境に求められる要件と現状の問題点について論じる。第3節では、問題解決のために我々が設計した CoOrbiter について概要を述べる。第4節では、CoOrbiter における情報検索方式について説明する。第5節では、JavaServer による CoOrbiter の実装例を示す。

## 2 情報共有環境に求められる要件

本節では、エクストラネットの情報共有環境に求められる要件と、既存技術を利用して、それらの要件を満たすときの問題点について論じる。

### 2.1 情報共有環境に求められる要件

エクストラネットによってVEを実現するとき、情報共有のための環境に求められる要件を列挙する。

#### 要件1 安価な構築・利用・運用コスト

元々エクストラネットは通信コストの安さが最大の魅力であり、コストが安価であることは重要である。

#### 要件2 セキュリティの保証

エクストラネットでは、参加企業間での情報共有を行うために、各社は、社内情報の一部

を他社に対して公開する。このため、セキュリティの保証は必須であり、(a) ユーザ認証とアクセス制御による、各社の内部情報に対する不正アクセスの防止と、(b) 暗号化による、イントラネット間の流通情報に対する盗聴の防止が必要となる。

#### 要件3 構成変更の柔軟性

エクストラネットには、参加企業各社が自由に離脱参加できるように、柔軟に構成を変更できることが求められる。

#### 要件4 参加企業各社の情報システムの独自性(他社の情報システムとの差異性)

エクストラネットには、各々異なる情報システムを持つ企業が参加するため、各社がエクストラネットに参加するにあたり、既存システムをできるだけ変更する必要のないことが重要である。

#### 要件5 参加企業間の関係の対等性

VEの実現形態の一つであるエクストラネットでは、情報の提供と利用は対等な関係で行われる。このため、参加企業各社の関係は、上下関係ではなく、対等な水平関係が前提となる。

### 2.2 現状の問題点

要件1～3は、インターネット標準の技術と市販製品や既存システムを利用することによって満たすことができる。

しかしながら、それらを利用するだけでは、要件4および5を満たすことは困難である。

例えば、エクストラネット用の情報共有環境の実現形態としては、現状では、情報共有用サーバを導入して、このサーバに共有情報を集める集中型の形態が一般的である。しかしながら、このような集中型の形態では、特定の社に情報共有用サーバの管理作業等が集中してしまうため、要件5の“参加企業間の関係の対等性”を満たすことができない。

要件4および5を満たすためには、各社に分散した情報を、特定のサーバに集中させることなく、分散したままの状態、利用者が分散を意識せずに共有できる“分散型の情報共有システム”が必要である。

### 3 エクストラネット用分散型情報共有システム—CoOrbiter—の設計

本節では、我々が問題解決のために設計したCoOrbiterについて概要を述べる。

#### 3.1 要件の機能要件へのマッピング

CoOrbiterを実現するために、第2節で示した要件4および5をそれぞれ機能要件にマッピングする。

要件1～3については、分散型の情報共有システムにおいても、インターネット標準の技術と市販製品や既存システムを利用することによって満たすことができるため、本稿では検討の対象外とする。

**機能要件4** 変更を社内に局所化して他社に影響を与えないこと、また、簡単に変更に対処できること

イントラネットが自由に離脱参入するエクストラネットでは、各社の社内情報システムに変更があった場合、変更の影響範囲を各社内に局所化し、エクストラネットに参加する他企業に影響を及ぼさないこと、また、各社は変更に対して簡単に対処できることが求められる。

**機能要件5** 分散型の情報共有機構

集中型の情報共有機構では、特定の企業に管理作業等が集中するため、参加企業間の対等な関係を実現するためには、分散型の情報共有機構が必要となる。

#### 3.2 機能要件へのアプローチ

CoOrbiterで前提とする情報共有の実現手段と、CoOrbiterで採用した機能要件4および5へのアプローチを説明する。

- 前提条件:

キーワード検索システムによる情報共有の実現

各社内における情報共有の実現手段には、各社の既存のキーワード検索システムを利用する。

キーワードベースの検索エンジンには、OSやWWWサーバに付属する単純なテキスト検索ツ

ルから、“Open Text INDEX”[1]のような全文検索システムや、“Oracle”[2]のようなRDBMSまで、様々な種類があるが、それらはイントラネット用の情報検索システムとして一般に普及している。

