

Web Index における配信型コンテンツを利用した自動ライブラリ更新システムの提案

大島 拓也[†] 遠山 元道[†]

[†] 慶應義塾大学理工学部情報工学科 〒 223-8522 神奈川県横浜市港北区日吉 3-14-1

E-mail: [†]nanadama@db.ics.keio.ac.jp, ^{††}toyama@ics.keio.ac.jp

あらまし Web Index(WIX) とは、単語 (keyword) とそれに関連する Web 文書の URL(target) の組み合わせであるエントリの集合 (WIX ファイル) を閲覧中の Web 文書に対して結合 (アタッチ) することによって Web 文書中のキーワードに対して指定された URL へのハイパーリンクを生成し、Web 情報資源の結合を実現するシステムである。本論文では、WIX システムにおいて、RSS のような複数の配信源より配信されるコンテンツを利用し、WIX ファイルのライブラリを追記的に自動更新するシステムについて提案する。

キーワード Web Index, Web 情報システム, RSS, XML

1. はじめに

近年、インターネットの普及により Web 上での情報検索のニーズが著しく増加し、ユーザは検索エンジンを使用することで必要な情報を得ることが可能となった。しかし、検索エンジンを用いた情報検索は、検索結果として得られた Web ページ内の単語の詳細情報を得たい場合には、新たに検索エンジンで再検索し、選択しなければならない。Web ページ作成者はハイパーリンクによって Web ページに関連する複数の情報を予め結合することが可能だが、これは Web ページ作成者が意図した関連のみで結合され、必ずしも閲覧するユーザのニーズに応えているとは限らない。そこで、著者は Web における利用者主導による情報資源結合を実現するために、Web Index(以下、WIX と呼ぶ) システムという情報資源表現形式の提案、開発を行っている [1, 2]。

キーワードと URL の組み合わせである、エントリの集合を XML 形式で記述したものを WIX ファイルという。WIX システムでは、WIX ファイルを用い、閲覧中の Web ページに結合 (アタッチ) をすることで、Web ページ内の文章に出現するキーワードをそれに対応する URL へのハイパーリンクに変換する。現在の Web 形態では、Web ページ作成者によって特定のアンカーテキストから特定のページへのリンクが関連付けられるという構造が一般的である。Web ページ内のリンクは常に既存の Web ページへしかリンクすることができず、その Web ページ作成後に作成される新たな Web ページへのリンクを作成することはできない。WIX では、アンカーテキストとリンクを Web ページから独立した「キーワードとリンク先の集合」として扱い、任意のドキュメントに対して、Web ブラウジングを行っているユーザ主導で「結合」することでドキュメント内のキーワードに対応する URL のハイパーリンクに自動的に変換することができる。WIX システムの利用により、Web ページ作成の時系列という壁を越え、古い Web ページからより新しい Web ページへアクセスすることが可能となる。

本論文では、ニュースサイトなどの日々コンテンツを追加、更新するサイトが配信している RSS を利用し、WIX システムのユーザが常に最新のコンテンツにアクセスすることを可能にする WIX ファイルの提供、更新、メンテナンスを行う WIX ファイルの更新システムについて提案するものである。

本論文の構成は以下の通りである。まず、2 章において本論文で説明する研究の目的を示す。3 章では、WIX システムの概要について述べる。4 章ではシステムの構成および機能について述べる。5 章では評価を示す。6 章で結論を述べる。

2. 研究の目的

現在提供されている WIX ファイルについて、その作成方法には、以下に示すような方法がある。

- ターゲットとなる URL と、対応付けられるキーワードとを手動で記述する方法
- ターゲットとなる URL を、対応付けられるキーワードをもとにして作成する方法

Wikipedia の見出し語をキーワード、その見出し語に該当する Wikipedia の記事ページの URL をターゲットとしているエントリの集められた Wikipedia の WIX ファイルや、英単語をキーワードとして、該当する Web 英和辞典の記事ページの URL をターゲットとしているエントリの集められた英和辞典の WIX ファイルなどが実際に生成されている。

- リンクの集合をひとまとめにしている Web ページを利用し、キーワードとターゲットとなる URL をまとめて取得して作成したもの

有名人の Ameba ブログのリンクをまとめた Web ページや、日本の有名企業の公式サイトのホームページのリンクをまとめた Web ページを利用した WIX ファイルが実際に生成されている。[3-5]。

