

フォーム自動記入のための Web サービス変換 フレームワークの開発

Design and Implementation of Web Service Transformation Framework
for Filling Automatically in a Form

藤原 克哉[†] 中所 武司^{††} 玉本 英夫[†]

Katsuya FUJIWARA Takeshi CHUSHO Hideo TAMAMOTO

[†] 秋田大学工学資源学部情報工学科

Dept. of Computer Science and Engineering, Akita University

^{††} 明治大学理工学部情報科学科

Dept. of Computer Science, Meiji University

fujiiwara@ie.akita-u.ac.jp

我々は、全ての日常的な仕事はコンピュータが代行すべきであるという観点から、エンドユーザが自分のエージェントを自ら作成し利用するための Web アプリケーションフレームワークの研究を行ってきた。本稿では、Web フォームへの記入を支援する自動記入エージェントの実現方式について述べる。これまでの自動記入方式では、記入する内容をあらかじめ利用者が登録できる範囲に記入対象が限られていた。そこで、記入内容を外部 Web サービスから取得し自動記入する方式を開発した。まずはじめに、異なるインタフェースを持つ Web サービス群を透過的に扱うために、共通インタフェースへ変換する Web サービス変換フレームワークを開発した。さらに、サービスの網羅性と可用性を高めるために、現在普及している Web アプリケーションを Web サービスにラッピングするフレームワークを開発した。最後にこの方式を用いて、従来方式では自動記入できなかった書籍情報が自動記入できることを確認した。

1 はじめに

近年、WWW を利用したオンラインショッピングや銀行・証券取引、旅行予約等のシステムがインターネットに広く実用化されている。また、最近では行政サービスの窓口を電子化する電子政府の実現が注目されている。これらの既存の WWW システムにおいて、利用者は、氏名や住所、メールアドレス等の同じような項目の入力を求められることが多い。このような電子フォームへの記入作業を支援する自動記入機能がいくつか実現されているが、実際に自動記入できるフォームが少なかったり、記入を間違えるなどの問題がある。自動記入機能は、適用範囲の広く、記入精度の高い方式が求められる。そこで、我々はこれまでに、記入精度を向上させるメタデータ形式の記入ルール定義言語と適用範囲を広くする記入ルールの自動抽出方式を確立した [1]。本稿では、記入内容を外部 Web サービスに問い合わせる自動抽出方式により、記入対象があらかじめ利用者が記入内容を登録できる範囲に限られていた問題を解決する。記入対象フォームの種類を増やすことで、自動記入できるフォーム数を増やすことができる。

2 フォーム自動記入技術

フォーム自動記入は、フォームにある氏名や住所、メールアドレスなどの決まりきった内容の記入を自動化することで、利用者のフォーム記入の手間を軽減することを目的とする。また、利用者による人為的な記入ミスを防ぐメリットがある。自動記入機能は、電子フォームと、「どの入力項目にどの内容を入力するのか」という記入ルール、「その内容の具体的な値」である記入内容の 3 つの入力を元に、フォームの入力項目に自動記入し、記入済みフォームを出力する。フォーム記入の自動化技術は以下の 3 段階に分類できる。

- (1) 記入の自動化
- (2) 記入ルール抽出の自動化
- (3) 記入内容抽出の自動化

(1) の段階では、あらかじめ用意された記入ルールと記入内容を用いて自動記入を行う。従来の自動記入方式は (1) の段階に分類されるが、適用範囲や記入精度に問題があった。我々は、まず (1) について、明示的に意味情報を定義する Semantic Web 技術 [2] を応用し、フォームのメタデータとして記入ルール

