

WWW 情報統合からエージェント統合へ

北村泰彦

大阪市立大学大学院工学研究科

kitamura@info.eng.osaka-cu.ac.jp

1. はじめに

WWW はインターネット情報処理のベースとなるプラットフォームである。WWW 上には膨大な量の情報が蓄積され、さまざまなサービスが提供されている。WWW 情報統合はコンテンツレベルで WWW 上の情報を統合する試みである。ここでは WWW 情報統合へのアプローチをプログラミング、データベース、エージェントの観点から述べる。

2. WWW 情報統合へのアプローチ

2.1. プログラミングアプローチ

WWW 情報統合における最初の問題は Web ページからの必要な情報の抽出である。この困難さは現在の Web ページ記述言語である HTML によるものである。すなわち HTML は Web ページをブラウザで表示する場合の視覚的構造を記述することを目的としており、そのページの意味的な構造を記述することを目的としていない。この問題を解決するために Web ページから HTML のタグの位置関係を基に、必要な情報を抽出するためのラッパー(wrapper)と呼ばれるソフトウェアモジュールを開発することにより対処できる。このためのプログラミング言語としては我々が開発した MetaCommander などがある。¹

しかしこのような情報抽出は XML をベースにした意味タグの導入により格段に容易なものとなる。またさらに情報抽出ラッパーの統合を行うための標準化やプラットフォーム開発として UDDI(Universal Description, Discovery, and Integration)^aや Web Service²などの試みがある。

2.2. データベースアプローチ

プログラミングアプローチでは手早く WWW 情報統合システムが実現できる反面、拡張性が乏しく、大規模な情報統合システムを実現することは困難である。

これに対してデータベースアプローチは WWW システムを一つの大規模な分散情報ベースとして扱おうとしている。この実現のためにはそれぞれの Web 情報源の持つ異種性を吸収する標準化手法が重要になる。Stanford 大学の TSIMMIS プロジェクト³では OEM という共通データ表現を採用することによりデータ表現の標準化を行っている。OEM ではデータを<名前、型、値>の三つ組みで表現している。このようにデータを値だけで表現するのではなく、その意味を表す名前や型を埋め込むことにより、異種情報源の統合が可能になる。これは現在の XML や RDF にも共通する概念である。情報統合はメディエータと呼ばれるモジュールにおいて行われる。メディエータは複数のラッパーから得られる情報を組み合わせる特定の領域に特化したクエリーに答えられるようにしたものであるが、この作業はハードコーディングとなる。この部分を自動化するためにはそれぞれの情報源からのデータの意味を統一的に扱える枠組みが必要になる。南カリフォルニア大学の Ariadne プロジェクト⁴では情報統合のために SIMS オントロジを採用している。オントロジはある領域の概念体系であり、それをを用いることにより情報統合の自動化が容易になる。このようなオントロジを WWW 上で容易に利用可能にしようとする試みが SemanticWeb⁵であるといえる。

2.3. エージェントアプローチ

データベースアプローチは WWW を世界規模の分散情報ベースとして扱えるようにする試みであった。しかしながら情報は蓄積するだけでは不十分であり、それをどのように使いこなすかが重要になる。特に情報統合に対する利用者の要求は多様であり、これらの要求に応じてゆく必要がある。まず、利用者の要求や制約を満たす情報源は一つとは限らず、複数の情報源から適切なものを選択する必要がある。それを選択する基準としては、アクセスに対する通信遅延や有料サイ

^a <http://www.uddi.org/>

ト等における費用を考慮する必要がある。また、株価情報やオークションなど情報更新が頻繁に行われる情報源が存在する。このような情報源に対しては更新頻度に応じて連続的にアクセスを行う必要がある。さらに、Web 情報源の情報は必ずしも正しいとは限らないので、複数の情報源の情報を重ね合わせることで、情報の精度を高めることも必要になるであろう。

以上のような要求に柔軟に応えるためには情報統合プロセスが十分に知的であることを要求する。情報エージェントは利用者の多様な要求に応えるべく知的に情報検索や収集を行うソフトウェアモジュールである。エージェントは、自発性(proactiveness)、適応性(adaptability)、社会性(sociability)といった特徴を持っている。

- 自発性とはエージェント自身が目標をもち、その達成に向けて自発的に行動することである。例えば、マサチューセッツ大学で開発された BIG⁶は情報収集に課せられた時間、コスト、精度などの制約を考慮して、情報収集のプランニングとスケジューリングを行った上で、その実行を行う情報エージェントの一つである。
- 適応性とはエージェントが環境の変化に対して適応的に行動することである。例えば、我々は航空便の乗り継ぎ問題を例題に、空席情報提供サイトの情報更新に応じて効果的に情報収集を行う動的情報収集メディア⁷を開発した。
- 社会性とはエージェントが他のエージェントと協力して行動することである。例えば、CMU では複数の情報エージェント、タスクエージェント、インタフェースエージェントが協力して情報統合を行う RETSINA⁸プラットフォームを開発した。これは一つのマルチエージェントシステムといえるが、その標準化作業は FIPA^bにおいて行われている。

3. エージェント統合へ向けて

これまで議論してきた WWW 情報統合技術は主に Web 情報源に起因する問題へ対処と見なすことがで

きる。XML や Semantic Web の浸透はこれらの問題の解決をさらに容易なものとするであろう。その一方で、我々は利用者側の課題にも目を向ける必要がある。WWW 利用者の底辺が広がるにつれて、情報統合システムの利用者インタフェースもより容易で自然なものへと進化させてゆく必要がある。従来の情報統合システムでは統合プロセスは利用者からはブラックボックスとして隠蔽されていたが、好みの情報源を選択し、その統合プロセスを自由にカスタマイズできるような仕組みが今後必要になるとと思われる。そのような試みの一つとして我々はマルチキャラクタエージェントインタフェース(MCI)⁹を開発している。MCI では各情報源はキャラクタインタフェースをもち、利用者は複数のキャラクタと自然言語を用いて相互作用しながら協調的な情報統合を実現できる。MCI はエージェント統合による情報統合システムのさきがけといえる。

参考文献

- ¹北村泰彦, 他. スクリプトに基づく WWW 情報統合支援システムとゲノムデータベースへの応用. 信学論, J81-D-I(5):451-459 (1998)
- ²McIlraith, S.A., et al.: Semantic Web Services. IEEE Intelligent Systems, 16(2):46-53 (2001)
- ³Chawathe, S., et al.: The TSIMMIS Project: Integration of Heterogeneous Information Sources. Proceedings of IPSJ Conference, 7-18 (1994)
- ⁴Knoblock, C.A., et al.: The Ariadne Approach to Web-based Information Integration. International Journal of Cooperative Information Systems, 10(1&2):145-169 (2001)
- ⁵Hendler, J.: Agents and the Semantic Web. IEEE Intelligent Systems, 16(2): 30-37 (2001)
- ⁶Lesser, V., et al.: BIG: An agent for resource-bounded information gathering and decision making. Artificial Intelligence, 118(1-2): 197-244 (2000)
- ⁷北村泰彦, 他. 動的情報メディアのための知的情報収集手法. 信学論, J84-D-I(8):1256-1265 (2001)
- ⁸Sycara, K., et al.: Distributed Intelligent Agents. IEEE Expert, 11(6):36-45 (1996)
- ⁹Kitamura, et al.: Interactive Integration of Information Agents on the Web. Cooperative Information Agents V, LNAI 2182, Springer. 1-13 (2001)

^b <http://www.fipa.org/>