

機関番号：12608

研究種目：基盤研究（C）

研究期間：2008～2010

課題番号：20500026

研究課題名（和文）

自動的部分情報抽出方式に基づく知識Webアプリケーションの連携・統合の研究

研究課題名（英文）

Methods for integration of Web applications dealing with knowledge information using automatic partial information extraction

研究代表者

徳田 雄洋（TOKUDA TAKEHIRO）

東京工業大学・大学院情報理工学研究科・教授

研究者番号：30111644

研究成果の概要（和文）：

従来 Web 上の一般的な Web サービス関数を連携させて1つの仕事を達成するにはプログラミングが必要であった。プログラミングせずにダイアグラムから作成する手法も存在したが、扱える Web サービス関数に強い制限があった。

本研究はプログラミングなしに、宣言的記述を用いて、Web アプリケーションおよび REST 型 Web サービスをデスクトップコンピュータやスマートフォン上で連携統合する方法を開発した。

研究成果の概要（英文）：

Traditionally integration of general Web services requires manual programming. Diagram composition approach without programming has strong limitations of applicable Web services.

This research enables us to construct integrated applications of Web applications and REST type Web services on desktop computers and mobile devices using declarative definitions without programming.

交付決定額

(金額単位：円)

| | 直接経費 | 間接経費 | 合計 |
|--------|-----------|-----------|-----------|
| 2008年度 | 1,900,000 | 570,000 | 2,470,000 |
| 2009年度 | 800,000 | 240,000 | 1,040,000 |
| 2010年度 | 800,000 | 240,000 | 1,040,000 |
| 年度 | | | |
| 年度 | | | |
| 総計 | 3,500,000 | 1,050,000 | 4,550,000 |

研究分野：総合領域

科研費の分科・細目：情報学・ソフトウェア

キーワード：Webアプリケーション、Webサービス関数、モバイルアプリケーション、連携統合、部分情報抽出

1. 研究開始当初の背景

研究当初、Web 上で Web サービス関数を連携統合することは広く行われるようになっていたが、主にデスクトップコンピュータ上

で地図情報、住所位置情報、検索結果情報、ネットストア検索情報などを扱うのが主であった。

連携実現法としては手動プログラミングによって Web サービス関数を統合する方法が

一般的方法として行われていた。また事前処理された Web サービス関数をダイアグラム方式で作成する方法も行われていたが扱える Web サービス関数が限定的であった。

したがって一般的な Web アプリケーションや Web サービス関数をプログラミングなしに連携統合する方法は存在していなかった。

2. 研究の目的

本研究の目的は、本研究グループが開発した手動プログラミングなしに Web ページの部分情報抽出を行う技術に基づき、一般的な Web アプリケーションや Web サービス関数を宣言的記述から連携統合する技術を開発することである。

Web ページの部分情報抽出技術は、Web サイト上で時間的に一連と考えられる、あるいは空間的に一連と考えられる Web ページについて、取得したい必要部分の情報を、手動プログラミングなしに、GUI で利用者が直接指定して取得する技術である。

例えば Web 上の百科事典のある 1 項目を選び、最初のパラグラフのみ、あるいはテキスト説明全体のみ、あるいは位置情報記述の部分のみ、などと指定し、百科時点の他の項目についても指定部分を自動取得することを実現する技術である。

この技術を用いれば、Web サービス関数を持たない Web アプリケーションと Web サービス関数を、相互に組み合わせ、知識や情報を扱っている各種サイト、百科事典、ニュース、辞書サイトなどの情報を連携統合させることが可能になる。

3. 研究の方法

現状では多くの Web アプリケーションが、対応する Web サービス関数を持っていない。したがってそのままでは Web サービス関数などと連携統合することが困難である。

そこで、部分情報抽出技術を利用する。すなわち、連携統合させるために、Web アプリケーションに引数を与えて得られる結果ページから、指定した必要部分を部分的に取り出すことにより、統合用コンポーネントを作る。

そしてこれらの統合用コンポーネントをプログラミングなしに統合するために、宣言的記述、すなわち XML 記述を与えて、一般の Web アプリケーションと Web サービスが直列的、並列的に自由に統合できるようにする。

これでプログラミングなしに、宣言的記述を用いて、Web アプリケーション、REST 型 Web サービスを柔軟に連携統合させることが可能となる。

2010 年はじめに、Android スマートフォン

の開発標準機である NexusOne が公開されたことにより、デスクトップコンピュータ上で開発した手法をさらにモバイルデバイス上でも実現した。

4. 研究成果

本研究により、デスクトップコンピュータ上やスマートフォン上で Web アプリケーションと REST 型 Web サービスを、プログラミングなしに、宣言的記述から統合することが可能となった。

