

MetaCommander2: エージェント指向 Web コンテンツ統合環境

MetaCommander2: Agent Oriented Web Contents Integration System

回り道康博, 北村泰彦, 辰巳昭治

Yasuhiro Mawarimichi, Yasuhiko Kitamura, and Shoji Tatsumi

大阪市立大学工学部情報工学科

Department of Information and Communication Engineering
Faculty of Engineering, Osaka City University

In the WWW, web sites are distributed and HTML documents consist without semantic tag like XML, so automated information gathering and integration is difficult. We propose a web contents integration system called "MetaCommander2". In this system, agents gather HTML documents, extract the required information, and integrate it by collaborating with each other. Moreover, the information is integrated into a portal page. This system has three characteristics: (1) information gathering, extracting and integrating by utilizing multiple agents, (2) information selection according to user's interest, and (3) robustness to changes of HTML documents.

1 はじめに

インターネットは無限に近い情報から成り、その情報源は分散して存在している。この分散した情報源から、利用者が興味ある情報を絞り込んでまとめることを情報統合と呼ぶ。例えば、日本の航空会社の空席情報は JAL, ANA, JAS の 3 社の Web サイトに分散して存在しているが、これらを 1 つにまとめることが情報統合である。

同様に、利用者が興味あるいくつかの領域の情報を 1 つの Web ページにまとめることを Web コンテンツ統合と呼ぶ。本稿ではこの Web コンテンツ統合を支援するシステムとして MetaCommander2 を提案する。

Web コンテンツ統合への過程においては、大きく分けて「情報抽出」「情報統合」「提供ページの生成」という 3 つの副処理が必要となる。まず、WWW の中から利用者が興味ある Web ページを収集し、そこから情報の抽出を行う。Web ページの内容はすべてが有用であるわけではなく、利用者にとって必要でない部分や、冗長な情報も含まれている。そういった不要な部分を省いて必要な情報だけを残すのが、情報抽出である。Web ページには、意味付けのされていない HTML タグが使

〒558-8585 大阪市住吉区杉本3-3-138
{ym,kitamura,tatsumi}@kdel.info.eng.osaka-cu.ac.jp
<http://www.kdel.info.eng.osaka-cu.ac.jp/>

用されており、単純に機械的な方法で情報抽出を行うのには困難が伴う。次に、抽出された情報を元にして情報統合を行う。同じ領域に属しているが内容が異なる情報（例：飛行機会社 3 社の空席情報）は、複数の Web ページに分散して存在しているが、これは 1 つにまとめられていることが望ましい。そして最後に、得られた情報の中から、利用者の興味と照らし合わせて重要度の高いものを絞り込み、1 つの Web ページに収める。このようにして Web コンテンツ統合が行われる。

また、Web ページでは改変や削除が頻繁に行われるため、この点も考慮に入れる必要がある。

MetaCommander2 の特徴は、以下の 3 点である。

1. エージェントによる情報収集・抽出・統合
2. 利用者の興味に合わせた提供情報の選択
3. Web ページの変化への対処

本稿では、2 章で MetaCommander2 の構成を示し、3 章で考察を行い、4 章でまとめとする。

2 MetaCommander2

2.1 Web コンテンツ統合

MetaCommander2 では、Web ページの収集から情報統合に至る過程を機能毎にモジュール化し、それぞれのモジュールを再利用可能にするという目的で、

RETSINA[1]で提唱された、3種のエージェントによるシステムアーキテクチャモデルを使用している。

- 情報エージェント (Information Agent)
- タスクエージェント (Task Agent)
- インタフェースエージェント (Interface Agent)

2.1.1 情報エージェント

情報エージェントは、Webコンテンツの収集と情報抽出を担当する。情報エージェントは1つないし複数のWebページを収集し、そこから情報を抽出して、1つのオブジェクトを出力する。この出力されるオブジェクトをWebオブジェクト (WO) と呼ぶ。WOは、再利用の観点からXMLにより表現される。情報抽出にはMetaCommander[2]を利用する。MetaCommanderは、Webソースのタグ構造を解析し、指定された位置にある情報を抽出して再構成するためのプログラムであり、スクリプト言語によって動作を記述する。情報抽出により、タスクエージェントによる情報の再利用が容易になる。