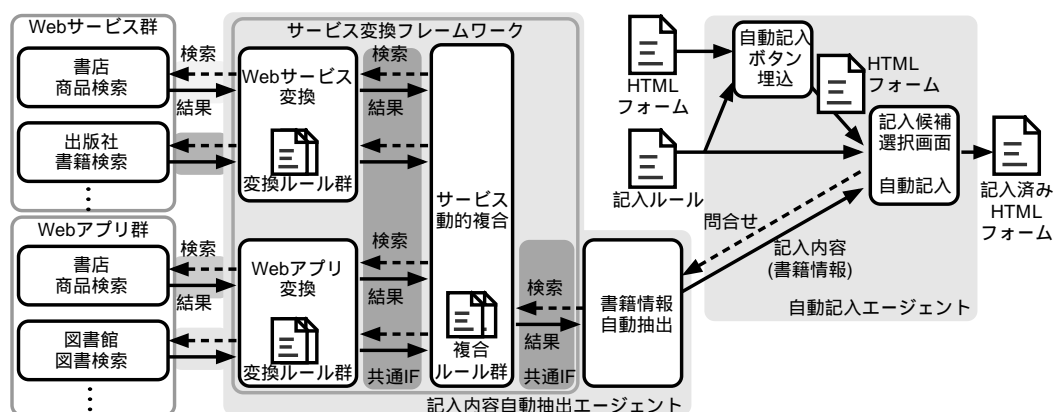


図 1: 応用システムのアーキテクチャ

を定義するための RDF ベースの記入ルール定義言語と定義ツールを開発した。フォーム毎に記入ルールを明示的に定義することで、記入精度を向上させた。しかし、(1) の記入自動化の段階では 2 つの問題があった。1 つ目は、記入ルール・記入内容を手動で定義・登録しておく必要があり、エンドユーザの負担が大きいためである。2 つ目は、記入できる内容もあらかじめ登録できる個人情報記入や、郵便番号からの住所記入などの範囲に限定されることである。記入できるフォーム数を増やすためにはこれらの問題を解決する必要がある。そこで、これまでに 1 つ目のルール定義の負担を解消するために、(2) を実現するマルチエージェント技術を応用した記入ルールの自動抽出方式 [1] を開発した。本研究では、さらに (3) の記入内容抽出を自動化する方式を確立することで 2 つ目の問題を解決する。

3 記入内容自動抽出による自動記入方式

3.1 実現方式の概要

本方式による自動記入の応用例として研究室の図書管理 Web システムの図書登録フォームに、書籍情報を自動記入するシナリオを設定した。これまでは、利用者が著者名や出版者名を自ら書き写す必要があった。そこで以下のような自動記入方式により、利用者の手動記入の手間を解消する。まず利用者は、書籍の ISBN 番号かタイトル等の一部を入力し自動記入ボタンを押す。次に自動抽出機能はその内容に該当する書籍情報を書店がホームページ上に提供している書籍検索機能に問い合わせ、結果を記入候補一覧として表示する。利用者が一致する内容を選択し

確定ボタンを押すと自動記入される。

このような自動記入機能を実現するために、本システムは図 1 に示すように 2 種類のエージェントからなる構成とした。自動記入エージェントは、フォーム、記入ルール、記入内容の 3 つの入力から、記入済みフォームを生成する。まず、自動記入ボタン埋込機能が、記入ルール定義に基づいて図書登録フォームに自動記入機能の呼出しボタンを埋め込む。利用者が、ISBN 番号等の書籍情報の一部を入力し自動記入ボタンを押すと、記入内容自動抽出エージェントに書籍情報を問い合わせ、記入候補選択画面を表示する。利用者が記入候補を選択すると、その内容を記入ルールに基づき図書登録フォームに自動記入し、記入済みフォームを表示する。記入内容自動抽出エージェントは、ISBN 番号等をキーワードに、外部サービス群に書籍情報を問い合わせ、その結果を記入内容候補として出力する。Web サービス変換と Web アプリケーション変換機能は、個々の外部サービスのインタフェースの違いを吸収する。サービス動的複合機能では、状況に応じてサービスを選択して問い合わせ、検索結果を複合する。図の Web サービス群は、書籍情報が検索できる SOAP ベースの Web サービスである。Web アプリケーション群は、書店などのホームページにある書籍検索機能で、Web ブラウザ向けに構築された一般の Web アプリケーションである。

