

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 3 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2014～2015

課題番号：26730158

研究課題名(和文) Web機能の再利用と統合に向けた大規模UIコンポーネントの検索方法の研究

研究課題名(英文) A Research on Retrieval and Reuse and Integration of Large-scale User Interface Components

研究代表者

韓 浩 (Han, Hao)

東京大学・医学部附属病院・助教

研究者番号：20600802

交付決定額(研究期間全体)：(直接経費) 2,000,000円

研究成果の概要(和文)：Webの構造・情報属性やWeb利用者の情報検索・閲覧行動を総合的に分析し、Web機能の再利用のため、大規模の統合サービスの実現に向けて安定的な情報抽出技術と系統的な統合方式及びテスト・メンテナンス手法等の開発を行った。その結果、ユーザインターフェースコンポーネントの認識方法と記述方法を提案し、検索方法と再利用・統合手法などに関する研究を行い、異なる利用者に向けた総合開発環境とクラスパッケージを開発した。さらに、Web視聴率データから利用者の検索行動を抽出してWeb情報の時効性と利用パターンの分析などを実施した。そして、研究結果を活用して健康情報学に応用として展開している。

研究成果の概要(英文)：For reusing and integrating large-scale web functionalities, we introduced and developed stable partial information extraction methods and various systematic integration/testing approaches based on a comprehensive analysis of web structure, information property and diverse browsing behavior patterns of client-side users. Consequently, we proposed UI component retrieval method and developed various integration environments (e.g., class packages for programmer and GUI for non-programmer) based on the recognition and description of UI components. Moreover, we extracted behavior patterns of panel users from net view data and analyzed query satisfaction levels for further applications on healthcare informatics.

研究分野：ウェブ情報学

キーワード：ウェブシステム ウェブ機能検索 ウェブ機能統合 UIコンポーネント

1. 研究開始当初の背景

(1) Web 上のサービスの効率的実現のため、既存の Web 機能 (Web アプリケーション・Web ページの一部として利用者の操作等に応じて動的にコンテンツを生成・提示する仕組み) を活用する技術が高度化してきており、複数の Web サービスを融合して新しいサービス・アプリケーションとして提供する形態を目指すマッシュアップの技術が目ざされている。

(2) しかし、現状では、Web 機能の利用については、ほとんどが Web サービス API に依存している。このため、Web サービス API を提供していない有用な Web 機能が非常に多くあるにもかかわらず、利用できる Web 機能が制限されているし、手動で Web 機能の検索を行うことを前提としている。ここで、従来の Web 検索技術が、主として Web ページのコンテンツに表現された情報を対象として研究されてきたことが問題となる。

(3) Web 機能を利用している個々の Web ページには Web 機能自体に関する十分な記述がないことが多いため、従来の検索エンジンは Web 機能を直接的な対象として検索するのに適しておらず、手動により Web 機能を大量の Web ページから網羅的に検索し、抽出・分析・比較等を行うために、最適な機能の発見は膨大な時間と専門知識が要求されることになる。また、Web 機能の統合方式と応用について、UI コンポーネントの適応性と信頼性を考えず、例外処理及びセキュリティを重視していない現状である。

(4) そこで、これをより広範囲な対象に対して大規模に適用可能とするため、Web サービス API を提供していない一般の Web 機能の再利用を可能とする手法と最適な Web 機能・UI コンポーネントを検索する手法と Web 機能の統合方式及び統合サービスの系統的な開発手法が求められていた。

(5) 利用者の行動分析に基づく、情報アクセス技術に展開する。技術の発達に伴うコストパフォーマンスの高いユーザーインターフェイスの効率を検証する。

2. 研究の目的

(1) 本研究では、まず Web 上の大量の Web ページから機能を含む Web ページを収集する技術、各種の Web 機能を代表する UI コンポーネントを抽出・総合分析する技術、および Web 機能・UI コンポーネントの検索に必要な情報の分析・生成を行う技術について研究を行う。

(2) 本研究では、コンポーネントのソースコードを分析し、タイプと構造、使われた言語、通信プロトコルなどの情報を表示する。様々

な観点から複雑性・安全性などの評価指標を導入する。なお、異なるサイトの様々な機能を再利用・統合するため、機能検索エンジンを活用して、コンポーネントの属性とコンテンツの構造の分析、インターフェースの定義、機能のエミュレーション、部分情報の抽出、コンテンツの統合と表現形式のパーソナライゼーション等の大規模の統合サービスの自動生成の必要な技術の高度化を図る。

(3) 統合サービスの系統的なテスト・メンテナンス方式とセキュリティチェック方法を開発する。

(4) UI コンポーネントの効率的な統合のため、総合的な開発環境を提供する。

(5) 単純な HTML で記述されたページよりも操作性や表現力に優れたリッチインターネットアプリケーションが増えている。クライアントレンダリングメカニズムに応用し、アプリケーション開発を計画する。

(6) 情報のアクセス方式と属性を分析して利用者の閲覧満足度などを総合的な統計を行う。

3. 研究の方法

具体的な解決策としては、大規模の Web ページを収集し、コンポーネントの認識・抽出・分類を行い、検索に有効な Web 機能の表現方式を明らかにして Web 機能検索エンジンを実装する。さらに、Web 機能のインターフェースのユニフォーム化と適応性の検証を行い、部分情報抽出手法を進化し、セキュリティポリシーとテスト・メンテナンス手法を用いる統合方式に展開し、大規模の統合サービスに適用する系統的な開発と評価方法を目指す。

