

既存 Web ページの利用を活発にするための Web ページ再活性化システム

土井 達也[†] 湯浅 康史^{††} 白松 俊[†] 大園 忠親[†] 新谷 虎松[†]
 名古屋工業大学大学院工学研究科情報工学専攻[†] 名古屋工業大学工学部情報工学科^{††}

1 はじめに

本研究では、Web ブラウザからの編集が不可能な既存の Web ページを Web サーバの設定変更無しに Web ブラウザ上から編集可能にすることを目的とする。Web ブラウザから編集できることにより複数ユーザでリアルタイムに情報共有を行うことで Web ページの利用が活発な状態となる。本稿では、既存の Web ページを Web ブラウザから編集可能にすることを Web ページの再活性化と呼ぶ。Web サーバに CGI 等を導入せずにコンテンツの編集結果を反映させることができる Web ページ再活性化システムを提案する。

複数ユーザでのリアルタイムの情報共有は作業効率が向上し生産性が高まる。Web ページをリアルタイムに編集を行ったり、離れた場所にいるユーザに対し同じコンテンツを閲覧させたり [1] するシステムが提案されている。Web 上での情報共有ではコンテンツの同期とシステムの導入の容易性が重要となる。複数ユーザが Web ページの編集を行うには、編集のタイミングが重なった場合の競合の問題を解決する必要がある。

Web ページを Web ブラウザから編集可能にする既存のシステムとして我々は Web サーバに CGI を設置する必要があるものを提案している [2]。CGI を動作させるためには Apache の httpd.conf などの Web サーバの設定ファイルを書き換える作業が必要である。また、レンタルサーバでは CGI の実行を許可していないものもあり、既存のシステムの導入は困難、もしくは不可能な場合がある。既存の Web ページを、別のサイトに Web ページを移転することにより、Web ブラウザからの Web ページ編集の実現は可能であるが URL が変わってしまう。検索サイトへの影響を考えれば異なるドメインの Web ページが Web サイトに混ざるのは避けたほうが良く、ユーザにも混乱を与えずに済む。

本稿では Web サーバとコンテンツ管理サーバを分割することで、Web サーバ上に CGI 等を導入せずに運用できる Web ページ再活性化システムを試作した。本システムでは HTML の script タグとして記述された専用のタグを既存の Web ページに埋め込むことで容易に編集可能状態にする。専用のタグによって対象 Web ページの編集する内容を Web サーバとは別のコンテンツ管理サーバで管理することを実現している。また Web ページ再活性化を行った後でも URL が変わらないような SEO 対策を解決した。

2 Web ページ再活性化

本研究の目的は Web ブラウザからの編集が不可能な既存の Web ページを Web サーバの設定変更無しに Web ブラウザ上から編集可能にすることである。既存の Web ページを Web ブラウザから編集可能にすることを Web ページの再活性化と呼ぶ。本研究で再活性化の対象とする Web ページは、wiki やブログなどの CMS がすでに導入されているものではなく、Web ブラウザ上から編集を行うことができない Web ページである。既存の Web ページとコンテンツの編集結果を管理

[†] A Web Page Resurrection System based on Realtime Synchronization Mechanism

Tatsuya DOI, Yasufumi YUASA, Shun SHIRAMATSU, Tadachika OZONO, and Toramatsu SHINTANI

Dept. of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology. Dept. of Computer Science and Engineering, Nagoya Institute of Technology.

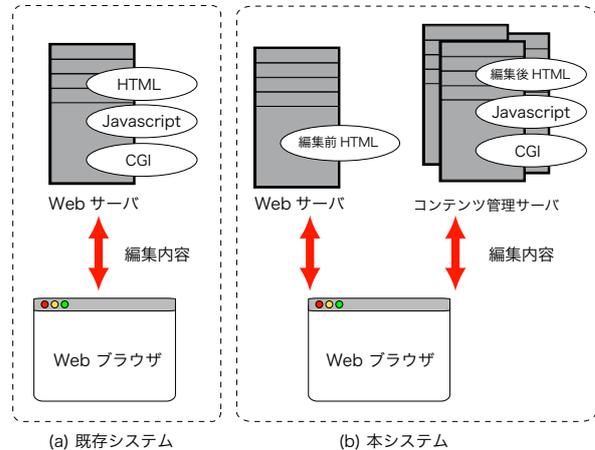


図 1: 既存システムと本システム

する機構は、同一 Web サーバ上に設置せず、Web サーバとコンテンツ管理サーバの構成である。そのため、Web サーバとの通信が必要な Ajax や複数の Web ページが必要な frame などを利用している Web ページは Web ページ再活性化の対象外とする。Web ページ再活性化を行う際には URL が変化しないようにすることを前提とする。