3.2 対象ドメインのオントロジ設計

記入ルールは、どのフォーム項目への記入ルールかを示す識別子、どのような内容を記入するかを示す情報の概念名、記入書式の詳細の 3 要素から構成

される。記入内容は、情報の概念名とその具体的な値の 2 要素からなる。これら 2 つに共通する情報の概念名は、記入対象ドメインの共通概念であるオントロジ (語彙) としてドメイン毎に設計する。

これまでの個人情報の自動記入方式の研究では、記入対象のフォーム項目から共通する特徴を抽出し、56 の基本項目からなるオントロジを設計した。基本項目以外の項目は合成項目と呼び、基本項目からの変換ルールと項目の組合せルールにより定義する。

今回は対象とする書籍情報について、記入対象のフォーム項目の他に、記入内容を抽出するサービス群の検索結果の形式も参考にし、共通する特徴から著者、訳者、タイトル、サブタイトル、シリーズ名、出版社、出版年月日、ISBN 番号、言語、ページ数、定価の 11 の基本項目からなるオントロジを定義した。さらに、著者と訳者を組合わせた項目などよく利用される約 30 の合成項目をあらかじめ定義した。

3.3 記入内容の自動抽出方式

記入内容抽出方式は以下の 3 つの特徴を持つ。

- (1) 複数サービスの動的複合
- (2) Web サービス変換機能
- (3) Web アプリケーション変換機能

本システムの記入精度は、記入内容の網羅性と記入ルールの正確さにより決まる。網羅性は、正しい記入内容が取得できる比率で表される。まずはじめに本方式では、網羅性を高めるために複数の外部サービスを組み合わせることで利用することとした。(1) の動的複合機能により複数のサービスに問い合わせ、結果を統合する。組み合わせると網羅性が向上するのは、例えば、新刊の書籍は図書館に登録されておらず、書店には登録される傾向があるなど、各サービスに特色の違いがあるためである。また、複数サービスを利用することで、耐故障性の向上やサービス提供側への負荷分散のメリットもある。

しかしながら Web サービス群は、似たサービスであってもインタフェースが統一されていないことが多い。このため個別に問い合わせ機能を構築する必要があり、このままでは扱いにくい。そこで、(2) のサービス変換機能を設けて、各サービスを共通インタフェースに変換することでこれらのインタフェースの違いを吸収する。サービスの動的複合機能からは共通インタフェースで Web サービス変換機能に問い合わせれば良い。

さらに、本方式では多くのサービスを利用するほど網羅性が高まると考えられるが、現状の Web サービスは数が少なく選択肢が限られる問題がある。そこで、現在普及している HTML ベースの Web アプリケーションを、Web サービスヘラッピングする (3) の Web アプリケーション変換機能を開発した。

図 1 に示すように、記入内容自動抽出エージェントは、これらの 3 機能に対応する 3 つのサブフレームワークからなるサービス変換フレームワークと、自動記入エージェントからの要求を受け付け結果を返す書籍情報自動抽出機能から構成される。

4 応用システムの実装

4.1 共通インタフェースの開発

まずはじめに、検索条件に一致する書籍の検索対象項目を検索する問い合わせに対し、結果として検索対象項目の具体的な値のリストを返す、共通インタフェースを開発した。検索対象項目は 3.2 節で定義した書籍のオントロジで表現される。

次に、各サービス機能の詳細を定義するサービス定義言語を開発した。各サービスは、変換前の Web サービス群と Web アプリケーション群の持つ機能の違いにより、利用できる共通インタフェースの機能が異なる。例えば、書店では検索結果に価格が含まれるが、図書館の検索結果には価格が含まれない。サービス定義では、利用できる検索条件と、検索できる対象項目を変換ルールごとに定義する。

