研究計画書
価値のある情報を発掘する情報流通基盤

慶應義塾大学環境情報学部
自署：＿＿＿＿＿＿＿＿＿＿
学籍番号 79956832
平成 14 年 4 月 30 日

概要

本研究では、コンピュータ内で埋もれている情報を手間をかけることなく公開でき、自分の欲しい情報を的確に探し出すことのできる情報流通基盤を構築する。

1　研究の背景と問題意識

1.1　背景

研究や講義に使われた資料やレポート、あるいはソースコードの断片など、コンピュータを利用して有用な情報が作られている。しかしこれらの情報の大半は用が済んでしまった後はそのまま公開もされず、作成者のコンピュータの中に埋もれてしまっている。これらの情報を活かす為には情報を整理し公開するという作業が必要となるが、その手間に比べて情報を提供する側のメリットが小さい。その結果、それ以前に誰かが作成したのと同じような内容の情報が再び手間を掛けて作成されるという無駄が起きてしまう。

1.2　問題意識

このようなディスクの肥やしとなっている情報をネットワークを通じて流通させる事により、これまで車輪の再発明に費やしていた時間を使って、より価値の高い情報を作り出す事が出来るだろう。

この問題に対し、情報や知識を集積して一括管理を行うナレッジ・マネジメントツールと呼ばれるソフトウェアが存在し、企業等で一つのシステムを利用して行う体系的な知識管理には向いているが、一般の人同士で使おうとした場合にその集中的なアプローチは適さない。また、Peer to Peer モデルを利用したアプリケーションを利用することで手間をかけずに情報を共有することはできるが、内容に対し検索を行ったり、公開範囲を制御したりといった機能を持たないため実用的ではない。

結果として、現状では価値のある情報が各々のコンピュータ内に埋もれている。

2　研究の目的

コンピュータ内で埋もれている情報を手間をかけることなく公開でき、自分の欲しい情報を的確に探し出すことのできる情報流通基盤を構築する。

3　既存の技術

3.1　WWW

ほとんどのコンピュータで利用可能で、情報の公開にもっとも頻繁に利用されている。基盤技術として、識別子の URL、転送規約の HTTP、記述言語として HTML が利用されている。それ自体は検索機能は無いため、情報を発見するには能動的に情報を収集してインデックスを作成する検索エンジンに依存せざるをえない。しかし、これらのサービスはインターネット全体を対象としているため、WWW サイトの増加に伴い情報を探すことが難しくなっている。また、主なインターフェースとしてリンクを利用しているため、情報を公開する際にはそれを整理し、各々をリンクで結ぶという手間がかかる。

3.2　Peer to Peer

Peer to Peer モデルを採用するアプリケーションは、二つの観点から分類することができる。

検索方法　ファイルの一覧とクライアントの位置の管理にあらかじめ指定されたロビーサーバを利用する方法を Hybrid P2P モデルと呼ぶ。ロビーサーバから目的のファイルを所有するクラ

イアントの IP アドレスを取得し、ファイルの転送のみを直接相手に接続して行う。このモデルではロビーサーバに障害が発生するとネットワーク全体が利用不可能となる。
ロビーサーバに頼らない方法として、検索要求をネットワーク全体にブロードキャストすることで検索を行う方法を Pure P2P モデルと呼ぶ。このモデルでは特定のノードに依存しないため、どのノードがネットワークから切り離されてもネットワーク全体としては機能しつづける。検索要求によってネットワーク全体が過負荷にならないように HTT(Hop To Live) と呼ばれる寿命を設定するが、ネットワークの規模が拡大すればするほど目的のデータを持つノードまでたどり着けない可能性が高まり、検索の正確性が落ちるという欠点がある。

キャッシュの有無 各ノードがデータを別のノードから取得した場合に、そのデータをノードとしてキャッシュしてさらに別の第三者に再配布するかどうかでネットワーク全体の性質が変化する。
キャッシュを行わない場合には各ノードが持つデータはたとえそれが他のノードから取得したものであっても別個のものとして扱われる。
キャッシュを行う場合は、最初の所有者がネットワークから切り離されてもそのデータを他にダウンロードしたノードから取得することができる。また、人気のあるデータが存在した場合にキャッシュが無い場合はその所有者にアクセスが殺到するが、そういったデータはキャッシュも増えるためボトルネックが発生しにくい。

3.2.1 Napster[1]

Napster 社によるサービスで、音楽データの共有を目的として設計されたアプリケーションである。Napster では、”Hybrid P2P” モデルを採用し、キャッシュは提供しない。また、Napster 社の提供するロビーサーバでは、著作権上の問題が有るデータの流通を制限している。

3.2.2 Gnutella[2][3]

”Pure P2P” モデル利用しているが、キャッシュは提供していない。

3.2.3 freenet[4]