Web に存在する情報資源のうち、ニュースサイトの記事、企業や団体のプレスリリース、個人が開設するブログの記事など、時系列で比較的頻繁に情報が記録、更新されていくコンテ

ンツに対する、Web ブラウジングを行うユーザの関心は年々高まっている。現在の WIX システムで提供されているこれらの WIX ファイルでは、このような、日々追加されていくコンテンツに直接アクセスすることができない。WIX システムの特徴である、古い Web ページの閲覧中に新しい Web ページへアクセスすることが可能であるという点をより有効に活用するには、このようなコンテンツを含む Web ページへのアクセスを WIX システムにおいて可能にする必要がある。本研究の目的は、WIX ファイルの自動生成、自動更新システムを構築し、WIX システムを利用して、日々追加されていくコンテンツを含む Web ページへのアクセスを可能にすることである。

3. Web Index システム

3.1 システムの概要

Web Index (WIX) とは、Web における結合可能な情報資源である。現在の Web では、特定のアンカーテキストから特定のページへのリンクがされるという構造が一般的である。これに対し、WIX では、アンカーテキストとリンクを Web ページから独立した「キーワードとリンク先の集合」(WIX ファイル)として扱い、任意のドキュメントに対しユーザ主導で「結合」を行うことでドキュメント内のキーワードを対応する URL へのハイパーリンクに変換する。

3.2 WIX ファイル

```
<?xml version="1.0" encoding="UTF-8"?>
<WIX>
  <header />
  <body>
    <entry>
      <keyword>錦織圭</keyword>
      <target>http://ja.wikipedia.org/wiki/錦織圭</target>
    </entry>
    <entry>
      <keyword>ノバク・ジョコビッチ</keyword>
      <target>http://ja.wikipedia.org/wiki/ノバク・ジョコビッチ</target>
    </entry>
    <entry>
      <keyword>ロジャー・フェデラー</keyword>
      <target>http://ja.wikipedia.org/wiki/ロジャー・フェデラー</target>
    </entry>
  </body>
</WIX>
```

図 1 WIX ファイルの例

WIX ファイルとは、XML 形式で記述されたキーワードと URL の組み合わせであるエントリの集合であり、WIX システムにおいて、アタッチ対象とリンク先を決定する元として用いられるコンテンツの 1 つである。記述例は図 1 のようになる。エントリ内には、キーワードとなる見出し語が keyword 要素として格納され、それに対応する参照先の URL が target 要素に格納される。header 要素には、WIX ファイルのファイル概要、カテゴリー、作者コメントなどのメタ情報を格納する。

3.3 アーキテクチャ

3.3.1 WIX ライブラリ

WIX ライブラリでは、全ての WIX ファイルを、過去のバージョンも含めて保存しており、ファイル単位での情報管理を行っている。アタッチの際には、アタッチエンジンが全ての WIX

表 1 entry テーブル

wid	eid	keyword	target
1	1	錦織圭	http://ja.wikipedia.org/wiki/錦織圭
1	2	フェデラー	http://ja.wikipedia.org/wiki/ロジャー・フェデラー
2	3	田中将大	http://ameblo.jp/tanaka-masahiro
3	6	前田健太	http://www.carp.co.jp/team14/member/p18.html

ファイルのエントリに対して辞書式マッチングを行うため、WIX ファイルをエントリ単位に分解し、WIX データベースに格納する。

3.3.2 WIX データベース

WIX データベースでは、WIX ファイルをエントリ単位に分割し、関係データベースにて管理する。WIX ファイルの情報は、entry テーブルに格納され、管理される。entry テーブルには、各エントリが所属する WIX ファイルの id (wid)、エントリの id (eid)、辞書語となる keyword と、それに対応する URL の target が属性として存在する。entry テーブルの例を表 1 に示す。

3.3.3 Find インデックス

Find インデックスでは、WIX データベースの entry テーブルからエントリ情報をメモリ上に展開し、高速なアタッチを実現する。WIX システムでは、Aho-Corasick 法に基づくオートマトンを構築し、辞書式マッチングを行う [6]。