このアプローチを取ると、各社のイントラネットには、キーワード検索に対応する検索エンジンがあることが前提となるが、単純なテキスト検索ツールであれば、OSやWWWサーバに付属しているため、この前提については問題ないと判断できる。

他の情報共有方式には、“Yahoo!”[3]のような目次形式のディレクトリシステムや、LDAPやX.500等のディレクトリシステムがあるが、現状のディレクトリシステムは、異種システム間の連携が困難であり、また、イントラネット用の情報共有方式としては、現状ではあまり普及していないため、CoOrbiterでは検討の対象外とした。

また、“infoseek”[4]のようなキーワード検索システムとディレクトリシステムの併用型システムの場合には、キーワード検索システムのみを利用することとする。

- 機能要件4および5へのアプローチ:

拡張WWWサーバの連携による他社の検索エンジンの隠蔽ならびに共有情報の検索

各社の拡張WWWサーバを連携させる分散型の情報共有機構によって、各社の利用者に対して他社の検索エンジンを隠蔽するとともに、社間に分散した共有情報の検索を実現する。WWWサーバの拡張はアダプタの組み込みによって行う。

このアプローチにより、各社の検索エンジンに変更があった場合でも、変更の影響範囲を各社内に局所化し、他社に影響を与えずに済む。また、各社では、検索エンジンの変更に対して、各社のアダプタを変更することにより簡単に対処することができる。

分散型の検索エンジンには、Harvest[6]やIngrid[7]があるが、これらはいずれも、インターネットの巨大性に伴う問題の解決を目的とした検索システムであり、各社のイントラネットに分散した異種検索エンジンの連携はできない。

“たーぼさーち”[5]のような、単一の検索インタフェースから、複数のインターネット用検索エンジンを利用できるメタサーチと呼ばれるサービスもあるが、メタサーチでは、検索対象がインターネッ

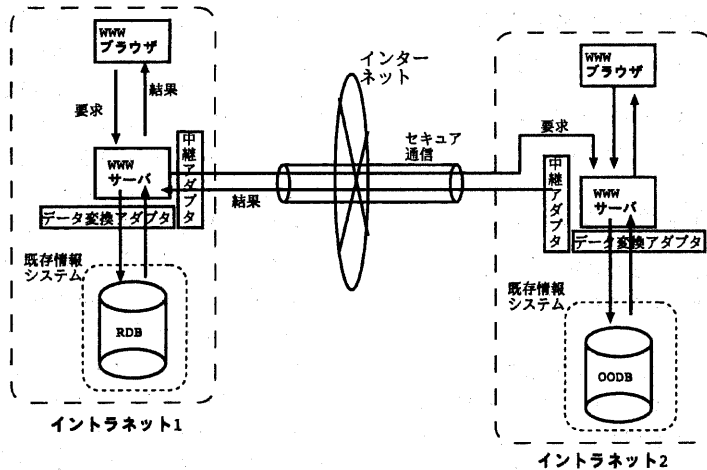


図 1: CoOrbiter の構成

トの公開情報に限られているため、イントラネットの内部情報の検索はできない。

### 3.3 CoOrbiter の特徴 (新規性) と構成

前節のアプローチに基づき我々が設計した CoOrbiter の特徴 (新規性) を列挙する。

- 全体の制御機構を持たない分散型情報共有機構
- WWW サーバと検索エンジンとの柔軟な連携が可能
- 利用者に他社の検索エンジンを意識させない情報検索方式

我々が設計した CoOrbiter の構成を図 1 に示す。

CoOrbiter の構成において、最も重要な役割を持つのはアダプタである。CoOrbiter では、WWW サーバに以下の 2 種類のアダプタを組み込むことによって、WWW サーバと検索エンジン間のデータ変換と、WWW サーバ間の中継を実現する。

**データ変換アダプタ** WWW サーバと検索エンジン間のデータ変換機能

図 1 では、データ変換アダプタは、WWW サーバと RDB、OODB 間のデータ変換を行う。

**中継アダプタ** WWW サーバ間の中継機能

図 1 では、中継アダプタは、WWW サーバが利用者から受け付けた要求を、他社の複数の WWW サーバに中継するとともに、中継先の複数の WWW サーバから送り返された結果を、要求元の WWW サーバに中継する役割を持つ。