匿名の情報保存・検索ネットワークとして設計された。”Pure P2P” モデルを採用しキャッシュも提供しているがそれをさらに発展させている。アップロードされたデータはネットワーク全体のノードに分散して保存され、データの提供者や取得者が分からないように設計されている。

また、Pure P2P モデルの欠点として検索要求によってネットワークに負荷がかかる点があるが、通信路をうまく動的に効率化することで、検索要求がより小さな範囲にとどまるように工夫されている。

freenet 単体では平坦な名前空間を持つ分散ファイルシステムを提供するが、freenet 上に掲示板を提供する frost や、Web を提供する FreeWeb などのアプリケーションが存在している。

匿名性を確保するため検索して発見されたデータ本体も複数のホストを中継して転送されるが、この経路制御にネットワークトポロジは反映されず見かけ上ランダムに結ばれるため実際のネットワーク的に遠いリンクが数多く存在し、経路中最も細い部分によって帯域幅が制限される。

4 Peer to Peer ネットワークの構築

4.1 要求事項

コンピュータの中に埋もれている創造活動の成果を共有する為には以下の条件を満たさなければならない。

4.1.1 ネットワーク透過

端末がネットワーク上のどこに居ても、その端末内のデータを検索して取得することができなければならない。

4.1.2 索引

既存のアプリケーションでは、検索のキーとしてファイル名を用いていることが多い。また、Napster 等では、それに加え音楽データにつけられたメタデータ [6] から曲名やアーティスト名、音質を示す符号化情報による検索もできる。

ファイル名やメタデータの使い方や正確さはそれを設定した人に大きく依存する。さらに、ファイルの内容は一切加味されないため、検索の条件としてはあまりに不十分である。本研究で実装するアプリケーションでは、より多角的な検索のできる索引を採用しなければならない。

また、必要なデータを確実に取得しなければならないことから、検索の確実性が必要不可欠である。データを取得する際に所有者の端末がネットワークに接続されていない場合、そのデータをキャッシュしている端末が他にあれば取得できなければならない。また、携帯端末を考慮してネットワークに接続されている間にデータをキャッシュする端末があればデータ取得の確実性が向上する。

また、複数のファイル間の関係を示すリンク情報を共有することで、このリンク情報を利用して検索精度の向上を図ることができる。また、たとえば引用が行われた場合に引用元のデータに対するリンクを作成することで、そのデータの閲覧者が引用元となった 1 次情報のデータを検索できるようになるといったことが考えられる。

4.1.3 ACL

Peer to Peer モデルでは Server へのアップロード作業が必要なくなるとはいえ、公開すべきデータと公開したくないデータが存在する。また、特定のグループに含まれる者には公開してよいが、それ以外には公開したくない、というケースが考えられる。必要な場合に相手を認証できなければならない。また、データ単位での ACL (アクセス制御リスト) も必要となるが、あまりに高度に設定可能なシステムは設定が難解となってしまう。したがって、容易に ACL を設定できるユーザインターフェースが必要不可欠である。

4.1.4 versioning 管理

テキストやグラフィックデータではデータの変更があり得る。既存のシステムでは少しでも変更されたデータは全く別の存在として扱われている。また、そうでなくても他の人が持つデータが同じかどうかは、md5 などのハッシュを利用するか、人間が判断するしかない。

また、マルチメディアデータにおいても、全く同じ内容にもかかわらず符号化方式やそのパラメータによって、同じ内容に対しバイナリとして異なるデータが無数に存在する。

このような似た内容で異なる内容のものをグルーピングして扱う事で求めるデータの発見がしやすくなりネットワーク全体の効率が上がると共に、例えば自分の持つデータに対し、より新しいものが存在するか調べる事が出来る様になる。

表 1: 既存のアプリケーションとの違い

	索引	ACL	Cache
WWW	検索エンジン	あり	なし
Napster	Server	なし	なし
Gnutella	Broadcast	なし	なし
freenet	Broadcast	なし	あり
本システム	分散 Server	あり	あり

4.2 設計

上記の条件を満たすため、新しいアプリケーションを設計・実装する。

ここで実装するシステムは、ネットワーク的に分散して存在する複数のノードから成り立つ。各ノード、以下に示す機能から成り立っているが、後述する通り必ずしも全ての機能を保有しない。図 1 として概念図を示す。