3 Web ページ再活性化システム

既存のシステムはページ遷移なしに Web ブラウザ上から Web ページを編集可能としている。Web サーバに専用の CGI を設置することで Web ページを編集し、その編集結果を同じページを見ているユーザにも反映させることを実現している。本システムと我々の既存のシステムとの違いを図 1 に示す。既存のシステムでは図 1 (a) のように、Web サーバには編集可能にする対象の HTML ファイル、HTML ファイルを書き換える CGI、Web ページの編集箇所を送受信する JavaScript ファイルが必要となりコストがかかる。

そこで、本稿では図 1 (b) のサーバ構成で Web ページを編集可能にすることを提案する。Web サーバ上にある既存の Web ページはそのまま利用する。Web ページを編集するために必要な CGI や JavaScript は Web サーバとは別のコンテンツ管理サーバに設置する。Web ページを再活性化するためには既存の Web ページに HTML の script タグとして記述された専用のタグを埋め込む。専用のタグによって Web ページの編集内容を Web サーバとは別のコンテンツ管理サーバで管理することを実現している。これにより Web サーバの CGI の設定を気にせず、Web ページ再活性化システムを導入することが可能となる。

3.1 システム構成

本システムの構成を図 2 に示す。本システムは既存の Web ページが置かれている Web サーバと編集結果を管理するコンテンツ管理サーバからなる。再活性化が行われた Web ページをユーザが閲覧する場合は、既存の Web を用いて再活性化前と同様に Web ページにアクセスする。既存の Web ページには HTML の script タグを埋め込む。埋め込むタグは Web ページ再活性化を行いたい各 Web ページで共通のものを使

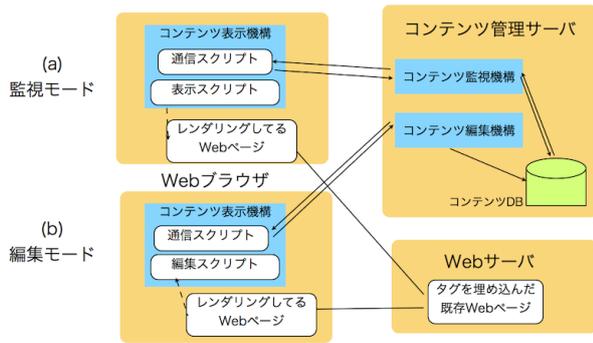


図 2: システム構成図

用する。Web ページ再活性化に必要なスクリプトはコンテンツ管理サーバからダウンロードする。Web ブラウザによって Web サーバ上の既存 Web ページにアクセスする。既存の Web ページにはコンテンツ管理サーバへアクセスするためのスクリプトが埋め込まれているため、コンテンツ管理サーバへアクセスを行う。コンテンツ管理サーバからコンテンツを表示するためのスクリプトがダウンロードされる。アクセス元の URL から固有の ID を特定し、再活性化に必要なスクリプトが出力される。

3.2 コンテンツの同期

コンテンツの同期は、大きく分けて 2 つのパートに分かれる。コンテンツの編集を行う編集モードと、編集されたコンテンツを見ている Web ページに反映させる監視モードにである。編集結果のコンテンツは DOM 要素の単位で同期を行う。対象とする DOM 要素はブロック要素とインライン要素の一部である。編集結果のコンテンツはコンテンツデータベースに key - value ストアとして保存されている。保存する内容は編集結果のコンテンツの DOM 要素の XPath と、コンテンツの内容、そして更新日時である。コンテンツの内容は、DOM 要素の innerHTML とレンダリングに関わるスタイル情報を含む。また、DOM 要素は半構造化されているため、情報をフラット化し保存に適した形に変換を行う。