3.4 ハイパーリンクの生成

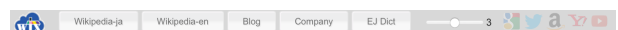


図 2 WIX ツールバーの例 (Google Chrome Extension)

ユーザは、WIX ファイルの集合をブックマークとして記録しておき、Web ページを閲覧中に、アタッチボタンをクリックすることで、Web ページ中の単語をハイパーリンクに変換することができる。この結合操作をアタッチと呼ぶ。WIX システムのクライアントサイドは、Firefox add-on や Google Chrome Extension によって実装されている。図 2 は Chrome

ソフトバンクから3人選出 ゴールデングラブ賞発表

2015/11/10 15:16

小中大 保存 印刷 リプリント 共有

プロ野球の守備のベストナインを選ぶ三井ゴールデングラブ賞が10日発表され、球団初の2年連続日本一に輝いたソフトバンクから松田宣浩三塁手、今宮健太遊撃手、柳田悠岐外野手の3人が選ばれた。松田はパ・リーグ最多の197票を集めて3年連続4度目の受賞となった。

14年ぶりにセ・リーグを制したヤクルトからは中村悠平捕手、川端慎吾三塁手が初受賞し、畠山和洋一塁手も選ばれた。中村は227票で両リーグ最多得票。初受賞は中村、川端に加え、日本ハムの中田翔一塁手、ロッテのクルーズ二塁手、清田育宏外野手の5選手。阪神の福岡孝介外野手は米大リーグから復帰後初めて9年ぶり5度目の受賞。

選考は全国の新聞、通信、放送各社のプロ野球担当者の投票で行われ、有効投票数はセが248、パが220だった。(共同)



ゴールデングラブ賞に選出された中田 (10日、天海) =共同

図 3 アタッチ前の Web ページ

Extension の一例である。図 3 の記事閲覧中に、アタッチボタンをクリックすることで、Web ページに含まれるキーワードのうち、WIX ファイルのエントリにキーワードとして含

れているものがあれば、WIX ファイルの target に記述された URL へのハイパーリンクが生成され、図 4 のようになる。



図 4 アタッチ後の Web ページ

4. 自動 WIX ライブラリ更新システム

4.1 システムの概要

自動 WIX ライブラリ更新システムは、WIX システムのユーザが、ユーザ自身の興味に応じて、キーワードに関連するコンテンツに容易にアクセスできるようにすることを目的とした、WIX ファイルの自動更新・管理システムである。本システムによって提供される WIX ファイルは、最新のコンテンツへの URL を含むものとなっており、ユーザが該当する WIX ファイルへのブックマークを含むアタッチボタンをクリックすることで、閲覧中の Web ページの単語に対し、それに関連する最新のコンテンツへのリンクが作成される。ユーザはこのリンクをクリックすることで該当するコンテンツにアクセスすることができる。

最新のコンテンツを取得するために、本システムは RSS を利用している。RSS とは、コンテンツを提供する情報提供者が、更新情報をひとまとめにして利用しやすい形式で配信する XML 形式の文書フォーマットである [7]。RSS 2.0 では、Web ページの更新情報をアイテムという単位で扱っている。アイテムに含まれる要素は、表 2 で示している。

表 2 RSS のアイテムに含まれる要素

要素名	説明
<title>	記事のタイトル
<link>	記事のリンク URL
<pubDate>	記事の更新日時

RSS を利用して最新のコンテンツを取得する方法は、直接 Web サイトをクローリングして最新のコンテンツを取得する方法より、以下の二点で優れている。一点目は、情報提供者がコンテンツとして配信している記事の含まれるページのみのリンクを、Web サイトの構造に依存せず取得できることである。直接 Web サイトをクローリングしてコンテンツを取得する場合は、サイトごとに異なる Web ページの構造を分析し、CSS セレクタなどを用いて適切に必要なリンクのみを取得する必要がある。一方、RSS を利用したコンテンツ取得では、そのよう

な Web ページの構造を分析して取得方法を指定する必要がない。二点目は、Web ページの更新日時の取得が簡単にできることである。ニュースサイトなどの Web サイトでは、ブラウザのリクエストに対し、PHP などのサーバサイドスクリプトを利用して HTML を動的に生成しているものが多い。このような Web ページの更新日時は、記事そのものの更新日時ではなく、サーバサイドスクリプトの実行日時になっており、記事の更新日時の取得には利用できない。一方、RSS では、その多くが表 2 に示すような更新日時を含んでおり、その取得は容易である。記事の更新日時は、古くなったコンテンツを、エントリ生成の対象から除外する処理において重要である。