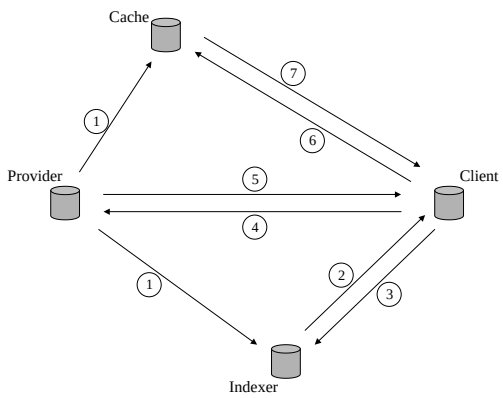

図 1: 構成図

プロバイダ データの所有者。クライアントからのリクエストに応じてデータを提供する。また、ファイルの検索を作成し、インデクサに送信する。ACL を維持し、制限がかかっている場合にはクライアントに対し認証を行う。

クライアント ユーザから要求されたデータをインデクサを利用して検索し、プロバイダあるいはキャッシュから取得する。

インデクサ プロバイダからデータの索引を収集し、クライアントからの検索に応じる。また、プロバイダの位置を追跡し、オフラインであれば利用可能なキャッシュを探す。

キャッシュ プロバイダの所有するデータを複製・

保存する。プロバイダがオフラインである場合にプロバイダに代わってクライアントにデータを提供する。

一般的に、ユーザが携帯する端末ではプロバイダとクライアントの機能を持ち、インデクサおよびキャッシュは自宅の端末やデータセンタなどのネットワーク的に固定された場所に設置することを想定している。

他の検索サーバを利用できない場合は任意の端末にインデクサとしての機能を持たせることで、独立したネットワーク環境においてもシステムを利用することができる。

5 評価方針

5.1 ソフトウェアの評価

本研究で作成するアプリケーションが要求事項を満たしていることを確認する為、以下の点から評価を行う。

- ノードの位置に関らず検索・データの転送が行えるかどうか。
- データの総数・容量あるいはノードの総数が増加した場合に実用的な範囲で検索ができるかどうか。
- 同様にキャッシュが十分に機能するかどうか。
- 認証およびアクセス制御は十分であったか。またその設定は容易だったか。
- アプリケーションのユーザ・インターフェースは十分使いやすいものか。

5.2 コミュニケーション基盤としての評価

本研究が目標とするコミュニケーションが実現されたかを確認するため、以下の点から評価を行う。

- それまで利用頻度の低かったデータが再利用されたか
- 自分の欲する情報を探し出すことができたか。

6 予想される成果

以上のようなコミュニケーション基盤を実装・運用することにより、最低限の操作で自分が持っている情報を公開することができ、自分の欲しい情報を探し出し、取得することができる。その結果、社会全体がより価値の高い情報の生産に専念でき、技術や文化の発展に貢献する。

7 これまでの活動

徳田・村井・楠本・中村・南合同研究会 環境情報学部入学当初より徳田・村井・楠本・中村・南合同研究会に所属し、高度なネットワークインフラの構築及び運用に携わっている。

pie[7],sprng[8] 大学 1 年春よりインターネットの普及に関する研究グループに所属し、インターネットインフラが普及する上での問題点や、普及が社会に与える影響について研究を行っている。

neco[9] 大学 3 年春よりネットワーク上のコミュニケーションに関する研究グループに所属しネットワーク上で行われる新しい形態のコミュニケーションの研究に携わっている。

8 志望理由

本研究では、新しい情報流通ネットワークモデルを提案し、構築する。これを実現する為には、ネットワーク技術のみならず、効果的な索引の作成や検索にかかわるマルチメディアデータベースや、アプリケーションとしてより多くの人に使ってもらう為のユーザインターフェースに関する議論が必要不可欠である。このように多角的な面から研究を進めるにあたり、これらの分野の研究が盛んに行われている政策・メディア研究科を志望したい。

9 共同研究者・関連団体

本研究では、慶應義塾大学大学院政策・メディア研究科 “モービル広域ネットワーク (MAUI)” プロジェクトにおいて、村井純教授、楠本博之助教授、中村修助教授の指導のもとに行う。また、本研究の基礎技術として必要不可欠なデータベースおよびユーザインターフェースの教員の指導を受けたいと考えている。

また、以下の団体と協力しながら研究を進める。

- 慶應義塾大学村井研究会内ワーキンググループ neco
- WIDE Project

参考文献

[1] Napster
http://www.napster.com

[2] Gnutelliums
http://www.gnutelliums.com

[3] Jnutella.org
http://www.jnutella.org

[4] FreenetProject.org
http://www.freenetproject.org

[5] Groove Networks, Inc.,
http://www.groove.net

[6] id3.org
http://www.id3.org

[7] インターネットの普及に関する研究グループ
http://www.sfc.wide.ad.jp/kg/pie/

[8] 通信基盤としてのインターネット整備戦略、およびその上での政策・社会環境の整備に関する研究
http://www.sfc.wide.ad.jp/kg/sprng/

[9] ネットワーク上のコミュニケーションに関する研究グループ
http://www.sfc.wide.ad.jp/kg/neco/