監視モードのシステムの流れを図 2 の (a) に示す。閲覧している最中に Web ページのコンテンツが編集された場合は、コンテンツ管理サーバのコンテンツ監視機構がコンテンツデータベースから編集された情報を取得し、Web ブラウザのコンテンツ表示機構にその情報を渡す。コンテンツ表示機構の通信スクリプトがコンテンツ管理サーバからの情報を解析し、書き換えスクリプトによってレンダリングしている Web ページの一部を書き換える。

編集モードのシステムの流れを図 2 の (b) に示す。Web ページの編集を行う際は、コンテンツ管理サーバのコンテンツ監視機構との通信を一旦停止する。そして、編集を行った Web ページのコンテンツを書き換えスクリプトによりフラット化する。通信スクリプトはフラット化されたコンテンツの編集結果をコンテンツ管理サーバに送信する。送信後は再びコンテンツ管理サーバとの通信を開始し、他のユーザがコンテンツを編集した情報を待つ。コンテンツ編集機構は Web ブラウザからのフラット化された情報を更新日時とともにコンテンツデータベースに保存を行う。

コンテンツの同期は DOM 要素単位で行うため、複数ユーザが同時に Web ページの異なる場所のコンテンツを更新した場合であっても競合せずに編集結果の反映を行うことが可能となる。同期を行うコンテンツの DOM 要素の粒度により、通信時間や同期のコストが異なる。本稿ではコンテンツの粒度を、編集された部分に一番近いブロック要素またはインラ

イン要素の一部となる親要素としている。

3.3 SEO 対策

本システムを導入する場合、元の Web サーバには古い Web ページが残されている。これは検索エンジンのクローラに編集前のページがそのまま収集されることとなる。この時、編集された最新の状態の Web ページは検索サイトからされなくなる。そこで、編集された Web ページの結果を書き出す機能が実装されている。書き出された編集結果の HTML ソースを Web サーバの HTML ファイルに置き換えることで、最新のコンテンツを検索サイトから発見してもらえるようにする。Web サーバの HTML ファイルの書き換えは手動で Web ページを管理しているユーザが行う。

また既存の Web ページとコンテンツ管理サーバとの通信で得られる Web ページの通信量が同じ場合、既存の Web ページを閲覧するときの 2 倍程度の時間がかかることとなる。これはユーザにとって Web ページの閲覧性を低下させることにつながり不都合である。そのため、既存の Web ページを表示した後は徐々に最新の編集結果を反映させるモードを追加した。このモードでは、ユーザは既存の Web ページを見ている最中にコンテンツ管理サーバから得られた最新の編集結果が徐々に反映されていく様子を見ることとなる。コンテンツ管理サーバで最近まで編集された Web ページの結果を書き出し、その情報が既存 Web ページに反映させていることが前提となる。SEO 対策を行わない場合は、既存の Web ページの情報量を最低限のものとすることで、最新の編集状況が反映されている Web ページが Web ブラウザで表示される時間を短縮できる。

4 まとめ

ページビューを増加という Web ページの利用の活発化のために、Web ページ再活性化を提案した。Web ページ再活性化によって、既存の Web ページは Web ブラウザから編集可能状態になる。Web ページ再活性化システムは Web サーバとは別のサーバとしてコンテンツ管理サーバを用いることで、編集結果の履歴などの管理も可能としている。編集結果の情報の粒度を調節することにより、複数人で Web ページの編集を行った場合でも競合が起こらない。本システムを利用することで、既存の Web ページを協調ワークスペース [3] として利用するが可能となる。今後、Web ページ再活性化を行うことにより、既存の Web ページに協調作業を行う機能の追加が容易となると期待できる。

参考文献

- [1] Dietwig Lowet, Daniel Goergen: "Co-Browsing Dynamic Web Pages", 18th International World Wide Web Conference (WWW2009), pp.941-950, Apr. 2009.
- [2] 西 健太郎, 新谷 虎松, 松尾 徳朗, 田代 慎治, 伊藤 孝行: "既存 Web ブラウザを利用したオンライン編集可能な Web ページの実現", 電気学会論文誌 (部門誌)C. vol.125, no.4, pp.660-665, Apr. 2005.
- [3] 湯浅 康史, 土井 達也, 白松 俊, 大園 忠親, 新谷 虎松: "Web ページ再活性化システムを利用した Web ページ上での協調ワークスペース", 第 73 回情報処理学会全国大会, Mar, 2011 (掲載予定).