本システムでは、以下の 3 ステップによって WIX ファイルの自動更新を行っている。

- (1) キーワード候補の登録
- (2) RSS ファイルの登録と RSS アイテムの抽出
- (3) キーワード候補単語と RSS アイテムからエントリ生成システムの概要とデータの入出力を図 5 に示す。

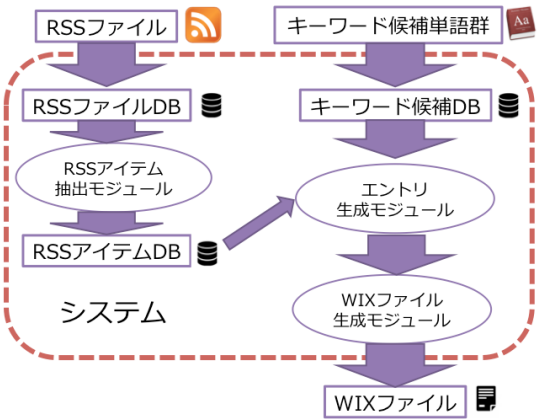


図 5 システム概要図

4.2 キーワード候補の登録

エントリのキーワードの候補となる単語をキーワード候補としてデータベースに登録する。キーワード候補の単語群には、助詞や代名詞などのストップワードが含まれないようにする必要がある。本システムでは、キーワード候補の単語群として、日本語版 Wikipedia の見出し語約 200 万語を取得し、ここから「ここ」「私」などのストップワードや、数字のみの見出し語、日付の見出し語などを取り除いて登録している。

4.3 RSS ファイルの登録と RSS アイテムの抽出

本システムでは、コンテンツの提供者が提供する、RSS フィーダの URL を登録する必要がある。RSS フィーダを登録すると、RSS アイテム抽出モジュールで定期的に RSS フィーダより配信される RSS ファイルの更新を確認する。RSS ファイルに更新があった時、追加された RSS アイテムを取得し、データベースに登録する。また、RSS アイテム抽出モジュールでは、古くなった RSS アイテムの削除も行う。RSS アイテムが古くなったと判定する基準は二つある。一つは RSS アイテムの示すリ

表 3 拡張された entry テーブル

wid	eid	keyword	target	title
4	1	錦織	http://www.nikkei.com/article/XXXXXX	錦織がベスト8入り...
4	2	錦織	http://news.tennis365.net/news/XXXXXX	錦織、アウェーに屈する
4	3	清田	http://www.nikkansports.com/XXXXXX	ロッテ清田&藤岡の東洋大コンビ...
4	4	清田	http://headlines.yahoo.co.jp/XXXXXX	ロッテ清田、東洋大の後輩・藤岡に...

リンクがリンク切れとなっている場合である。もう一つは RSS アイテムの示す記事の更新日時から、一定時間以上経過していた場合である。いずれの場合も、古くなった RSS アイテムは削除され、エントリの生成の対象外となる。

4.4 エントリ生成と WIX ファイルの更新

RSS アイテム抽出モジュールにおいて、RSS アイテムの更新が行われると、続いてエントリ生成と WIX ファイルの更新を行う。エントリ生成モジュールでは、データベースに登録されている全ての RSS アイテムに対し、各 RSS アイテムに含まれている記事のタイトルに、同じくデータベースに登録されているキーワード候補の単語が含まれているかどうかを調べる。タイトルにキーワード候補の単語が含まれていれば、含まれていた単語をキーワード、RSS アイテム中の Web ページの URL をターゲットとしたエントリを生成する。この操作をマッチングと呼ぶ。本システムでは、大量のキーワード候補からトライ木を生成し、それを利用してマッチングを行った。一つの RSS アイテムに対応する単語が複数ある場合は、それぞれの単語をキーワードとしたエントリを生成する。全ての RSS アイテムに対し、対応するエントリの生成が終わったら、エントリ全てを含む WIX ファイルを新たに作成する。RSS アイテム抽出モジュールで削除された古い RSS アイテムを含むエントリが残らないように、WIX ファイルの更新は追記型ではなく、完全な上書き型で行う。

