

## 科学研究費助成事業 研究成果報告書

平成 30 年 6 月 5 日現在

機関番号：14401

研究種目：若手研究(B)

研究期間：2014～2017

課題番号：26730155

研究課題名(和文) Web・クラウド技術の利用事例検索のためのソフトウェアトレンドマイニング

研究課題名(英文) Mining Software Trends for Case Study of Web and Cloud Technologies

研究代表者

松本 真佑 (Shinsuke, Matsumoto)

大阪大学・情報科学研究科・助教

研究者番号：90583948

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：本研究ではソフトウェアトレンドの提示による開発支援を目的として、Web/クラウド技術の利用事例に基づいたトレンドマイニング手法を提案した。本研究の遂行により、Webエンジニアに対して各種技術のトレンドや動向の把握、ライブラリの利用事例検索といった様々な開発支援サービスを提供することができる。提案手法の適用実験として、Webリソースの最適化技術のトレンド把握、およびブラウザ上で発生するJavaScriptの実行時エラーの収集システムを開発した。

研究成果の概要(英文)：The goal of this study is to realize a technique of software trend mining for supporting web and cloud system development by showing the current trends of these tools and frameworks. Our proposed concept provides variety of development assistant systems such as, visualizing rise and fall of some web technologies and searching actual use cases of software libraries. As a experimental evaluation, web apply our system to analyze techniques of web resource optimization. Furthermore, we develop a system for collecting and notifying JavaScript errors which are occurred on a Web browser.

研究分野：ソフトウェア工学

キーワード：ウェブマイニング トrendマイニング Web・クラウド

### 1. 研究開始当初の背景

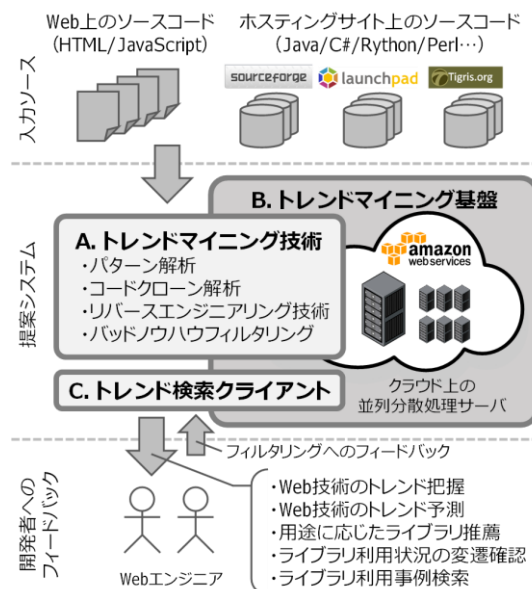
(1) 日々、目まぐるしく進化する Web やクラウドの世界において、様々な Web 技術の「トレンド」を把握することは Web エンジニアにとって極めて重要なスキルの一つである [1]。特に HTML5 や CSS3, JavaScript をはじめとする Web/クラウドの技術は“Living Standard” (生きた標準) と呼ばれており、Web の標準そのものが世の中のトレンドやデファクトスタンダードを反映して策定されている。さらに昨今のライブラリやツールはサービス指向化 (あるいは WebAPI 化) が浸透しており、単純な API の組み合わせだけで高度な Web アプリを実現することも容易になってきた。そのため Web エンジニアは個々の技術に関する深い知識を得るのみならず、世の中全体のトレンドや動向、その利用事例や用途、長所短所を幅広く把握し整理できるスキルが必須となる。

(2) しかしながら激動の最中にある Web の世界において、最先端の技術を追従・把握し続けることは容易ではない。その技術の進化の早さが一つの要因であるほか、標準自体が「生きている」こともトレンド把握の難しさの一つの要因である。また Web 上に公開されているライブラリやツール、フレームワークも多種多様であり、類似目的であっても様々な優れたライブラリが公開されている。例えば JavaScript の利用を簡便化するフレームワークとしては、jQuery や Prototype.js, Dojo, MooTools 等様々なものが存在する。これらは当然ながらそれぞれ長所短所、向き不向きが存在しており、どのような目的・用途の場合にどのフレームワークを使うべきかを適切に選択できる能力も必要となる。さらにバージョン更新によって、ライブラリの利用方法が大きく変化するケースも数多く見受けられる。2013 年 3 月に実施された TwitterAPI のバージョン更新では、全てのサードパーティ Twitter クライアントにおいて改訂作業が必要であった。Web/クラウド技術は日進月歩の勢いで進化する一方で、Web 上に残された古い情報は長期に渡って消えることがない。そのため、Web 上の旧バージョンの情報は役に立たないどころか、ライブラリ利用の妨げや不具合の要因となる可能性も考えられる。

