

# エージェントによる分散情報の管理と探索\*

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‡

近年の World-Wide Web (WWW) の普及によるインターネットへの関心の高まりは、目覚ましいものがある。特に World-Wide Web による個人での容易な情報の発信は、瞬く間にインターネット上に膨大なあらゆる情報を無秩序に氾濫させるに至り、ユーザが自分の必要とする情報をどのようにして探し出すかが大きな問題となって来ている。そのため、World-Wide Web 上の情報を対象とした様々な検索システムが盛んに研究・開発されている。しかしながら、いずれのシステムも単一ホストによる集中管理方式を採用しており、現在のような肥大化した情報資源全体を管理するには限界があるように思われる。

そこで本研究では、World-Wide Web などの情報資源を提供するサーバ自身が検索サーバとしての役割をもつ分散情報の自己管理方式を提案する。本方式では、エージェントのマイグレーションによって、分散配置された情報資源を相互に結び付け情報探索に利用する。これにより、インターネット上の情報資源を効率的に管理し、ユーザの必要とする情報を探索することが可能になると考えている。本文では、システムの基本概念と各エージェントの役割、コミュニケーション・プロトコルなどについて報告する。

## 1 はじめに

近年のインターネットの一般的社会への普及、すなわち World-Wide Web の利用者の増加は目覚ましいものがある。今では、インターネット = World-Wide Web という図式が成り立つほどに World-Wide Web の利用者が急増している。特に、個人による自由な情報の発信が可能ということで、多くの利用者によってインターネット上に情報が無秩序に分散・蓄積され続けている。それにより、既存のインターネット上の情報資源が World-Wide Web の影に隠れてしまうほどである。

そのため現在では、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来ている。最近では、膨大な World-Wide Web の情報源から目的の情報を見つけ出すために多種多様な検索システムが研究・開発され続けている。一般に、この問題を解決する方法として、情報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(情報カタログ)

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 World-Wide Web や NetNews などの情報量を見てもわかるように、単一ホストで情報資源全体を把握するのは困難である。このようにインターネット上の情報資源の極度の分散化、大規模化が現在の情報検索における大きな問題となって来ている。

本研究では、従来の単一ホストによる集中型の情報管理方式に対して複数のホストによる分散型の自己情報管理方式を提案する。しかし、そのような分散型自己情報管理方式においては各情報資源間の効率的な検索、つまり相互に情報を結びつける方法が必要がある。そこで、本研究では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を利用する。さらに遺伝的アルゴリズムにもとづく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、分散環境上における情報の効率的な情報資源探索を実現する。本研究の究極的な目標は、World-Wide Web などに限らずインターネット上の全ての情報資源の効率的共有の実現である。

本文は、以下のように構成される。まず最初に、イ

\*An Information Retrieval and Management System on Distributed Computer Networks based on Agent Model

†Kouichi ABE, Toshihiro TAKETA, Kuninobu TANNO

‡Department of Electrical and Information Engineering, Yamagata University

インターネット上における情報資源について述べ、現在の情報資源の管理方式と本研究で提案する分散した情報資源の自己管理方式とそれに応じた情報検索方式について述べる。次に、この方式にもとづいたシステムの概要について詳細に説明する。最後に、今後の研究の方向性について述べる。

## 2 インターネット上の情報資源

### 2.1 情報資源と利用の仕方

インターネット上で利用できる情報資源には様々なものが存在する。その中で現在、非常に盛んに利用されているのが World-Wide Web である。この節では、インターネット上で広く利用されている情報資源とその利用形態について簡単に述べる。

- 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。ブラウザと呼ばれるクライアントを利用して個人毎に自分の興味のある見出しを辿って行くことにより情報を得る。専用の検索サーバを持つ。

- NetNews:

インターネット上の世界規模の電子掲示版システム。様々なニュース・グループがありニュース・リーダと呼ばれるクライアントを利用し、自分の興味を持つニュース・グループの記事を読むことにより情報を得る。専用の検索サーバを持つ。

- anonymous FTP Service:

インターネット上でソフトウェアやドキュメントなどを提供するサービス。UNIX などの OS には標準で ftp という名のクライアントが存在し、それを利用して必要なソフトウェアなどを得る。ただし、事前に目的のソフトウェアなどの保存しているサーバなどを、archie などの専用の検索アプリケーションを用いて検索を行う必要がある場合が多い。

これらの情報資源は、それぞれ独立したサービスとしてインターネット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(ユーザ・インタフェース)を使用しなければならない。