具体的には、イントラネット 1 において、WWW ブラウザから出された要求を WWW サーバが受け付けると、中継アダプタは、イントラネット 2 の WWW サーバに要求を転送する。次に、イントラネット 2 の WWW サーバからイントラネット 1 の中継アダプタに対して結果が送り返されると、中継アダプタは、結果を編集してから、要求元の WWW ブラウザに編集済みの結果を提示する。

## 4 CoOrbiter の情報検索方式

CoOrbiter では、利用者から他社の検索エンジンを隠蔽するために、検索エンジンのフロントエンドである WWW サーバをアダプタにより拡張し、拡張 WWW サーバを連携させるというアプローチを取っている。

アダプタを利用することのメリットは3点ある。(a)各社の検索エンジンに変更を施さなくて済むこと、各社のイントラネットに構成変更等があった場合、(b1)各社のアダプタの変更によって対処できること、(b2)他社のアダプタや検索エンジンに影響を及ぼさないことである。

このアプローチにより、各社の利用者は、WWWブラウザを利用して、従来通り、自社のイントラネットの情報検索ページにアクセスして、他社の検索エンジンを意識することなく、社内情報だけでなく他社内情報までも含めた情報検索を行うことができる。

#### 4.1 検索処理のプロセスについて

利用者から検索要求が入力されると、どのような処理プロセスを経て、最終的に利用者のもとに検索結果が提示されるのか、検索処理のプロセスを、図2と関連付けながら以下に示す。

**ステップ1** 各社のWWWサーバが、利用者から検索要求を受け付ける(図2(a))

**ステップ2** データ変換アダプタが、利用者から入力された検索条件を自社の検索エンジン用に変換する(図2(b'))

**ステップ3** データ変換アダプタが、変換された検索条件を自社の検索エンジンに渡す(図2(c))。

**ステップ4** 必要があれば、中継アダプタが、変換された検索条件を他社のWWWサーバに渡す(図2(a'))

**ステップ4.1** WWWサーバが、他社のWWWサーバから検索要求を受け付ける(図2(a'))

**ステップ4.2** データ変換アダプタが、渡された検索条件を自社の検索エンジン用に変換する(図2(b'))

**ステップ4.2** データ変換アダプタが、変換された検索条件を自社の検索エンジンに渡す(図2(c'))

**ステップ4.3** 検索エンジンは、渡された検索条件に基づき、indexファイルを参照し(図2(d'))、検索結果を取得する(図2(e'))

**ステップ4.4** 検索エンジンは、取得した検索結果を、データ変換アダプタと中継アダプタ経由で、ファイル形式で検索要求元のWWWサーバに送り返す(図2(f'))(f'')

**ステップ5** 各社の検索エンジンは、渡された検索条件に基づき、indexを参照し(図2(d))、検索結果を取得する(図2(e))

**ステップ6** 各社の検索エンジンは、取得した検索結果を、データ変換アダプタ経由で、ファイル形式で要求元の社内のWWWサーバに送り返す(図2(f))

**ステップ7** 各社の中継アダプタは、社内および検索要求先の他社の検索エンジンから取得した複数の検索結果ファイルを編集し、一つにまとめてから、利用者に検索結果を提示する(図2(g))

##### 4.1.1 検索要求配布と検索結果取得のためのプロトコル

WWWサーバ間での検索要求配布と検索結果取得のためのプロトコルは、次の理由から"http1.0"とする。

1. 検索要求→検索結果という単純なやり取りである
2. インターネット標準である
3. ツールや情報が揃っている

検索要求については、インターネット用のキーワード検索サービスと同様に、利用者から入力された検索条件をPOSTメソッドにより検索エンジンに渡す。検索結果は簡単に処理できるようにするため、XML形式[8]とする。

#### 4.2 検索条件について

CoOrbiterでは複数の検索エンジンの利用を目的としているが、利用者はそれを意識せずに利用できなければならない。CoOrbiterでは、各検索エンジンにおける、他の検索エンジンから渡された検索条件の変換は、検索エンジンが各々個別にデータ

