

# 進化的情報探索エージェントについて

阿部康一 武田利浩 丹野州宣

E-mail: kouichi, takeda, tanno @etn.yz.yamagata-u.ac.jp

山形大学工学部電子情報工学科

近年におけるインターネットの急激な普及により、情報の急激な増加と分散化が進んでいる。本研究の目的は、このような大規模なコンピュータ・ネットワーク上の多大な情報資源から、ユーザが必要とする情報を効率良く見つけ出すこと、また、それにとりまなう情報管理方式について考えることである。本文は、マルチエージェント・モデルを用いた分散情報共有システムのプロトタイプ・システムを提案し、その構成と機能について示す。さらに、実際に情報を探し出す情報探索エージェントについて述べる。

## An Evolution of Information Retrieval Agents on Distributed Computer Networks

Kouichi ABE, Toshihiro Taketa, Kuninobu Tanno

E-mail: kouichi, takeda, tanno @etn.yz.yamagata-u.ac.jp

Department of Electrical and Information Engineering, Faculty of Engineering, Yamagata University

The current exponential growth of the Internet precipitates the increase and the dispersion of information. The purpose of our research is to build an efficient information sharing by means of an original distributed information management and retrieval methodology. In this paper, we suggest a prototype system using multi-agent model, and describe its structure and functions. We describe the information retrieval agents as well.

## 1 はじめに

ここ数年のインターネットに対する関心の高まりと急激な普及、特に World-Wide Web (WWW) のユーザの増加により、誰もが容易にかつ自由に情報を発信しコンピュータ・ネットワーク上に蓄積された様々な情報を検索し、収集することができるようになった。しかしそれにともない、コンピュータ・ネットワーク上に爆発的な勢いであらゆる情報が無秩序に分散、蓄積されるような状況が生じている。そのため、ユーザが自分の必要とする情報をこの膨大な情報資源からどのようにして探し出すかが最大の問題となってきており、近年多くの研究開発が行なわれている [1]。

この問題を解決する方法として、カタログを利用した検索方式が広く使用されている。この方式では一般に、あらかじめ対象とする情報に関するデータベース(カタログ)を作成しておき、ユーザの検索要求を満たす情報をカタログから検索して、その情報の一覧をユーザに提示する。しかしこの場合、ユーザが情報検索サービスを行なうことのできる検索サーバーを明確に知っているということが前提である。また現在のように、WWW や USENET News などの情報量を見てもわかるように、単一のホストで情報資源全体を把握するのは困難である。このように情報資源の極度の分散化、大規模化が現代の情報検索における大きな問題となってきている。

本研究では、従来の単一ホストによる集中型の検索情報管理法に対して複数のホストによる分散型の検索情報管理法を提案する。しかし、そのような分散型情報管理法においてはユーザが必要な情報資源を効率的に見い出すことは容易でない。そこで、マルチエージェントモデルを導入し、その自律的なマイグレーション機能の利用および遺伝的アルゴリズムによるエージェントの自律再生産と最適エージェントへの進化を介してより効率的な情報資源探索を実現する。

本文は、以下のように構成される。まず最初に、コンピュータ・ネットワーク上における情報資源について述べ、分散した情報資源の管理方式とそれに応じた情報検索方式について提案する。次に、この方式に基づいて現在構築中であるプロトタイプ・システムの構成と機能を示す。続いて、プロトタイプ・システムにおける情報探索<sup>1</sup>で使用する情報探索エー

ジェントについて説明する。最後に、既存のシステムと本システムとを比較し、今後の研究開発の方向性について述べる。

## 2 コンピュータ・ネットワーク上の情報資源について

### 2.1 情報資源と利用法

コンピュータ・ネットワーク上における情報資源としては、現在以下のようなものが広く利用されている。

- World-Wide Web: インターネット上の分散された情報をハイパーテキスト・ネットワークでまとめた分散情報システム
- USENET News: インターネット上の世界規模の電子掲示版システム
- anonymous FTP Service: インターネット上のフリーウェアやドキュメントなどを提供するサービス

これらの情報資源は、それぞれ独立したサービスとしてコンピュータ・ネットワーク上に数多く存在している。ユーザは、目的の情報資源にアクセスするために、各サービス毎に異なるクライアント・アプリケーション(ユーザ・インタフェース)を使用しなければならない。

最近では、WWW 用のブラウザがこれらの情報資源にアクセスするための共通のユーザ・インタフェースとして使用されている。

### 2.2 分散情報管理

現在の情報検索システムのほとんどが、一つのサーバーで各情報資源の情報カタログを管理している(図1)。しかし、インターネットのような大規模なコンピュータ・ネットワーク上の全情報を一つのサーバーで管理するには限界がある。本研究では、以上の問題を解決するための一つの方法として、図2に示すようなシステム構造を提案する。このシステム構造

