

# エージェント理論に基づく インターネット検索エンジンに関する研究

能登研究室

中野良太 (46062)

## 1 はじめに

現在, WWW(World Wide Web) 上には, 何億という URL(Uniform Resource Locator) が公開されており, 情報検索のための検索エンジンが不可欠となっている. しかしながら, 検索エンジンを利用して, 知りたい情報や有用な情報を得ることは非常に困難である. 本稿では, 巡回エージェントを提案し, よりよい検索エンジンの構築について報告する.

## 2 検索エンジン

WWW 上で知りたい情報を得るためには検索エンジンを利用するのが一般的である. 現在, 主に検索エンジンには2種類あり, 本節では, その特徴を述べる.

- ディレクトリサーチ  
Yahoo! などに代表される検索エンジンで, 情報を発信する側(ホームページの作成者)がその URL と内容を検索エンジンに登録し, 検索エンジンの管理者側がデータベースを作成し, そこからキーワードによって該当する情報を検索するものである.
- ロボットサーチ  
goo などに代表されるロボット型の検索エンジンは, ロボットと呼ばれるソフトウェアが WWW サーバにアクセスし, WWW サーバに含まれている HTML(Hyper Text Markup Language) ファイルからキーワードを抽出する仕組みである.

ロボットサーチにおける問題点を以下に挙げる.

1. 表示された URL が存在しない.
2. 重複 URL の表示.
3. 時間の経過による情報の劣化.
4. ミラーサイトなどの情報の重複.

本研究ではロボットサーチの延長線上で, 巡回エージェントを提案し上記のような問題点を解決できるような検索エンジンを構築する.

## 3 エージェント

一般的なエージェントの特徴として以下の4つが挙げられる.

- 自律性: エージェントは人間や他システムからの直接的な干渉を受けることなく行動する. このため, 自己の行動や状況を制御する機能をもつ.
- 社会性: エージェントは, 交互に理解できる言語やプロトコルを介して他のエージェントや人間と相互作用を行う.
- 反応性: エージェントは, 自分のおかれた外部環境を認識し, その変化に適切に応答する.
- 自発性: エージェントは, 単に外部環境に応じて行動するだけでなく, ある目標を目指して, 主体的に行動する.

エージェント理論に基づき以下のような特徴をもつ, 巡回エージェントを提案する.

- 代行性: 人間のかわりにブラウザを用いることなく HTML ファイルを処理できる機能.
- 自発性: エージェント自ら, ユーザの要求を満たそうとする機能.
- 適応性: 刻々と変化する WWW の中でもっとも効率よくアクセス可能な情報源を動的に決定できる機能.
- 協調性: エージェント同士で協調して情報の重複やサーバの負荷状況など伝え合う機能.

## 4 システムの設計

本研究で提案する検索エンジンの流れを以下に示す.

1. WWW 上をエージェントが巡回.
2. エージェントが HTML ファイルなどを取得.
3. エージェント同士の協調により情報を交換.
4. HTML ファイルなどをファイルシステム(データベース)に保存.
5. 各エージェントが取得してきた重複 URL の除去.
6. 保存されたデータの有無を定期的に確認.
7. ユーザから要求された情報の表示.

本システムの流れを図式化すると図1の様になる.

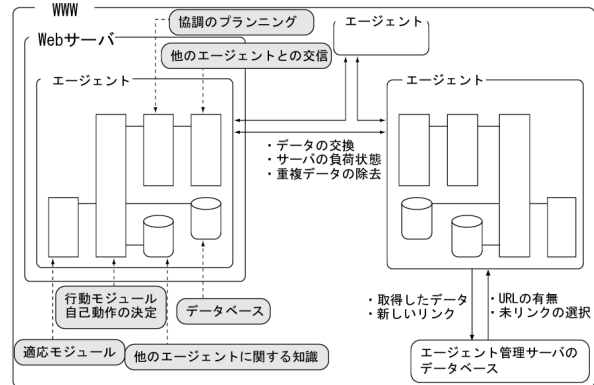


図1: システムの流れ

図1の各部分は以下のとおりである.

- ◇ 行動モジュール: 環境モジュールに応じて自己をプランニングし, 実行する. HTML ファイルなど取得し, データベースに格納する. WWW をどのように巡回するかを決定する.
- ◇ 適応モジュール: 上記のモジュールの背景で動作し, 外部環境に応じて自己の動作の最適化をはかるモジュール.
- ◇ 他のエージェントとの通信: 実際, 他のエージェントと通信を行う部分.
- ◇ 協調のプランニング: 他のエージェントと協調するためのプランニングを行う.
- ◇ 他のエージェントに関する知識: 上記の協調のプランニングを効率よく行うための部分.
- ◇ データベース: HTML ファイルなどのデータを格納する部分. 未リンクや URL の有無など情報を保存しておく部分.

## 5 おわりに

本研究では, インターネット検索エンジンの向上のために巡回エージェントを提案し, 大まかな流れやプロトタイプシステムを構築した. 巡回エージェントにより URL や内容の重複を防ぐことができ, ユーザによりよい情報を発信することができる.

今後は, WWW 上に巡回エージェントを走らせて実際にシミュレートする予定である.