4.2 サービスの動的複合方式

サービスの動的複合では、利用するサービスの選択、選択したサービスへの問い合わせ、結果の複合の 3 つの処理を行う。サービス複合の目的は、サービスを組み合わせることで、全体としての網羅性を高めることである。サービスの選択方法には、単純に全てのサービスに問い合わせる方式もあるが、網羅性は高い反面、サービス数が増えるとネットワークや処理の負荷が高い問題がある。最小限のサービスの組み合わせで実現できる選択方式が求められる。そこで本システムは、各サービスのヒット率を集計し、ヒット率の良いサービスを選択する方式とした。ヒット率は、最近 100 回の ISBN からの書籍検索で、1 つ以上の結果が得られた回数割合とした。ヒット率が高いサービスは書籍データの網羅性が高いと考えられる。まず、初期値として研究室の図書 100 冊

を検索した結果の割合を設定し、後は実際の実行結果を反映していく。書籍情報自動抽出の具体的な選択ルールの適用手順を以下に示す。

- (1) 全ての検索対象項目を検索できるサービスがあれば、そのヒット率上位から順に合計ヒット率 200 % を上限にサービスを選択する。この時点で 200 % に達した場合は以下は実行する必要が無い。
- (2) 検索対象項目ごとにその項目を検索できる利用可能サービスをリストアップし、利用可能サービスが少ない項目から順に以下を実行する。
- (3) 項目の利用可能サービス中で、既に他で選択されたサービスも含めて、合計ヒット率が 200 % 以上になるようにヒット率上位から順に選択する。

200 % は応用システムにおいてこれ以上では結果が変わらずこれ以下では悪くなるという経験値から決定した。

次に、選択された全てのサービスに同時に書籍情報を問い合わせる。問い合わせ結果はスコア順にソートし、記入候補選択画面では最もスコアの高い候補が選択された状態で表示される。サービスのスコアは、利用者が最終的にそのサービスから抽出された記入内容候補を選んだ回数で、よく選ばれたサービスほどスコアが良くなる。なお、複数サービスへ同時に問い合わせるのは処理全体の応答時間を短くするためである。結果を待ち、見つからなかった場合に次のサービスへ問い合わせる方法も考えられるが、2 回目が終わるまで処理時間が 30 秒以上かかり実用面で劣る。また本システムでは、さらに応答時間を短縮するために、応答に 30 秒以上かかっているサービスは待たずに、先に応答のあった結果のみを利用する。30 秒の時点で結果が空の場合にのみさらに応答を待つことにした。

4.3 Web サービス変換方式

Web サービス変換は、要求メッセージ変換と、応答メッセージ変換の 2 つの機能からなる。要求メッセージ変換では、書籍情報の検索条件を共通形式からサービス固有形式の要求メッセージへ変換する。要求メッセージのひな形と、ひな形への ISBN 番号などの検索条件の埋め込みルールを定義しておく。応答メッセージ変換では、サービス固有形式の書籍情報から共通形式へ変換する。応答メッセージからの



図 2: 自動記入候補の選択画面例

書籍情報の切り出しルールを定義しておく。要求メッセージ変換の、ひな形のどの場所に何を埋め込むかというルールは、フォームのどの項目に何を記入するかという記入ルールと基本は同じである。応答メッセージ変換の、応答結果のこの場所に何が書かれているかというルールは、要求メッセージ変換ルールや記入ルールの逆変換である。そこで、Web サービス変換ルール記述言語は、RDF ベースの自動記入ルール記述言語 [1] をベースに開発した。

4.4 Web アプリケーション変換方式

Web アプリケーション変換では、Web アプリケーションを共通インタフェースの Web サービスとしてラッピングする。本システムでは、Web アプリケーションへの要求を、人が Web ブラウザを操作するのと同じ手順で行う。まず、Web アプリケーション変換のためのルールを、リンクやボタンのクリック、フォーム項目への記入などからなる Web ブラウザの操作手順と、Web ページからの情報の切り出し手順により定義する。そして、GUI を持たない仮想 Web ブラウザを用いてこれらのルールを実行する。変換ルールの定義には、Web アプリケーションの操作手順は手続き的な要素が大きいため、スクリプト言語の JavaScript を用いて記述することとした。以下に図書館図書検索の問い合わせ手順例を示す。

- (1) メニューページを開く
- (2) 検索フォームへのリンクをクリック
- (3) 検索フォームに検索条件を入力

- (4) フォームの送信ボタンをクリック
- (5) 結果一覧の 1 つ目の結果へのリンクをクリック
- (6) 結果ページの切り出し
- (7) 次の結果について (5) から繰り返し