現在は、World-Wide Web 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利用できる。これにより、上記の情報資源への統一したユーザ・インタフェースの提供がなされることになり、初心者にとってはかなり便利な環境になりつつある。実際、大多数のユーザによって様々なブラウザがこれらの情報資源へのインタフェースとして利用されている。

### 2.2 情報管理形式と情報検索方式

現在の情報検索システムのほとんどが、一つのサーバで各情報資源の情報カタログを管理している(図 2.1)。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方法として、図 2.2 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になることである。このとき各サーバでの情報の検索は、専用の情報探索エージェントを用いて実行される。

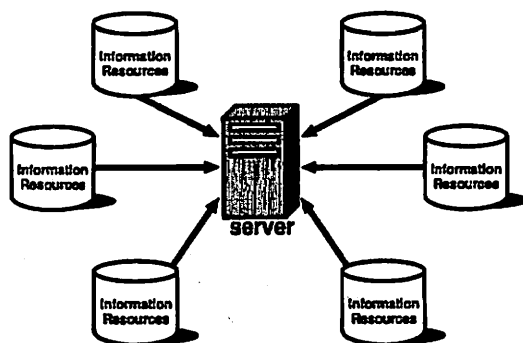


図 2.1: 一般的な検索システムの構造

また、インターネット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、インターネット上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えている。

しかし、現在のような肥大化したインターネット上の情報資源の情報を単一のホストで管理するには限界がある。さらに、実際に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ともあげられる。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。

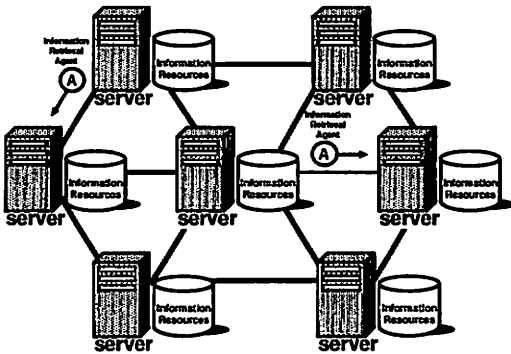


図 2.2: 本研究で提案する検索システムの構造

### 3 分散情報の共有

本システムの基本概念は、分散されている情報がある特定の単一ホストで集中管理するのではなく、その分散されている情報毎に自己管理し、それらを相互に効率的に結び付けることである。

#### 3.1 マルチエージェント・モデル

本研究では、これらの検索サーバ間の相互結合に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(図 3.1)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。本システムでは、ユーザの必要とする情報を探し出すことが最終目的にあたる。

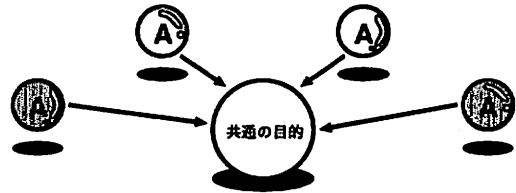


図 3.1: マルチエージェント・モデル

#### 3.2 システムの構造

インターネット上における本システムの構造は、図 3.2に示すモデルとして表すことができる。図 3.2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される。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。これらは大まかに言って、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にユーザとのやり取りを行う。情報探索エージェントは、実際にインターネット上で情報資源を探査し、目的の情報を対象とする情報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

本システムはインターネット上のあらゆる情報資源を効率的に共有することを究極的な目標としている。情報共有の基礎は、いかに効率的に目的の情報を探し出すことができるかである。そのため現段階では、分散情報の自己管理方式による検索サーバに依存しない、動的にホスト環境に適応する情報探索システムの構築および実験を目標としている。実装においては、既存の環境を変更することなく本システムを導入でき、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と情報資源名を入力する