### 2. 研究の目的

本研究の目的は、Web/クラウド技術の利用事例に基づいたトレンドマイニング手法の確立である。トレンドマイニングでは、Web 上に公開された様々なソースコードを対象とした超大規模マイニングを実施し、Web 上に散らばった「知識の宝庫」を発掘する。本研究の遂行により、Web エンジニアに対して各種技術のトレンドや動向の把握、ライブラリ利用事例検索といった様々な開発支援サービスを提供することができる。

提案手法の概要を図 1 に示す。図上部はシステムへの入力ソースとなるデータであり、Web 上のソースコードやホスティングサイト上の各種リポジトリが示されている。図下部は提案システムの出力結果であり、開発者は Web 技術のトレンド把握や、ライブラリ利用事例検索のサービスを得ることができる。図中央が研究対象となる提案システムを構成する以下 3 つのコンポーネントを表す。



A. トレンドマイニング技術は入力ソースとなる様々なソースコードから、トレンド情報やライブラリ利用事例を発掘するコンポーネントである。ソフトウェアリポジトリのマイニング研究分野で提案されているソースコードのパターン解析技術やコードクローン解析技術を応用する。B. トレンドマイニング基盤はトレンドマイニングを実現するための実行基盤に該当する。トレンドマイニングは単一プロジェクトを対象とした一般的なマイニング手法と異なり、複数のプロジェクトに横断的な超大規模マイニング実行環境が必須となる。

### 3. 研究の方法

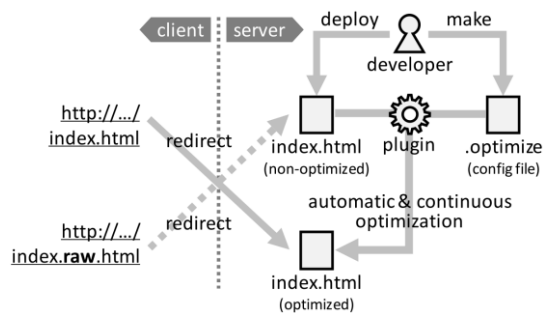
(1) トレンドマイニングの実現に向け、(a) Web リソースのクローリング技術、(b) クローリング結果の統合データベースの設計、及び、(c) a と b を活用した JavaScript 最適化技術の普及率の調査、の 3 つの研究に取り組んだ。(a) と (b) では Web 上に存在する様々な Web リソースを読み込み、マイニング処理しやすい形に処理を加えた上で大規模データベースに保持する仕組みである。膨大な数の Web データを処理するためのスケールアウト可能な設計となっている。動作確認として、Amazon の仮想 Web サーバに本コンポーネントを設置し複数台並列でのクローリングを実施中である。(3) では(1) と(2) の活用事例として、ある Web トレンド技術のマイニング、及び調査に取り組んだ。具体的な対象は

JavaScript Minification と呼ばれる技術であり、JavaScript に対して事前に最適化や圧縮処理を加えることでネットワーク通信量の削減を狙う手法である。さらに HTTP 圧縮と呼ばれる、HTTP 通信のパケットを圧縮して転送する技術も同様に調査対象の Web 技術とした。

(2) さらに、トレンドマイニング技術の適用実験として、Web リソース実装において広く用いられている JavaScript ライブラリ jQuery、Web のエコシステムにおいて不可欠な Web 広告、および Web 上で発生する JavaScript 実行時のエラーの3つを題材とした実験を行った。特に JavaScript 実行時エラーは Web リソースの開発における重要な問題である。Web の仕様は常に進化しており、過去に正しく動作していたページが動作しなくなるということも少なくない。これは、仕様やブラウザに対するリソースの相対的な劣化と見なすことができる。このリソース劣化によるエラーの発生という問題を解決するために、トレンドマイニング技術を応用した、ユーザ参加型エラー収集システムを開発した。