統合例を 2 つ示す。第 1 の例は Web 上の洋書の価格比較である。ネット上の 3 つの洋書ストアに、購入したい本の ISBN 情報を与えて、価格を調べ、最新の換算レートに基づいて、自国通貨に変換し、価格を比較することが、宣言的記述から自動生成できる。

第 2 の例はスマートフォン上の Web アプリケーションと REST 型 Web サービスの統合である。洋書の ISBN 情報のバーコードをスマートフォンのカメラで読み取り、対応する書名をネット書店の検索から取り出し、さらに対応する書評を書評サイトから得る連携統合アプリケーションを、宣言的記述から生成できるようになった。

これらの研究成果は、米国、フィンランド、オーストリアなどの国際会議において、成果発表を行い、多くの参加者から高い評価を得た。またデモでは代表的なスマートフォンである Android スマートフォンを用いて、Web 上の各種 Web アプリケーションと Web サービスを連携統合するいくつかの実例を示すことにより行った。

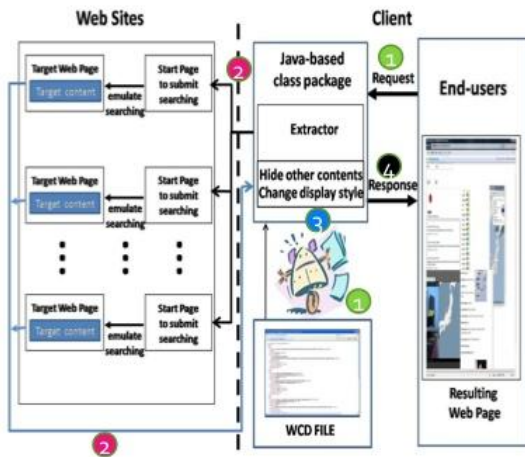
主要論文の概要は以下のとおりである。

(1) 従来の部分情報抽出手法では不可能だった JavaScript 等のスクリプトによって生成される動的コンテンツの抽出手法を開発した。その手法は以下のとおりである。

- ① ユーザは、WCD (Web Content Description) ファイルを作成し、検索キーワードと共にリクエストを送信する
- ② 抽出器は、検索機能をエミュレートし、対象の Web ページを取得する
- ③ WCD ファイルをもとに抽出箇所を特定し、可視属性の制御を行う
- ④ 変更した Web ページをユーザに返す

WCD ファイルでは、抽出対象の Web ページ、抽出箇所を示すパスや型等を指定する。この指定は GUI ベースのツールを利用して行うため、プログラミングの知識等は一切必要としない。本手法により、複数の Web ページから静的・動的コンテンツを自動抽出し、1 つの Web ページに

統合することが可能となった。



- (2) Web アプリケーション、Web サービス、モバイルアプリケーション間の連携統合手法を開発した。例えば、スマートフォンのGPS機能を利用して現在地を検出し、周辺情報や気象情報をWeb上から取得するアプリケーションを作成できる。作成したアプリケーションを、GPS機能を持たない他のスマートフォンから利用することも可能である。



ユーザは MID (Mobile Integration Description) ファイルによって連携統合する Web アプリケーション、Web サービス、モバイルアプリケーションを記述する。Web アプリケーションを連携させる場合には、部分情報抽出技術を利用して抽出箇所を指定する。本手法により、プログラミングなしに容易に連携統合させることが可能となった。

- (3) 部分情報抽出技術を利用し、下記のことを可能にする手法を開発した。
- ① Web アプリケーションから Web サービスを自動生成する
 - ② デスクトップ PC 上で、Web アプリケーションを Web サービスと自動連携統合する
 - ③ スマートフォン上で、モバイルアプリケーションを Web アプリケーションや Web サービスと自動連携統合する

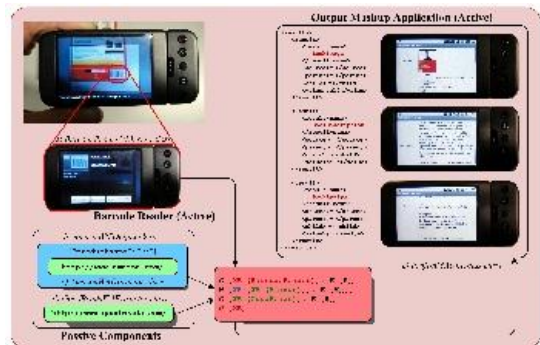
Web サービスの自動生成では、Web アプリケーション上の抽出したい箇所を指定するだけで、Web サービスを作成でき

る。作成した Web サービスはプロキシサーバ上で動作し、抽出結果はテーブルに保存される。Web サービスはユーザからの要求を受けて結果テーブルを検索し、結果を返す。