本システムでは、同一のキーワードに対し複数の URL が対応するようなエントリが存在する。このような場合には、Web ページ中のキーワードを直接ハイパーリンクに変換するのではなく、ポップアップの形で複数のリンクをまとめて表示し、ユーザがユーザ自身の興味に応じてリンクを選択できるようにしている。図 1 で示したように、WIX ファイルに含まれる情報は、キーワードとリンク先となる URL に限られているが、これらの情報だけではユーザがリンクを選択する上で不十分である。そこで、本システムでは、表 3 のように WIX ファイルを拡張し、各エントリごとにエントリのタイトルを設定するようにした。

WIX ファイルを拡張し、各エントリにタイトルを設定することで、図 6 に示すように、ポップアップによって表示された複数のリンクは、リンク先のページのタイトルが表示され、一目でどのようなページにリンクするかわかるようになった。

更新された WIX ファイルは、3 章で説明した、他の WIX ファイルと同様に、WIX DB に登録され、Find インデックスを構築して、アタッチ可能な状態となる。

手も選ばれた。中

、川端に加え、且

手、[The links of 清田育宏](#)

グから

ロッテ清田&藤岡の東洋大コンビ爆笑トークショー(日刊スポーツ)
前園真聖「hitomi似美人」と半同棲生活!で、結婚は!? (friday)
【ロッテ】清田、東洋大の後輩・藤岡に「もっと活躍すると思った」(スポーツ報知)
五郎丸は?プロ契約選手は?ラグビー選手の気になる年収を調べてみた! (friday)

図 6 ポップアップされた複数のリンク

5. 評価

5.1 評価の方法

評価に当たって、Yahoo!JAPAN ニュースで提供されている RSS ファイル 443 個を登録した。本システムで RSS アイテムの取得を行ったところ、2015 年 12 月 28 日 20 時から、30 日 20 時までの 48 時間に、9,879 個の RSS アイテムを取得できた。

本評価では、エントリ生成のためのマッチング処理における、探索範囲の違いによるリンクの変化を比較するために、以下に示す 2 種類の探索範囲を用意した。

- RSS アイテムに含まれる、記事のタイトルのみ。
- 記事のタイトルに加え、RSS アイテムのリンク先の Web ページを取得し、その本文全体。

Web ページ本文の取得にあたり、HTML のスクレイピングに特化した Java のライブラリである、Jsoup を利用した。Yahoo!JAPAN ニュースで提供されている Web ページからは、広告やナビゲーションなどの、コンテンツとは関係のない部分を取り除いた部分を、CSS セレクタで "div#main" と指定することで取得できる。このようにして取得した Web ページの一部を、以下 Web ページの本文と呼ぶ。取得したデータは、データベースの rssitem テーブルに、text 型の content 要素として格納した。本文を取得する場合は、RSS アイテム一つにつき一回の HTTP リクエストとレスポンスとが必要になるため、RSS のみを取得して解析するのに比べ、RSS アイテムの取得にかかる時間は非常に大きくなる。

5.2 特定のキーワードについての結果と考察

探索範囲にタイトルのみを利用した場合と、本文全体を利用した場合のそれぞれについて、幾つかの単語に対し生成されたエントリ数を表 4 に示す。

探索範囲がタイトルの場合、"ソフトバンク"、"ヤクルト"のような、複数の記事ページで多く見られるキーワードについては、図 7 のように十分なエントリ数が得られた。一方、"Linux"、"川端慎吾"のような、限られた記事ページでしか取り上げられないようなキーワードについては、エントリを得ることはできなかった。

探索範囲を本文全体とした場合、複数の記事ページで多く見られるキーワードについては、非常に多くのリンクが得られ

表 4 探索範囲別のキーワードごとのエントリ数

キーワード	タイトルのみ	本文全体
朝日新聞デジタル	118	124
野球	28	233
総理	31	206
ソフトバンク	15	114
ヤクルト	6	48
フェデラー	3	22
Linux	0	17
川端慎吾	0	2

