

オーバレイネットワークを用いた 個人サイト特化型分散 Web 検索システムの設計

豊田 正隆[†] 山崎 賢悟^{††} 勅使河原 可海^{††}

創価大学工学部[†]

創価大学工学研究科^{††}

1. はじめに

現在のサーチエンジンは、インデックスの更新間隔の長さのため、公開されて間もない情報の検索を行うことができないという問題を抱えている。多くのサーチエンジンでは、多数の人にとって価値の高いニュースサイト等を頻繁にクロールすることで問題の部分的解決を図っているが、少数の人にとってのみ価値の高い個人サイト等に関しては、逆にクロール頻度が下がってしまう。個人サイトにおいて公開される情報にも価値の高いものは多く、それらの情報を即座に検索することへの要求も多い。そこで、これまでに個人サイトで公開されて間もない情報を発見することを目的とした分散 Web 検索システムの研究を行ってきた[1]。

分散検索システムは、システムを構成するサーバがどのようなオーバレイネットワークを構成するかによって性質が変化する。本稿では、個人サイトを検索対象とした際に有効であると考えられるオーバレイネットワークの検討について述べる。

2. 個人サイトの特徴

個人サイトには以下の特徴があると考ええる。

- ・何らかのテーマとカテゴリに分類できる

本研究におけるテーマとは個人サイトが中心的に扱う話題、カテゴリとはテーマをさらに細分化したものと定義する。個人サイトは JAZZ やプログラミングといった、何らかのテーマを持っている。またテーマだけではなく、JAZZ ならレビューを扱うサイトや社会人バンドのオフィシャルサイト、プログラミングならソフトウェアを公開しているサイトや Tips を集めたサイトなどといったカテゴリに分類することが可能である。

- ・数が膨大である

総務省による平成 15 年通信利用動向調査[2]によると、世帯・個人でパソコンを用いてインターネットを利用している人口は 6,164 万人に達

している。また、インターネット白書 2004[3]によると、自宅からインターネットを利用している 16 歳以上の男女個人で現在個人ホームページを開設運営している人の割合は 15.7% となっており、個人が運営している Web サイトは決して少なくないことが分かる。現在個人ホームページを開設していない人でも、今後開設する予定がある人は 19.5% となっており、今後も個人が運営する Web サイトは増えていくと見られる。

3. システムの形態

分散 Web 検索システムでは、特定のカテゴリに属するサイトのみに関して検索を行えば、求める情報を効率的に発見できると考えられる。そのため、システムを構成するサーバを、各カテゴリに関する検索を行うグループに分けることで、全てのサーバに対して検索要求を転送する必要がなくなる。

そこで、本システムを構成するサーバは図 1 に示すオーバレイネットワークを構築する。各サーバはリンクで直接結ばれているサーバの情報を保持する。システムに送られた要求は、主にこのネットワークにおけるリンクを辿って転送される。

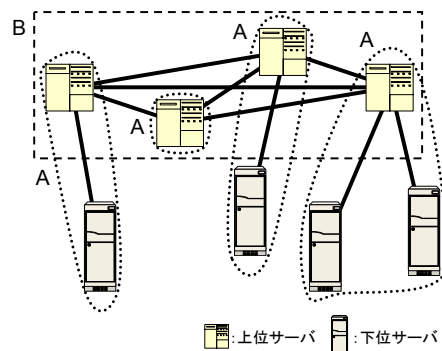


図 1 システムで用いるオーバレイネットワーク

本システムを構成するサーバは、図 1 の A に示される複数のスター型トポロジを構成する。スター型トポロジの中央サーバを上位サーバ、それ以外のサーバを下位サーバと呼び、1 つの上位サーバと上位サーバに直接接続されている下位サーバ群から構成される単位をブロックと呼ぶ。ブロックは前述したカテゴリ毎に分けられた Web サイトを担当するサーバのグループに相

A Design of a Distributed Web Retrieval System by Using an Overlay Network for Personal Web Sites