<sup>1</sup>本研究では、情報資源を求めるためにコンピュータ・ネットワーク上を探し回ることを探索と言ひ、情報資源から情報を探すことを検索と言ひ。

では、各々の情報資源が各自の情報カタログを所有する情報検索サーバーになることである。このとき情報の探索は、専用の情報探索エージェントを用いて実行される。

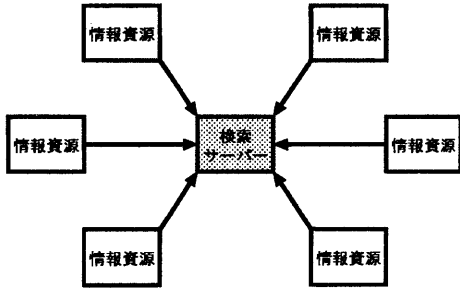


図1: 一般的な検索システムの構造

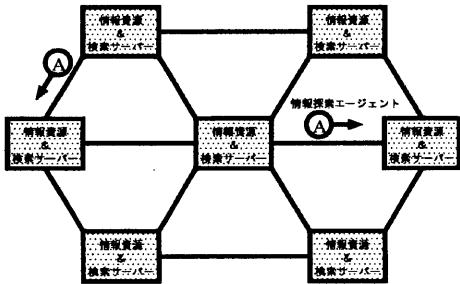


図2: 本研究で提案する検索システムの構造

### 2.3 情報検索システム

コンピュータ・ネットワーク上に存在する各種情報検索サービスは、一般にある限られた情報資源を対象としたシステムであることが多い。そこでは、同一の情報を提供する情報資源から定期的にそのホストが所有している情報のリストを受け取り、コンピュータ・ネットワーク全体とはいかないが、大規模な情報カタログを作成し、それを使用することでユーザからの要求に応じている。

しかし、ユーザがこれらの情報資源から必要な情報を探し出すためには、各情報資源毎に専用の情報検索サービスを使用しなければならない。ここで、問題となるのはユーザがそれらの検索サービスを提供しているホストを予め知っていなければならないことである。また、各検索サービス毎にユーザ・イ

ンタフェースも異なることも問題として上げられる。本研究では、これらの問題を情報の分散管理と情報探索エージェントによる検索方式で解決を図る。

## 3 分散情報共有システム

### 3.1 マルチエージェント・モデル

本研究では、システムの実現に際してエージェントの概念を導入する。現在、エージェントという用語は人工知能分野とコンピュータ・ネットワーク分野において広く使用されている [3]。本研究では、エージェントを次のように考える。

『人間や他のものの直接的な支配・制御を受けることなく動き(自律性)、ある種のコミュニケーション言語を通じて、その他のエージェント(あるいは人間)と相互に影響し合う(社会性)もの。』

またマルチエージェント・モデルとは、複数のエージェントがお互いに協力してある目的を達成するモデルをいう。

### 3.2 システムの構造

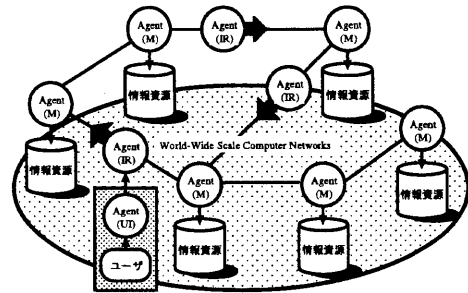


図3: システムの概念図

コンピュータ・ネットワーク上における本システムの構造は、図3に示すモデルとして表すことができる。図3からも分かるように、本システムは3つの機能別領域に分割することができる。すなわち、ユーザ・インタフェース領域と情報探索領域、情報管理領域である。これらは、それぞれユーザ・インタフェース・エージェント、情報探索エージェント、情報管理エージェントにより機能する。ユーザ・インタフェース・エージェントは、情報検索を行なう

際のユーザへの窓口を担う。情報探索エージェントは、ユーザ・インタフェース・エージェントから呼び出されコンピュータ・ネットワーク上で情報資源を探索し、目的の情報をそらの情報資源から検索する。情報管理エージェントは、各情報資源を管理するエージェントである。

### 3.3 システムの機能

本システムは検索サーバーに依存しない、動的にホスト環境に適応する情報探索システムを、究極的な目標としている。これは、どのようなプラットフォームにおいても特別な設定なしで、そのプラットフォームに適した情報探索環境を構築することである。また、ユーザから見れば単に必要な情報に関するキーワードを入力するだけで、目的の情報を入手することができることが理想である。