デスクトップ PC 上での Web アプリケーションと Web サービスの自動連携統合では、直列的および並列的な連携統合が可能である。例えば、国名を入力として与えると、Localtimes.info からその国の首都の現地時刻を、BBC Country Profile からその国の首都や人口等の基本情報を取得し、さらに取得した首都名を利用して、Weatherbonk.com からその都市の天気を、ABC News からその都市名を含む記事を取得するアプリケーションを作成できる。



スマートフォン上でのモバイルアプリケーションと Web アプリケーション、Web サービスの自動連携統合では、スマートフォンが持つ機能を利用した連携統合アプリケーションを作成できる。例えば、スマートフォンのカメラを利用し、バーコードリーダーで本の ISBN コードを読み取って Amazon で書名を検索し、それをもとに GoodReads で書評を取得するモバイルアプリケーションを作成できる。



これらの連携統合は、すべてプログラミング知識を必要としない。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕 (計 1 件)

- ① Hao Han, Tomoya Noro, and Takehiro Tokuda. “An Automatic Web News Article Contents Extraction System Based on RSS Feeds”. Journal of Web Engineering, 8(3), pp.268-284, Rinton Press, 2009. 査読有

〔学会発表〕 (計 1 3 件)

- ① Junxia Guo, Hao Han, and Takehiro Tokuda. “Towards Flexible Mashup of Web Applications Based on information Extraction and Transfer”. WISE2010, pp.602-615, Dec. 14 2010, Hong Kong, China. 査読有
- ② Junxia Guo, Prach Chaisatien, Hao Han, Tomoya Noro, and Takehiro Tokuda. “Partial Information Extraction Approach to Lightweight Integration on the Web”. ComposableWeb2010, pp.372-383, Jul. 5 2010, Vienna, Austria. 査読有
- ③ Junxia Guo, and Takehiro Tokuda. “Description-Based Mashup of Web Applications”. ICWE2010 Doctoral Consortium, pp.545-549, Jul. 5 2010, Vienna, Austria. 査読有
- ④ Prach Chaisatien, and Takehiro Tokuda. “A Description-based Approach to Mashup of Web Applications, Web Services, and Mobile Phone Applications”. EJC2010, pp.182-201, Jun. 1 2010, Jyvaskyla, Finland. 査読有
- ⑤ Hao Han, Junxia Guo, and Takehiro Tokuda. “DEEP Mashup: A Description-Based Framework for Lightweight Integration of Web Contents”. WWW2010, pp.1109-1110, Apr. 28 2010, Raleigh, USA. 査読有
- ⑥ Hao Han, and Takehiro Tokuda. “A Layout-Independent Web News Article Contents Extraction Method Based on Relevance Analysis”. ICWE2009, pp.453-460, Jun. 26 2009, San Sebastian, Spain. 査読有
- ⑦ Hao Han, Junxia Guo, and Takehiro Tokuda. “Towards Flexible Integration of Any Parts from Any Web Applications for Personal Use” ComposableWeb2009, pp.69-80, Jun. 23 2009, San Sebastian, Spain. 査読有
- ⑧ Junxia Guo, Hao Han, and Takehiro Tokuda. “A New Partial Information Extraction Method for Personal Mashup Construction”. EJC2009, pp.157-170, Jun. 2 2009, Maribor, Slovenia. 査読有

- ⑨ Hao Han, and Takehiro Tokuda. “A Quick Web News Article Extraction Approach Based on the Existing News Search Engines”. 日本ソフトウェア科学会第 25 回大会, 7B-4, Sep. 12 2008, 東京. 査読無
- ⑩ Prach Chaisatien, and Takehiro Tokuda. “Personalizable Mobile Web: An Approach to Mobile Mashup with Limited Devices”. 日本ソフトウェア科学会第 25 回大会, 6C-1, Sep. 12 2008, 東京. 査読無
- ⑪ Hao Han, and Takehiro Tokuda. “A Method for Integration of Web Applications Based on Information Extraction”. ICWE2008, pp.189-195, Jul. 17 2008, Yorktown Heights, USA. 査読有
- ⑫ Hao Han, and Takehiro Tokuda. “WIKE: A Web Information/Knowledge Extraction System for Web Service Generation”. ICWE2008, pp.354-357, Jul. 16 2008, Yorktown Heights, USA. 査読有
- ⑬ Hao Han, Yohei Kotake, and Takehiro Tokuda. “An Efficient Method for Quick Construction of Web Services”. EJC2008, pp.197-213, Jun. 3 2008, Tsukuba, Japan. 査読有

6. 研究組織

(1) 研究代表者

徳田 雄洋 (TOKUDA TAKEHIRO)
東京工業大学・大学院情報理工学研究科・教授
研究者番号：30111644

(2) 研究分担者

野呂 智哉 (NORO TOMOYA)
東京工業大学・大学院情報理工学研究科・助教
研究者番号：80401553

(3) 連携研究者

()

研究者番号：