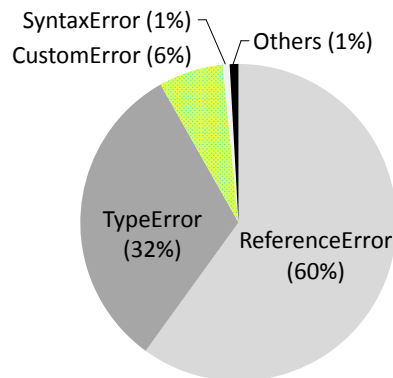
#### 4. 研究成果

(1) まず、JavaScript 最適化技術のトレンド調査の結果として、アクセスランキングの高い、すなわち需要の高い大手 Web サイトであっても、Minification 処理を徹底している割合は 6%にとどまっていることが明らかとなった。また、HTTP 圧縮は 90%近い普及率を持っている一方で、アクセスランキングが高くない、その他の Web サイトでは 20%にとどまっていることが分かった。普及率が低い一つの要因としては、最適化技術と Web のオープン性の齟齬が考えられる。一つの解決策を以下の図に示す。一般利用においては Web リソースの最適化を施し、ソースコードそのものにアクセスする際には未最適化のファイルへアクセスさせるという方法が有効であると考えられる。

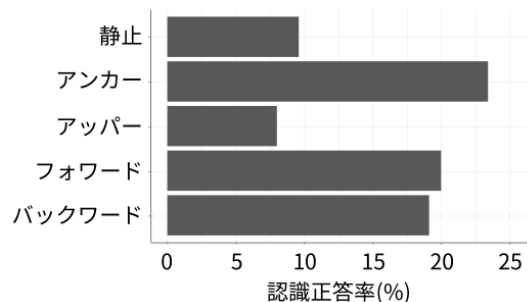


(2) Web 上で発生する JavaScript 実行時のエラーの傾向を以下の図に示す。この結果は被験者 8 名に専用のエラー収集プローブを利用してもらい、得られた 4,172 件の JavaScript

エラーの種類の傾向を表している。図より、Web 閲覧時に発生するエラーは参照エラーと型エラーが大半を占めることが分かる。参照エラーは<link>タグや<script>タグによってロードされるリソースが何らかの理由によりアクセスできない状態を表しており、CSS が崩れる等の Web ページが期待通りに動作していない可能性がある。提案ツールによりこのようなエラーを開発者に能動的に提示することが可能となり、エラーの解決に繋がると考えられる。また型エラーは関数でない変数を関数とみなして呼び出すといった、型の不整合により発生するエラーである。JavaScript では実行時に暗黙的な型変換が行われる言語であり、実行するまで型エラーが発生するかどうかは把握できない。TypeScript 等の型検査が可能な技術の普及によりこれらのエラーの発生を低減出来ると考えられる。

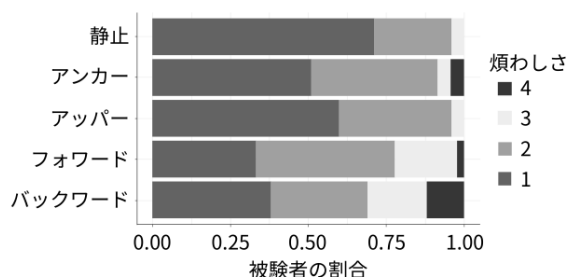


(3) Web のコンテンツ提供者が、広告をページ上のどこに、どのように配置するかは重要な問題である。特にスマートフォン等の画面領域の限られたデバイスでは、広告によってページの UI/UX が大きく低下する可能性がある。そこで、Web 広告配置方法のトレンド、及びその影響について、クラウドソーシングサービスを用いてその効果を定量的な調査を行った。以下の図は 5 種類の広告配置方法に対して、広告そのものをどの程度正しく認識できたかを表す結果であり、正答率が高いほど宣伝効果が高いとみなすことが出来る。



また、次の図は広告に対する被験者の感じた「煩わしさ」を計測した結果である。煩わし

さが高いとはすなわち広告配置による UI/UX



の低下度合いとみなすことが出来る。

結果より、最も幅広く利用されている静止配置とアンカー配置は正反対の性質を持っており、宣伝効果を取るかUI/UXを優先するかがトレードオフになっていることが分かる。また新たに提案した配置方法であるアッパー配置(画面の上部に常に表示する戦略)は静止配置と同程度の煩わしさと宣伝効果を持っており、今後のWeb広告配置の一つの戦略となりうると考えられる。

#### [引用文献]

[1] A. Aurum, "Innovative Thinking in Software Development" Encyclopedia of Information Science and Technology, pp. 2061-2065, 2009.