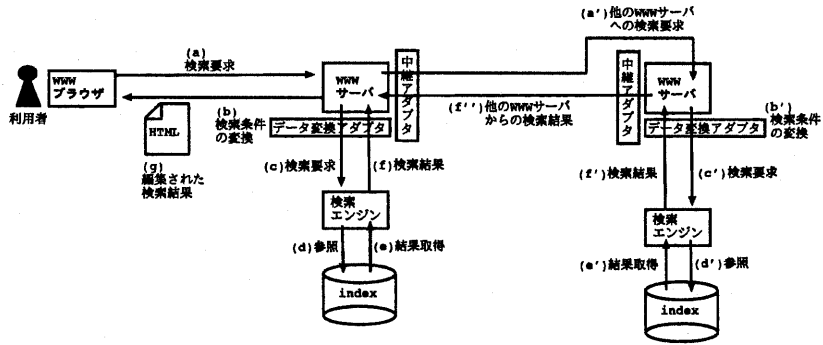


図 2: 検索処理の順序

変換アダプタをプログラミングすることによって対処することを基本方針としている。

本小節では、CoOrbiter で想定する一般的な検索条件の扱いについて説明する。

#### 4.2.1 検索条件

検索処理対象として想定する検索条件を列挙する。

##### 1. 検索文字列

- ワイルドカードを含む文字列
- 文字列間の関連付け  
{ "AND", "OR", "NOT", "NEAR" }

##### 2. 検索オプション

- 期間指定  
検索対象ドキュメントの作成日時に対して、期間指定により検索結果を絞り込む。

##### 3. 検索対象 WWW サーバ

検索対象とする WWW サーバを、{(a) 社内の WWW サーバのみ、(b) 利用者によって選択された特定の WWW サーバ、(c) すべての WWW サーバ}の中から指定する。デフォルトでは(c)とする。

##### 4. アクセス権限

検索対象ドキュメントに設定されたアクセス権に対して、ユーザIDとパスワードの2項目によって、検索時の利用者の権限を設定する。

#### 4.2.2 各検索エンジンへの検索条件の変換

各検索エンジンにおける検索条件の基本的な変換方針を、検索条件ごとに示す。

##### 1. 検索文字列

- ワイルドカード "\*" を含む文字列:

検索エンジンが "\*" に対応している場合には、 "\*" をそのまま利用する。対応していない場合には、 "\*" の前後に固定テキストがあるときには、 "\*" を "NEAR" に変換し、固定テキストがないときには、 "\*" を無視する。

- 文字列間の関連付け:

"AND", "OR", "NOT" は必ずあるので、そのまま利用する。 "NEAR" が無い場合には、 "AND" に置き換える。

##### 2. 検索オプション

- 期間指定

検索エンジンが期間指定に対応している場合には、検索結果の絞り込みを行う。対応していない場合には、期間を指定せずに検索を実行し、検索結果に含まれる作成日時を基に検索結果の絞り込みを行う。

#### 4.2.3 各検索エンジンに期待する検索結果

各検索エンジンは、検索結果をファイル形式で検索要求元の WWW サーバに返す。検索結果として、以下の情報が得られることを前提とする。

- 検索キーワードを含む各ファイルに対して：
  - － 検索キーワードにマッチした文字列の前後文字列
  - － ファイルの URL
  - － ファイルの作成日時
- 検索結果ファイルに対して：
  - － 検索エンジン名とバージョン

#### 4.3 検索結果のまとめ方について

検索対象の WWW サーバが複数ある場合、その数だけ検索結果ファイルが得られる。この場合の検索結果のまとめ方についての基本方針を以下に示す。

1. 検索エンジンごとに検索結果を分ける
2. 各検索エンジンの検索結果をマージする

検索結果をマージする場合、検索結果の整理には2種類の方法がある。

##### (a) 検索結果のソーティング

検索結果のソーティングは、さらに2種類の方法に分類できる。

- i. 各検索エンジンの検索結果のスコアリング情報を利用し、検索エンジン間で検索結果のスコアリング情報を再調整し、検索結果を再ソートする。
- ii. 検索キーワードと検索キーワードを含むファイルとの類似度を新たに計算し<sup>1</sup>、類似度の高い順に検索結果を再ソートする。