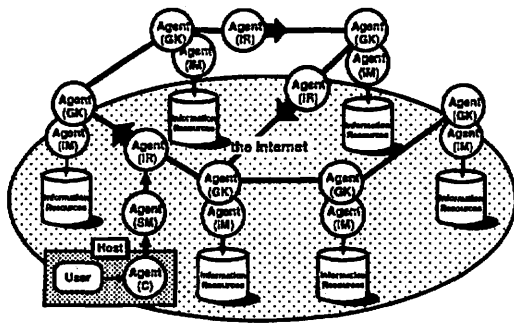


図 3.2: システムの概念図

だけで、目的の情報を入手することができるシステムの実現を目標に研究を進めている。

現在、プロトタイプ・システムとして以下の機能を実装中である。

1. 検索サーバに独立な情報探索機能:  
決まった検索サーバ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を決定し、検索しに行くこと(これを情報の探索と呼ぶ)。
2. 動的な環境適応機能:  
情報探索エージェントがホスト環境とユーザからのリアクションによって、その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:  
World-Wide Web、NetNews などの情報資源を区別することなく、一つのサーバ上で情報カタログを用いることによって統一して管理すること。つまり、これは World-Wide Web 専用の検索サーバとか NetNews 専用の検索サーバとかと言う区別がなくなることを意味する。
4. ブラウジング機能:  
検索結果である情報資源からの情報を、ユーザに分かりやすい形式に変換(アイコン化や URL など)して、ユーザはそれらを選択するだけでその情報を得ることができるようにすること。

## 4 システム・エージェント

本節では、システム上で稼働する各エージェントについて説明する。

### 4.1 ユーザ・インタフェース・エージェント

ユーザ・インタフェース・エージェントは、実際にユーザから情報探索の要求を受け取るエージェントである。そのため、初心者でも操作体系が容易に理解できるインタフェースを備える必要がある。現時点では、World-Wide Web の CGI のスクリプトとして実装することを考えている。本システムでは、このエージェントを“Communicator”と呼ぶ。

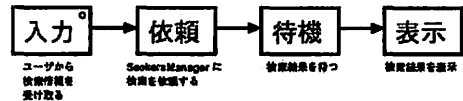


図 4.1: Communicator 状態遷移図

### 4.2 情報探索エージェント管理エージェント

情報探索エージェント管理エージェントは、情報探索エージェントの統括、ユーザ・インタフェース・エージェントからの探索依頼の下請け、情報探索エージェントの遺伝的操作を行うエージェントである。本システムでは、このエージェントを“SeekersManager”と呼ぶ。

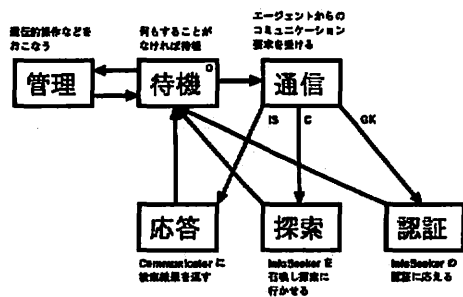


図 4.2: SeekersManager 状態遷移図

各情報探索エージェントは、図 4.3 で示す形式で管理される。

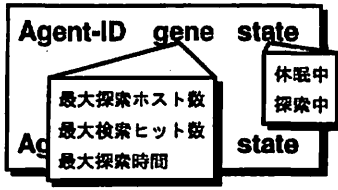


図 4.3: InfoSeeker 管理フォーマット

### 4.3 情報探索エージェント

情報探索エージェントは、実際にインターネット上の各検索サーバ間を移動(マイグレーションと呼ぶ)しながら、目的の情報を探し出すエージェントである。このエージェントは、各クライアント・ホストで一定の数だけ生成され、その時探索するサーバ、検索結果、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、一定サイクル毎に遺伝的アルゴリズムにもとづく遺伝的操作により、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、自分の生存するホスト環境とユーザに適合していく。さらに、検索結果に対するユーザからの評価や探索経験によって、より効率的な情報探索を行うための学習機能も持つ。本システムでは、このエージェントを“InfoSeeker”と呼ぶ。

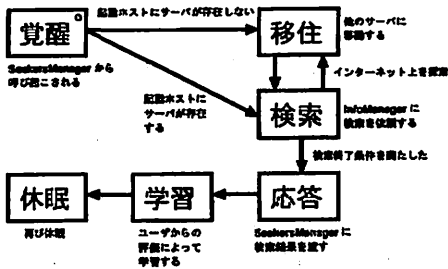


図 4.4: InfoSeeker 状態遷移図

### 4.3.1 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 4.5 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、この遺伝子型にもと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。

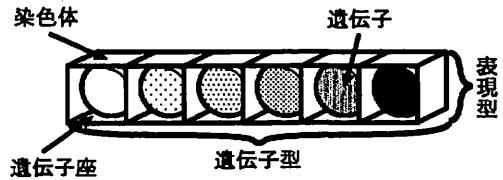


図 4.5: 遺伝的アルゴリズムにおける個体

### 4.3.2 マイグレーション

一般的なサーバ・クライアント・モデルでは、クライアントはサーバのホストに接続して、各種提供されるサービスを受ける。これに対して、マイグレーションとは、クライアントとしてのエージェントが実際にサーバ・ホストに移動し、そこでサービスを受けたり、何らかの行動を起こしたりする。そしてまた、出発したホストに戻って来る仕組みである。