#### 5. 主な発表論文等

##### [雑誌論文] (計 1 件)

[1] 松尾 裕幸, 榎本 真佑, 楠本 真二, "省電力プログラムの実現に向けて: 実装手段の違いによる実行時間と消費電力量の比較調査," 情報処理学会論文誌, volume 59, number 4, pp.1262-1272, 2018.

##### [学会発表] (計 11 件)

- [1] Naoto Ogura, Shinsuke Matsumoto, Hideaki Hata, and Shinji Kusumoto, "Bring Your Own Coding Style," In Proc. IEEE International Conference on Software Analysis, Evolution, and Reengineering (SANER), pages 527-531, March 2018.
- [2] 田中 紘都, 榎本 真佑, 楠本 真二, "Java プロジェクトにおける関数型イデオロギの実態調査," 電子情報通信学会技術研究報告, volume 117, number 477, pages 39-44, 2018.
- [3] 中島 弘貴, 榎本 真佑, 楠本 真二, "モバイル端末における Web 広告の配置方法に対する一検討," 電子情報通信学会技術研究報告, volume 117, number 388, pages 69-74, 2018.
- [4] Hiroyuki Matsuo, Shinsuke Matsumoto, and Shinji Kusumoto, "What Makes Software Energy-Efficient?: Make It

Faster," In Proc. IEEE Computer Society Signature Conference on Computers, Software and Applications (COMPSAC), pages 274-275, July 2017.

- [5] 山本 将弘, 榎本 真佑, 楠本 真二, "Web フロントエンド開発者のためのユーザ参加型エラー収集システムの提案," 電子情報通信学会技術研究報告, volume 117, number 271, pages 1-6, 2017.
- [6] 山本 将弘, 榎本 真佑, 楠本 真二, "Web 閲覧時における JavaScript ライブラリ使用による副作用の調査," 情報処理学会研究報告, volume 195, number 7, pages 1-8, 2017.
- [7] 松尾 裕幸, 榎本 真佑, 楠本 真二, "実装手段の異なるプログラムの実行時間と消費電力量に関する調査," 情報処理学会研究報告, volume 2017-SE-195, number 16, 2017.
- [8] 中島 弘貴, 松尾 裕幸, 榎本 真佑, 楠本 真二, "Web 広告が閲覧デバイスに与える副作用の実証的調査," 電子情報通信学会技術研究報告, volume 116, number 287, pages 39-44, 2016.
- [9] 小倉 直徒, 榎本 真佑, 畑 秀明, 楠本 真二, "共同開発環境におけるコーディングスタイルの変更に関する調査," 電子情報通信学会技術報告, volume 116, number 277, pages 103-108, 2016.
- [10] Shinsuke Matsumoto and Masahide Nakamura, "A Preliminary Study of Size Optimization for Text-Based Web-Resource," In International Workshop on Empirical Software Engineering in Practice (IWESEP), 2016.
- [11] Yasutaka Sakamoto, Shinsuke Matsumoto, Seiki Tokunaga, Sachio Saiki, and Masahide Nakamura, "Empirical Study on Effects of Script Minification and Http Compression for Traffic Reduction," In The Third International Conference on Digital Information, Networking, and Wireless Communications (DINWC), pages 127-132, 2015.

#### [図書] (計 0 件)

#### [産業財産権] (計 0 件)

#### [その他]

エラートレンド収集の試作ツール  
<https://tyr.ics.es.osaka-u.ac.jp/erepo/>

#### 6. 研究組織

##### (1) 研究代表者

榎本 真佑 (Matsumoto, Shinsuke)  
 大阪大学・情報科学研究科・助教  
 研究者番号: 90583948