##### (b) 検索結果のグルーピング

<sup>1</sup>ベクトル空間モデル等により計算する

## 5 Java Server による実装

Java Server(TM)[9]は、Javaで記述されたWWWサーバであり、Java servletによる拡張が可能である。Java Serverを利用することのメリットとして、移植性、httpとの相性のよさ、拡張性の3点があげられる。

##### ● 移植性

servletはJavaにおけるサーバサイドAPIの標準であり、NetscapeなどのWWWサーバでもサポートされる予定である。

##### ● httpとの相性のよさ

URL表記でservletを起動できる。

##### ● 拡張性

servletをプログラミングすることにより、Javaの機能を活かした拡張が簡単に行える。

servletによりデータ変換アダプタと中継アダプタを実装したときのCoOrbiterの実装例を図3に示す。

## 6 おわりに

本稿では、分散した各社のイントラネットが緩やかに連結するエクストラネットでは、情報共有の環境が問題となることを指摘し、分散型の情報共有システムの必要性を述べ、我々が問題解決のために設計したCoOrbiterにおける情報共有機構と情報検索方式の提案を行い、JaverServerによるCoOrbiterの実装例を示した。

CoOrbiterでは、WWWサーバー間のルーティング方式、ディレクトリシステムおよびマルチメディアコンテンツ検索システムへの対応については検討中であり、今後の研究課題である。

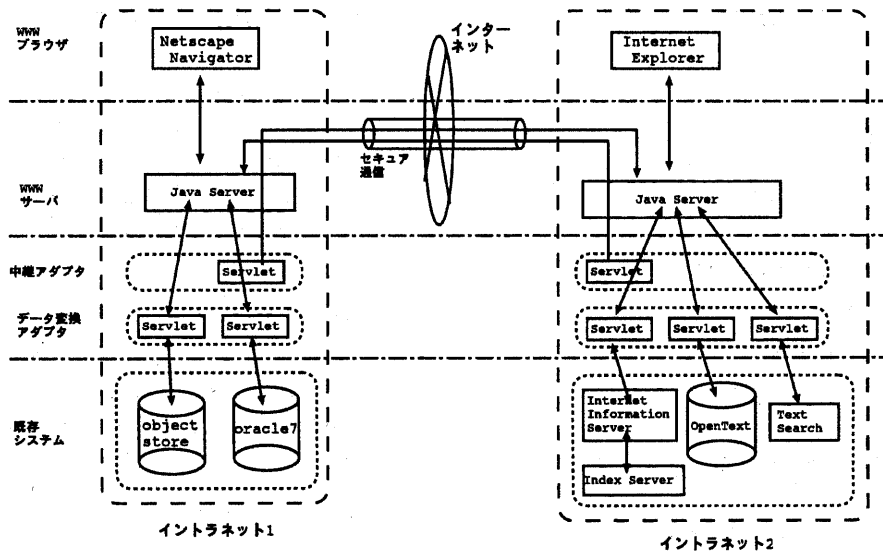


図 3: Java Server と servlet による CoOrbiter の実装例

## 参考文献

- [1] "Open Text INDEX"  
<http://index.opentext.net/>
- [2] "Oracle7"  
<http://www.oracle.co.jp/oracle7/index.html>
- [3] "Yahoo!"  
<http://www.yahoo.com/>
- [4] "infoseek"  
<http://guide.infoseek.com/>
- [5] "たーぼさーち"  
<http://www.st.rim/~wakasaya/>
- [6] C.Bowman, P.Danzig, D.Hardy, U.Manber, M.Schwartz: The Harvest information discovery and access system, Computer Networks and ISDN Systems, Vol.28,pp.119-125, 1995  
[\(http://harvest.transarc.com/\)](http://harvest.transarc.com/)
- [7] P.Francis, 神林, 佐藤, 清水: 次世代情報検索インフラストラクチャ Ingrid, NTT R&D, Vol.45, No2, pp.159-166, 1996
- [8] "XML"  
<http://www.w3.org/XML/>  
<http://www.ijnet.or.jp/FXIS/XSoft/sgml/xml/>
- [9] "Java Server"  
<http://jeeves.javasoft.com/products/java-server/>