マイグレーション・モデルを使用することにより、エージェントを用いた並列の情報探索と情報探索におけるシステムの負荷分散を行うことが可能となる。つまり、従来のように検索サーバに集中的に負荷がかかることもなく、またエージェントを通じた各情報資源間の情報共有にも大きな役割を果たす。

### 4.3.3 情報探索エージェントの構造

情報探索エージェントは、ヘッダ部とボディ部から構成される。ヘッダ部は、

Searched	検索したホスト名の履歴
Pre-Host	直前に検索したホスト名
Migration-Errors	マイグレーションに失敗した数
Home-Host	探索終了条件時に戻るホスト名
Max-Hosts	探索する検索サーバ数の上限
Max-Results	検索結果の上限
Time-Limit	探索時間の上限
Hosts	現在の探索サーバ数
Results	現在の検索結果数
Times	現在の探索時間
Date	探索を開始した日付
Agent-ID	情報探索エージェント認証コード
User	探索を依頼したユーザ名 (E-Mail address)
Resources	検索する情報資源名
Keywords	検索用キーワード群

で構成される。また、ボディ部は、検索結果を各情報資源に応じた URL (Uniform Resource Locator) で記述したものである。

#### 4.4 情報管理エージェント

情報管理エージェントは、情報資源を提供するサーバにおける情報の統一した管理、情報探索エージェントからの検索依頼の処理を行うエージェントである。また、現在のシステムの実装は、UNIX 上を前提としている。このため、World-Wide Web のサーバを Macintosh や Windows 上で動作させている場合は、情報管理エージェントなどが存在できない。そのため、そのようなマシンの代理的な情報管理を行う機能も備えている。本システムでは、このエージェントを "InfoManager" と呼ぶ。

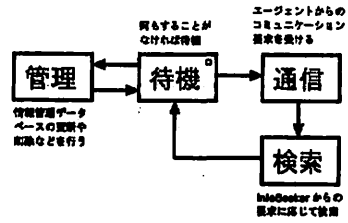


図 4.6: InfoManager 状態遷移図

##### 4.4.1 情報管理方式

各情報資源の情報は、全て統一したフォーマットで記述された情報カタログによって管理される。この情報カタログは、InfoManger によって各情報資源毎に、自動的に作成・更新される。現在、World-Wide Web、anonymous FTP service、NetNews を対象に実験を進めているが、このとき各情報資源毎のカタログ作成時における情報の公開制限を行うための設定ファイルを利用することもできる。情報カタログのフォーマットを、図 4.7 に示す。

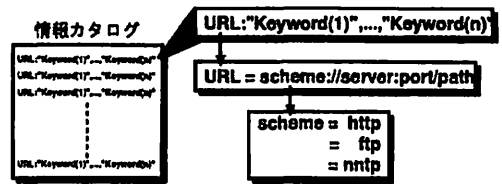


図 4.7: 情報カタログのフォーマット

また、各情報資源の情報カタログのための情報とキーワードは、それぞれ以下の規則に従って自動的に対象となる情報ファイル (HTML 文書、ニュースの記事など) から取り出される。

- World-Wide Web: HTML 文書中のタイトル、ヘッダ、アンカー・タグ・フィールドから取り出す。
- anonymous FTP service: ls-lR などのテキスト・ファイルから取り出す。
- NetNews: ニュース記事のヘッダ・フィールド (Message-

ID, Newsgroups, Subject など) と、本文から取り出す。

#### 4.4.2 情報検索方式

情報検索のパターンは、

- (1) 英語の大文字・小文字に関係なく、その文字列を含む情報を検索する。
- (2) 英語の大文字・小文字を区別して、その文字列を含む情報を検索する。
- (3) 正規表現を用いて情報を検索する。
- (4) 与えられた文字列に正確に適合する情報を検索する。

以上の4パターンである。また、検索文字列の結合規則は、

- (1) and 結合: キーワード群すべてを必ず含む情報を検索する。
- (2) or 結合: キーワード群の少なくとも一つのキーワードを含む情報を検索する。
- (3) not 結合: キーワードに完全にマッチする情報を除外する。
- (4) exact 結合: キーワードに完全にマッチする情報を検索する。