Masataka Toyoda[†], Kengo Yamazaki^{††}, Yoshimi Teshigawara^{††}

[†] Faculty of Engineering, Soka University

^{††} Graduate School of Engineering, Soka University
mttoyoda@soka.ac.jp

当する．単一のカテゴリに関する検索要求は，主に1つのブロック内のみで転送される．

上位サーバは図1のBに示される完全グラフ型トポロジを構成する．複数のカテゴリに関する検索要求や，システム全体に影響する情報は完全グラフ型トポロジを介して転送される．

4. システムの構成要素

システムを構成する要素について解説する．

4.1 上位サーバ

上位サーバは下位サーバを1つも配下に持たない，あるいは1つ以上の下位サーバを配下に持つ．上位サーバは以下に示す機能を持つ．

- ・ 同じブロックのサーバに対して検索要求を転送する
- ・ Web サイトの登録要求があった際に，同じブロックのサーバから最適なサーバを選択し，Web サイトを担当させる
- ・ Web サイトの脱退要求があった際に，Web サイトの情報をシステム内の各サーバから削除させる
- ・ サーバの追加／削除要求を受け付ける

4.2 下位サーバ

下位サーバは必ずいずれか1つの上位サーバの配下となる．下位サーバが担当するサイトのカテゴリは，その下位サーバが属する上位サーバが担当するサイトのカテゴリと同じである．

上位サーバと下位サーバは以下の機能を共通して持つ．

- ・ 自身が担当する Web サイトに関してクローリング，インデックス作成，検索を行う
- ・ 上位サーバからの要求を受けて検索やサイトの登録，保持している情報の更新を行う

4.3 サイト

本システムによる検索の対象となるサイトは，その運営者によってシステムが扱っているカテゴリのいずれかに登録される．複数のカテゴリに跨って登録することも可能である．

サイトの登録を行う際，サイト運営者は登録したいカテゴリを扱う上位サーバのいずれか1つに対してサイト登録要求を送る．サイト登録要求を受け取った上位サーバは同じブロック内のサーバから最も負荷の少ないサーバを選択し，そのサーバにサイトを担当させる．複数のカテゴリに登録する場合，その他のカテゴリを扱っているブロックにサイトを担当しているサーバの情報を通知し，検索時にはそのサーバに検索要求を転送するようにする．

5. 期待される効果

このようなオーバレイネットワークを用いることで，以下の効果が期待される．

- ・ 負荷が軽減できる

サイトを登録する際に，登録するカテゴリを扱うブロック内のサーバのうち，最も負荷の小さいサーバがそのサイトのクローリングを行うため，負荷分散が可能である．また，検索時には検索するカテゴリを扱うブロック以外には要求が転送されないため，負荷が軽減できる．したがって，登録されるサイトが増大した場合でも負荷が大きくなりにくい．

- ・ 拡張性が高い

システムに上位サーバを追加することで，システムで扱うことが可能なカテゴリを簡単に増やすことができる．また，システムに下位サーバを追加することで，1つのカテゴリで扱うことが可能なサイト数を簡単に増やすことができる．サーバを追加することによる負荷の増加は上位サーバによる要求の転送だけであるため，サーバの追加による負荷の増加は少ないと考える．

- ・ 応答時間が比較的短い

分散型検索システムは検索要求を転送するため，システムを構成するサーバが増加すると要求の転送回数も増加し，応答時間が長くなる傾向にある．本システムは要求を転送する回数がサーバの数に比例しないため，分散型検索システムとしては短い応答時間を維持できる．

6. まとめと今後の課題

本稿では，我々が研究している個人サイト特化型分散 Web 検索システムで採用するオーバレイネットワークの形態と動作，利点を述べた．

今後は，今回検討したオーバレイネットワークについて，実際に運用して有効性の評価を行っていききたい．また，自律的にネットワーク構成を変更できるような手法を考えていきたい．

参考文献

- [1] 豊田正隆，山崎賢悟，勅使河原可海：RSS の存在に依存せず新設／更新された情報を即座に収集する分散 Web 検索システムの提案，第3回情報科学技術フォーラム一般講演論文集第4分冊，pp.53-54，2004.9
- [2] 平成15年通信利用動向調査：
http://www.soumu.go.jp/s-news/2004/040414_1.html
- [3] 財団法人インターネット協会：インターネット白書2004，インプレス，2004.7