2.1.2 タスクエージェント

タスクエージェントは、情報エージェントが抽出した情報の統合を担当する。タスクエージェントは、利用者から与えられた要求を、複数の情報エージェントとの協調によって遂行する。最も簡単な情報統合は、複数の情報エージェントからの結果を単純に繋ぎ合わせることである。さらにそれより複雑な協調として、以下の2種類が挙げられる。

OR型協調 (分岐) ある情報エージェントが返した結果によって、以降の動作を分岐させる (例: 休日タスクエージェントは、天気情報エージェントの返した結果が「晴れ」であれば行楽情報エージェントを、「雨」であれば映画情報エージェントを使って、その情報を利用者に提供する)

AND型協調 (制約充足) 複数の情報エージェントが返した結果から、要求されている条件を満たす解を求める (例: 旅行プランニング・タスクエージェントは、飛行機情報エージェント、鉄道情報エージェント、ホテル情報エージェントの情報を集めて旅行のプランニングを実行する)

AND型協調は、必要に応じて情報エージェントの再実行が行われる。

OR型協調やAND型協調よりさらに複雑な要求は、この2つの組み合わせにより表現することができる。

2.1.3 インタフェースエージェント

インタフェースエージェントは、WOの統合を行う。ブラウザを用いた単純な情報収集では、同時に複数の情報を閲覧したい場合、利用者が手動でウィンドウをリサイズするなどして、その状態を整えなければならなかつ

た。インタフェースエージェントは、この作業を自動的に、利用者の興味に合わせて行う。

インタフェースエージェントは、WOの情報を利用者に提供する。MetaCommander2では、利用者の興味に合わせて選択された情報を収めたポータルページを提供する。ポータルクリエイターというインタフェースエージェントを用意している。ポータルクリエイターは参照頻度や利用者自身の格付けから利用者の興味を常に学習しており、それに応じてWOを一枚の画面上に配置する。この時、利用者の関心が高いWOの表示領域は大きく取り、逆に関心の低いWOの表示領域は狭くする (あるいは表示しない)。

例えば、普段から野球情報を熱心に見ている利用者のポータルページでは野球情報が大きく表示されるようになる。一方ほとんど関心を示さない利用者のポータルページでは、野球情報の表示領域は日を追うごとに狭くなっていき、やがては表示されなくなる。

2.1.4 エージェント・サーバ

エージェント・サーバは、情報エージェントやタスクエージェントのオリジナルを保存しておくためのダウンロード・サーバである。利用者は、エージェント・サーバから自分の興味に合ったエージェントをダウンロードしてクライアント・プログラムに組み込み、利用する。

さらにエージェント・サーバは、Webページの変化への対処も行う (後述) 。

2.2 実行例

情報エージェントとタスクエージェントは、エージェント・サーバからクライアント・プログラムにダウンロードされ、エージェント・マネージャによって管理される (図1)。この2つのエージェントは利用者の設定に応じて定期的に行われる (例: ニュース情報エージェントは3時間毎、休日タスクエージェントは毎週土曜日午後8時、など) 。

情報エージェントはWWWからWebページを収集し、情報抽出を行ってWOを出力する。またタスクエージェントは、利用者の要求を満たす結果を得るために情報エージェントと協調し、同様にWOを出力する。

WOはポータルクリエイター (インタフェースエージェント) の元に集められ、利用者の興味に応じて、最終結果であるポータルページが出力される (図2) 。

2.3 Webページの変化への対処

現在のWWWでは、絶えずどこかのWebページで削除や改変が行われている。エージェントにとって、参照していたWebページを正常に取得できないことは致命的であり、もしそういった事態に陥った時には、速やかにエージェント自身の内容が更新される必要がある。MetaCommander2では、エージェント・サーバとクライアント・プログラムの両方で、Webページの変化への対処を行っている。