応答時間の短縮のため、(1) のメニューページは利用する検索フォームへのリンク部分が常に同じなのでキャッシュしておいたものを用いる。また、(5) から先は並列に実行する。なお、この例では (2) の検索フォームは毎回異なる ID が埋め込まれているためキャッシュを用いることはできなかった。

4.5 自動記入エージェントの利用手順と実装環境

図 1 の自動記入エージェントは HTTP プロキシとして動作する。本システムは、利用者端末または組織内のプロキシサーバで稼働させることを想定している。主な利用手順は以下の通りである。

- (1) Web ブラウザが取り寄せている Web ページに対応する記入ルールがある場合、フォームに自動記入ボタンを埋め込み、Web ブラウザに渡す。
図 2 左上の新規図書登録フォームにある ISBN 記入欄の右の「本の情報を検索」ボタンが埋め込まれた自動記入ボタンである。
- (2) 利用者がフォームの一部を入力し自動記入ボタンをクリックすると、自動記入エージェントが呼び出される。
- (3) 自動記入エージェントは、記入内容自動抽出エージェントに一部入力された内容を渡し、書籍情報を問い合わせる。
- (4) 記入内容自動抽出エージェントは複数のサービスから書籍情報を検索し結果一覧を返す。
- (5) 自動記入エージェントは、図 2 右下に示すような自動記入候補の選択画面を、第 1 候補を選択した状態で表示する。
- (6) 利用者が記入候補から適切なものを選択し記入ボタンを押すと、自動記入済みのフォームが Web ブラウザに表示される。

一連の手順で実際に研究室の図書登録フォームで自動記入できることを確認した。また、新たに図書館の図書リクエストフォームの記入ルールを作成し、自動記入できることを確認した。

本システムの実装には、Java 1.4.1 と、SOAP に対応した Web サービス用フレームワークである Apache

WSIF(WebService Invocation Framework) 2.0 を用いた。また、Web アプリケーション変換の仮想 Web ブラウザ機能の実現には Web アプリケーションテスト用フレームワーク HttpUnit 1.5.3 を用いた。

5 考察

5.1 Web 全文検索を利用した自動抽出

検索機能の Web アプリケーションを用意していない Web サイトでも、レイアウトが統一されていれば、Web 全文検索の Web サイトを指定する機能を用いて、Web アプリケーションと同じ手法で自動抽出が可能である。この方式では、ISBN 番号の場合や、長いタイトルなどの固有のフレーズが含まれる場合は、短時間で結果を抽出できるが、短いタイトルの場合等では全文検索の検索結果が多くなり、結果の処理時間が長くなる傾向があった。

5.2 変換可能サービスの動的発見

図 1 の Web サービス群はサービス変換ルールが対応している WSDL に一致するサービスを、UDDI により動的に発見し利用することができる。今後は、さらに変換ルールも動的に発見できるようにした、UDDI ベースのディレクトリサーバを検討していく。

6 まとめ

本研究では、Web サービス変換により記入内容を自動抽出する自動記入方式を確立した。Web サービスの普及に伴いインタフェースの標準化は進むが、今後も同分野のサービスが異なるインタフェースで提供されるケースがなくなることはない。本研究の Web サービス変換フレームワークの基本概念は、フォーム自動記入に限らず広く Web サービスを利用したアプリケーション開発に適用可能と考えられる。また、Web アプリケーションのラッピング機能は、既存のシステムを変更することなく Web サービス化できるため、コストがかかる等の理由で Web サービス化されない分野に有効である。

参考文献

- [1] 藤原克哉, 中所武司: 窓口業務アプリケーションフレームワーク wwHww におけるルール生成を自動化した自動記入エージェントの実現方式, 情報処理学会論文誌, 43, 6, pp.1653-1662 (2002).
- [2] Tim Berners-Lee, James Hendler and Ora Lassila: The Semantic Web, Scientific American (2001).