である。情報管理エージェントは、上述の結合規則にもとづいた検索文字列を情報探索エージェントから受け取り、検索を行う。

#### 4.5 サーバ・アクセス監視エージェント

サーバ・アクセス監視エージェントは、情報探索エージェントに次の検索サーバ名を伝えたり、検索候補のサーバ名を管理するエージェントである。本システムでは、このエージェントを“GateKeeper”と呼ぶ。

##### 4.5.1 検索ホスト・テーブル

情報探索エージェントのマイグレーション用に使用する検索ホスト・テーブルのフォーマットは、図4.9に示す構造を持つテキスト・ファイルである。

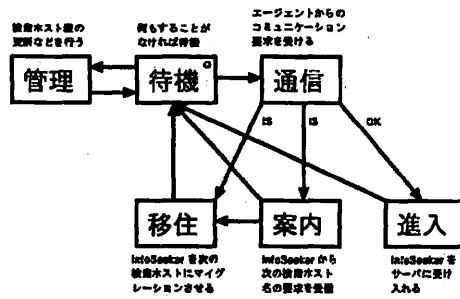


図 4.8: GateKeeper 状態遷移図

検索サーバ名	情報資源名	検索頻度
www.etcn.yz.yamagata-u.ac.jp	WWW	0
ftp.yz.yamagata-u.ac.jp	FTP	0
news.yz.yamagata-u.ac.jp	NEWS	0
⋮		

図 4.9: 検索ホスト・テーブルのフォーマット

## 5 むすび

インターネット上の情報資源は、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、この問題に対する解決法として分散情報の共有を前提とした情報探索システムを提案し、その構成と機能について述べた。さらに、システムにおける各エージェントの機能などについて詳細に説明した。

今後の研究方針は、大学内の LAN(Local Area Network) 上でのプロトタイプ・システムによる実験結果を収集し、評価することである。その後、WAN(Wide Area Network) 上での実験に拡張していく予定である。

#### 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.

## 付録 A エージェント・コミュニケーション・プロトコル

付録として、エージェント間のコミュニケーション・プロトコルについて説明する。説明の便宜上、ユーザ・インタフェース・エージェントを "[C]"、情報探索エージェント管理エージェントを "[SM]"、情報探索エージェントを "[IS]"、情報管理エージェントを "[IM]"、サーバ・アクセス監視エージェントを "[GK]" と示す。また、太文字のアルファベットはコミュニケーション・コマンドである。

### 付録 A.1 探索時におけるプロトコル

- (1) [C] → [SM]:  
**RETRIEVAL**(User)(Resources)(Keywords)
- (2) [SM] → [IS]:  
InfoSeeker を呼び出す
- (3) [IS] → [C]:  
検索結果ファイル名を伝える
- (4) [C]:  
検索結果ファイルが作成されるまで待機

### 付録 A.2 認証時におけるプロトコル

- (1) [GK] → [SM]:  
**AUTHENTICATION**(Agent-ID)
- (2) [SM] → [GK]:  
認証が成功 (OK) あるいは失敗 (NG) かを伝える

### 付録 A.3 応答時におけるプロトコル

- (1) [IS] → [SM]:  
**RESPONSE**(Agent-ID)
- (2) [SM] → [IS]:  
情報探索エージェントを受け入れる

- (3) [IS] → [C]:  
検索結果ファイルを作成
- (4) [SM] → [IS]:  
InfoSeeker を休眠させる

### 付録 A.4 案内時におけるプロトコル

- (1) [IS] → [GK]:  
**GUIDANCE**(Home-Host)(Agent-ID)  
情報資源名 (Resources) を送信  
検索ホスト履歴 (Searched) を送信  
**OVER**
- (2) [GK] → [IS]:  
マイグレーションの準備完了 (OK) あるいは前の検索ホストに戻る (NG) かを伝える

### 付録 A.5 移住時におけるプロトコル

- (1) [IS] → [GK]:  
**MIGRATION**(Home-Host)(Agent-ID)(Next-Host)  
InfoSeeker 自身を送信する  
**OVER**

### 付録 A.6 進入時におけるプロトコル

- (1) [GK] → [GK]:  
**ENTRANCE**(Agent-ID)  
InfoSeeker を受け入れ、ホスト内に解放する  
**OVER**

### 付録 A.7 検索時におけるプロトコル

- (1) [IS] → [IM]:  
**SEARCH**(Agent-ID)
- (2) [IM] → [IS]:  
**OK**
- (3) [IS] → [IM]:  
情報資源名 (Resources) を送信  
キーワード (Keywords) を送信
- (4) [IM] → [IS]:  
検索結果を送信  
**OVER**