エージェント・サーバは、情報エージェントが参照しているWebリソースを定期的に監視している。エー

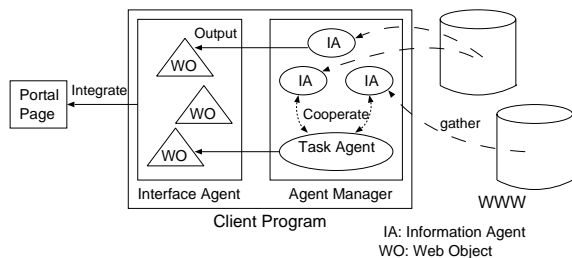


図 1: Web-Contents Integration

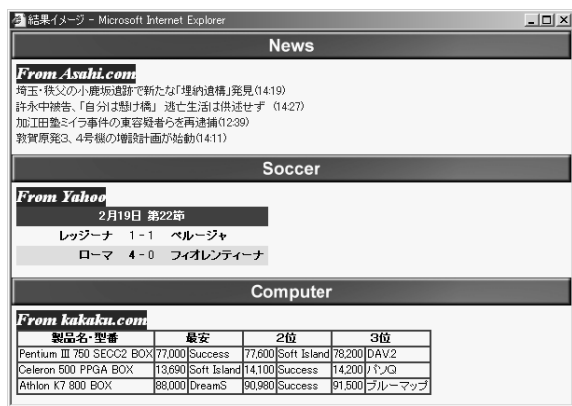


図 2: Portal-Page Image

エージェント・サーバは、Web ページの削除、タグ構造の変更、あるいは Web ページが置かれているサーバのダウンなど何らかの障害が発生した場合、その Web ページを参照しているエージェントをダウンロード禁止とする。もし障害が時間の経過によって復旧しないものである（削除やタグ構造の変更）ならば、エージェントの開発者へ電子メールによる通知を行い、エージェントの修正を要請する。障害が復旧するか、エージェントの開発者によって修正されたエージェントのアップロードが行われれば、ダウンロードの禁止を解除する。

またクライアント・プログラム側での対処は以下の通りである。情報エージェントが Web ページから情報を正常に取得できなかった場合には、エージェント・マネージャがそのエージェントを凍結状態へ移行させる。一方でエージェント・サーバの方へ定期的にお問い合わせを行い、障害が復旧していないか、あるいは修正されたエージェントがアップロードされていないかを確認し、正常にエージェントが利用できる状態になれば凍結解除を行う。

こういった機能により、エージェント・サーバに登録されているエージェントは、常に信頼性が保証される。このため、利用者は Web ページの変更を気にすることなく、エージェントを使用することができることになる。

なお、エージェントの開発は誰でも行うことができ、開発したエージェントをエージェント・サーバへアップロードすることで、一般の利用者が使用できるように

なる。

3 考察

MetaCommander2 と同様のシステムに RETSINA がある。本章ではこの RETSINA と MetaCommander2 の差異を示す。

RETSINA の情報エージェントは通常の Web ページと同じくインターネット上に存在しており、タスクエージェントがその情報エージェントを利用する。これに対し、MetaCommander2 の情報エージェントはオリジナルのものがエージェント・サーバにあり、利用者はそれを手元にダウンロードして、以後はそれを利用することになる。いずれも利用者間での情報エージェントの共有を目的としているが、MetaCommander2 ではシステムの運用に必要なモジュール（エージェント）をすべてクライアント・プログラム上に集めているのに対し、RETSINA ではインターネット上にも分散している。

提供情報の選択については、[1] の中では特に触られていなかった。従ってインタフェースエージェントが、利用者の興味に合わせて提供する情報を変更するというのは、MetaCommander2 で付加された機能である。

Web ページの変化への対応については、RETSINA では「ある情報エージェントが使用できなくなった場合には、同様の機能を持つ別の情報エージェントを探して利用する」ことで対処している。しかしこの方法では各情報エージェントそれ自体の信頼性を保証することにはならず、エージェントの開発者への通知やダウンロードの禁止により信頼性を保証する機能を備えた MetaCommander2 の方に分があると考えられる。

4 まとめ

エージェント指向の Web コンテンツ統合環境として MetaCommander2 を提案し、RETSINA への優位性について述べた。

MetaCommander2 は、複雑な手続きを踏むことなく、興味ある情報を収集することができ、その利便性は高い。

今後の課題としては、Java による MetaCommander2 の実装、またその有用性を示すために既存の WWW 上のシステムとの比較、あるいは一般に公開しての実験検証が挙げられる。

参考文献

- [1] Katia Sycara. "In-Context Information Management through Adaptive Collaboration of Intelligent Agents," Matthias Klusch (Ed.), Intelligent Information Agents, pp.78-99, Springer Verlag, 1999.
- [2] 北村泰彦, 野崎哲也, 辰巳昭治. "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", 電子情報通信学会論文誌, J81-D-I(5):451-459, 1998.