現在のところ、以下のような機能の実現を目標としている。

1. 検索サーバー独立な情報探索機能:  
決まった検索サーバーに問い合わせるのではなく、情報探索エージェント自らが検索サーバーを決定し、検索しに行くこと。
2. 動的な環境適応機能:  
情報探索エージェントがクライアント・ホストの環境とユーザからのリアクションによって、そのクライアント・ホストとユーザに対して最適な情報探索エージェントへと進化していくこと。
3. 情報カタログによる統一した情報管理機能:  
WWW、USENET News などの情報資源を区別することなく、一つのサーバーで情報カタログによって統一して管理すること。つまり、これは WWW 専用の検索サーバーとか USENET News 専用の検索サーバーとかと言う区別がなくなることである。
4. ブラウジング機能:  
探索結果である情報資源からの情報を、ユーザに視覚的に表示すること。

## 4 情報探索エージェント

### 4.1 遺伝的アルゴリズム

遺伝的アルゴリズム (Genetic Algorithms: GA) は、生物進化(自然淘汰、突然変異)の原理に着想を得たアルゴリズムである。遺伝的アルゴリズムでは、図4に示すように問題の解候補を遺伝子として表現する。この遺伝子によって構成されるのが染色体であり、染色体上の遺伝子の位置を遺伝子座、遺伝子の組合せパターンを遺伝子型という。また、この遺伝子型に基づいて形成された個体を表現型といい、これらの個体が進化することによって解を求めるアルゴリズムである。個体の進化は、3種類の遺伝的操作(選択淘汰、交叉、突然変異)を使用することによって行なわれる [2]。本研究では、情報探索エージェントが個体として扱われる。

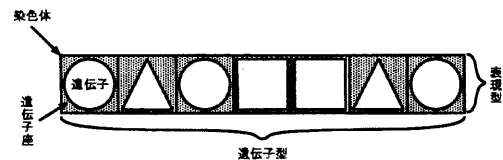


図4: 遺伝的アルゴリズムにおける個体

### 4.2 情報探索エージェントの機能

本システムでは、コンピュータ・ネットワーク上から必要な情報を見つけ出すのに、情報探索エージェントというものを使用する。これは、クライアント・ホストで生成されコンピュータ・ネットワーク内の各検索サーバー間を移動して、目的の情報を探し出す。

情報探索エージェントは、各クライアント・ホストで一定の数だけ生成される。そのとき、実際に探索するサーバーの上限と探索結果の上限、探索時間の上限などをパラメータ(遺伝子)として持つ。そして、あるサイクル毎に情報探索エージェントは遺伝的操作により、次世代の情報探索エージェントを生成する。このサイクルを繰り返すことにより、情報探索エージェントは徐々にクライアント・ホストとユーザに適応していく。

そのために、情報探索エージェントは以下に掲げる機能を有する。

- **マイグレーション機能:**  
自律的にコンピュータ・ネットワーク上のサーバー間を移動して、目的の情報を探索する。
- **再生産機能:**  
より優れた次世代の情報探索エージェントを生成する。
- **学習機能:**  
ユーザからのフィードバックや探索した経験によって、より効率的な情報探索を行なう。

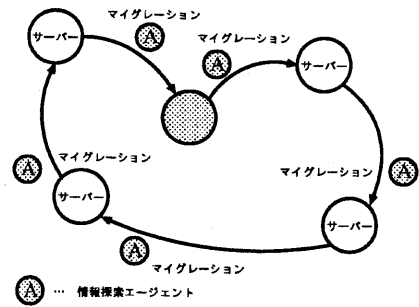


図 6: マイグレーション・モデル

#### 4.2.1 マイグレーション (Migration)

一般に情報カタログを作成するためのロボットと呼ばれるものは、図5で示されるようなモデルで表すことができる。すなわち、実際にはコンピュータ・ネットワーク上を動き回るのではなく、対象とするサーバーに接続して必要とする情報を得るという仕組みである。

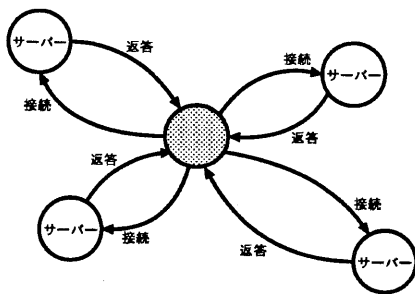


図 5: 非マイグレーション・モデル

これに対し、マイグレーションとは図6で示すようなモデルとして表現できる。これは、エージェントと呼ばれるものが各サーバーを移動し、実際に必要な情報を収集して回る仕組みである。

図5と図6のモデルにおける大きな違いは、図5では実際に仕事をするのはロボットを動かしているホストであり、図6ではエージェントが移動したサーバーであるということである。つまり、マイグレーション・モデルを使用することにより、並列的情報探索と情報探索におけるシステムの負荷分散を行なうことが可能となる。