(1) 手動でコンポーネントを含む Web ページのタイトルとメタデータ、コンポーネントが使うフォームとタグの情報、構造とクエリ送信方式などの情報を収集した。そして、収集した情報を機能によって分析・分類して、各 Web 機能の特徴を認識パターンとして取得した。視聴率データにより解析したパターンと分類情報を利用して、大規模の検索と分析実験を行った。さらに、様々な基準からコンポーネントの複雑性・機能の適応性・コンテンツの多様性・情報流れの安全性などを評価した。

(2) コンポーネント・機能の様々な再利用技術を開発し進化させ、Web サービスの自動生成、大規模のメタサーチとディープ Web などへの応用を拓いた。さらに、異なる Web 機能を統合して、大規模の統合サービスの実現に向けて、安定な情報抽出技術と系統的な統合方式及びテスト・メンテナンス手法等の開発

を行った。

(3) データフロー記述を GUI 化するシステムとプログラミングパッケージの提供を行い、適用性の高い統合システムを実装し、統合サービスの開発・管理環境を構築した。

(4) 利用者の視点から情報の属性より社会的な応用可能性を拡張した。クラスタリング・機械学習などの手法を利用して、ユーザの閲覧行動・傾向の比較検討やデータマイニングを行って、利用者の検索満足度評価に基づき、情報の時効属性を分析した。

(5) クライアントサイドの XML と XSLT 技術に基づき、複数のデータモデルに対応した柔軟なテンプレートエンジンを提案した。このテンプレートエンジンの利用をケース-スタディーとして分析して、クライアントサイドレンダリングメカニズムの適用性・有効性を示した。さらに、開発コスト・システム負荷の最適化を目指して進化する Web プログラミング支援システムの開発を行った。

(6) さらに、アプリケーション統合手法を利用して健康情報収集・分析システムを試作した。

4. 研究成果

(1) 基盤になる各種データ（視聴率データ、Web ページなど）を導入して UI コンポーネントの認識と記述方式の検討・実験を行った。また、データフロー記述を GUI 化するシステムとプログラミングパッケージを開発し、適用性の高いコンポーネント検索・統合システムを実装した。

(2) ブラウジング行動フィーチャーの有効性の解析に基づく Yahoo! 知恵袋のクエリ満足度レベルの探索的予測手法を提示し、24498 名のパネルユーザによって実行された 120182 件のクエリプロセスのデータセットに基づく実験を介して様々なカテゴリの満足度を示した。また、これらの差に関する潜在的な理由について説明し、予測結果に基づくフィーチャーの有意性について検討した。

(3) 健康に影響する潜在要素を把握するために、主に普段の飲食生活の行動をより簡単な方法でデジタル化して収集する手段の研究を行った。モバイルアプリケーションの統合と画像の認識技術の実現性や、利用者の自主的記録方法などの方向を考案し、アプリケーションの試作を行った。

(4) さまざまなプラットフォーム上で動作する Web アプリケーションの情報を利用するため、Web 上の部分情報抽出などの技術を利用し、サービス指向アーキテクチャの基盤技術である Web サービス生成への支援について

研究した。

(5) 斬新な Web アプリケーションのテンプレートエンジン「Pagelet」を設計し、Web プログラミングを支援するシステムの開発を行った。クライアントサイドレンダリングメカニズムの適用性・有効性を検証した。

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 5 件)

韓浩, 中山堯. モバイルアプリケーションの統合に基づく食と健康の情報の収集・分析システムの設計と試作, Science Journal of Kanagawa University, Vol.27, 2016 (査読なし)

韓浩, 郭俊霞, 中山堯. Q&A コミュニティに注目したブラウジング行動に基づく検索満足度の予測分析, Science Journal of Kanagawa University, Vol.26, 2015 (査読なし)

Hao Han, Hidekazu Nakawatase, and Keizo Oyama. Evaluating credibility of interest reflection on Twitter, International Journal of Web Information Systems, Vol.10, No.4, pp.343-362, 2014 (査読あり)

Hao Han, Yinxing Xue, Keizo Oyama, and Yang Liu. Practice and Evaluation of Pagelet-Based Client-Side Rendering Mechanism, IEICE TRANSACTIONS on Information and Systems, Vol.E97-D, No.8, pp.2067-2083, 2014 (査読あり)

Junxia Guo and Hao Han. A Method for Facilitating End-user Mashup Based on Description, International Journal of Web Engineering and Technology, Vol.9, No.2, pp.99-124, 2014 (査読あり)

[学会発表](計 3 件)

Hao Han, Junxia Guo, Takashi Nakayama, and Keizo Oyama. Towards Serving "Delicious" Information within Its Freshness Date, Proceedings of 24th International World Wide Web Conference, pp.41-42, 2015 (査読あり)

Hao Han and Yinxing Xue. Analysis and Evaluation of UI Component Integration for Large-Scale Web Content Reuse, Proceedings of 7th International Conference on Creative Content Technologies, pp.18-24, 2015 (査読あり)

Junxia Guo, Hao Han, Cheng Gao, Takashi Nakayama, and Keizo Oyama.

Prediction of Query Satisfaction Based
on CQA-Oriented Browsing Behaviors,
Proceedings of 6th International
Conference on Information Science
and Applications, pp.1005-1012, 2015
(査読あり)

6 . 研究組織

(1)研究代表者

韓 浩 (HAN, Hao)

東京大学・医学部附属病院・助教

研究者番号：20600802