#### 4.2.2 再生産 (Reproduction)

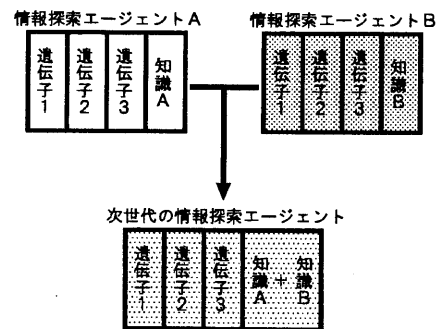


図 7: 次世代情報探索エージェントの生成

情報探索エージェントは、より優れた探索効率を得るために遺伝的アルゴリズムの概念に基づいて進化していく(図7)。すなわち、先に述べた探索パラメータや探索方法をユーザからの評価(ユーザが実際に必要とした情報をどのくらい見つけることができたかなど)や、予め定められている評価規則によって変化させていくことである。実際には、その評価によって全情報探索エージェントをランク付けし、そのランクに応じて次世代の情報探索エージェントを生成するという仕組みである。このことは、情報探索システムが動的にユーザとそのホスト環境に適應していくことを意味する。

#### 4.2.3 学習 (Learning)

情報探索エージェントは、ユーザからのフィードバックによって各ユーザの好みや探索する情報の分野を学習する。また、実際の探索における結果から

情報と情報資源の関係を学習していく(図8)。

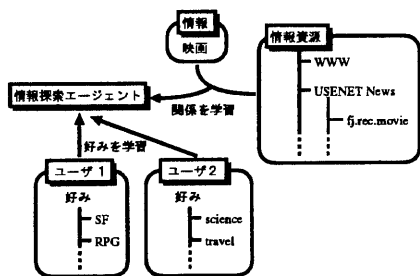


図8: 情報探索エージェントの学習

## 5 既存のシステムとの比較

現在利用されている情報検索システムと本システムを簡単に比較して、その構造や機能の違いを見る。

### 【情報資源の種類】

管理対象となる情報資源の種類は、従来のシステムでは WWW、USENET News などの情報資源のどれか一つをメイン・サーバーで管理する仕組みである。つまり、一つのメイン・サーバーでは WWW と USENET News の2つの情報資源の情報を管理することができない。それに対して、本システムでは WWW、USENET News、anonymous FTP Service が同時に一つのサーバーで行なわれていた場合、全てを統一して情報を管理することができる。

### 【情報の管理方式】

現在のシステムは、各サーバーの情報を一つのメイン・サーバーで管理を行なう一局集中型の管理方式である。本システムは、情報を持つ各サーバーがそれぞれ独自の情報を管理する分散管理方式である。

### 【情報の検索方式】

現在のシステムでは、ユーザから与えられたキーワードをメイン・サーバー上の情報カタログ内から検索する方式か、あるいは見出し語を再帰的にたどっていくナビゲーション方式である。本システムは、ユーザから与えられたキーワードを持った情報探索エージェントが各サーバーを移動し、その上で情報カタログ内から検索する方式である。

## 【特徴】

### ● 長所:

現在の検索システムは、メイン・サーバーによる一局管理のため情報検索が早い。それに対して、本システムでは分散情報管理のためロバスト性に優れている。また、情報資源が検索サーバーであるため最新の検索情報を保つことが可能である。

### ● 短所:

現在のシステムは、一局管理のためロバスト性に劣る。また、検索情報の更新を頻繁に行なえないため、検索情報が古くなる可能性がある。逆に本システムでは、情報探索エージェントが各サーバー間を探索して回るため検索が遅い可能性がある。

## 6 むすび

コンピュータ・ネットワーク上の情報資源は、日々更新され増え続けている。そのため、それらの情報資源からどのようにして必要な情報を見つけ出すかが、重要な問題となってきている。本研究では、この問題に対する一つの解決法を提案し、それに基づくシステムの構成と機能について述べた。また、今回はそのシステムにおける情報探索の中心的部分である情報探索エージェントに焦点を絞り、その機能などについて詳しく述べた。さらに、現在の検索システムと本システムを比較することにより、本システムの構造と機能の違いを述べた。

今後の研究方針としては、大学内の LAN(Local Area Network) 上にプロトタイプ・システムを構築し、評価することである。

## 参考文献

- [1] Robert Armstrong, Dayne Freitag, Thorsten Joachims, and Tom Mitchell, "Web Watcher: A Learning Apprentice for the World Wide Web", School of Computer Science, Carnegie Mellon University, January 20, 1995.
- [2] M. Srinivas and Lalit M. Patnaik, "Genetic Algorithms: A Survey", IEEE, June 1994, pp.17-26.
- [3] Michael Wooldridge and Nicholas R. Jennings, "Intelligent Agents: Theory and Practice", Knowledge Engineering Review, October 1994.