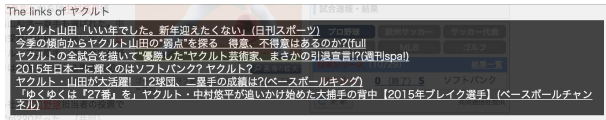


図 7 探索範囲がタイトルをみの場合の，“ヤクルト”でのアタッチ結果

る．“ヤクルト”の場合のアタッチ結果を図 8 に示す．このような場合は，ユーザが多くのリンクの中からユーザ自身の興味のあるページを選択するのはやや負担となると考えられる．一方，“Linux”，“川端慎吾”のような，限られた記事ページでしか取り上げられないようなキーワードについても，図 9 に示すようにエントリが 2 つ得られた．

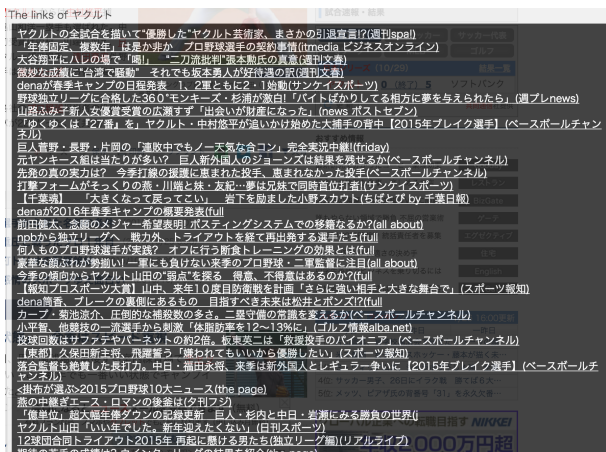


図 8 探索範囲が本文全体の場合の，“ヤクルト”でのアタッチ結果

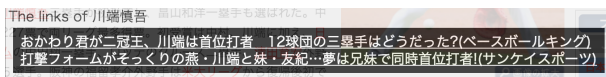


図 9 探索範囲が本文全体の場合の，“川端慎吾”でのアタッチ結果

RSS アイテムのタイトルには，その RSS を提供している Web サイトの名前が含まれているものも多い．“朝日新聞デジタル”というキーワードはその一例で，Web ページの内容そのものとは関連のない Web ページまで対応付けられてしまう場合が多い．

タイトルのみの探索範囲で十分に関連するコンテンツへのリンクが得られる場合と，本文全体まで探索してようやく関連するコンテンツへのリンクが得られる場合とがあると考えられる．

5.3 全体的なキーワードについての結果と考察

探索範囲を，Web ページのタイトルのみとした時と，Web ページの本文全体とした時の，エントリの生成数を比較した結果を図 5 に示す．

表 5 探索範囲，キーワード数とエントリ数の関係

探索範囲	エントリ数	キーワード数	平均エントリ数 *1	最大エントリ数 *2
タイトルのみ	45,266	10,888	4.16	45
本文全体	3,276,577	59,903	54.70	238

*1 エントリ数 / キーワード数．一つの単語あたりの平均エントリ数．*2 一つの単語に対し生成された最大エントリ数．

ハイパーリンクの数がおおよそ 20 個以下であるようなキーワードに対しては，WIX による結合が行われたことで，ユーザが興味のある Web ページに容易にアクセスすることができるようになると考えられる．ただし，探索範囲がタイトルのみであるうと，本文全体を含んでいようと，生成されたエントリ数が大きくなっている要因として，“野球”，“総理”など，ユーザがあまり注目しないほど一般的なキーワードを含むエントリを生成しているという点があげられる．これらのキーワードに関してはユーザの注目の対象になることが少ないため，エントリがいくら生成されても，ユーザが WIX システムを利用する上であまり妨げになるような要素はないと考えられる．一つのキーワードに対し，多数のエントリが生成される場合には，エントリの生成上限数を指定し，更新日時の新しいコンテンツへのリンクを含むエントリのみを生成することや，あらかじめ 2 つの探索範囲でエントリ生成を行い，キーワードごとに生成されたエントリ数に応じて，最終的な WIX ファイルに記述するエントリを変更することによって，このような問題を解消することが可能であると考えられる．

6. おわりに

6.1 結論

本研究では，ニュースサイトやブログの記事，企業や団体の公式サイトで配信されるプレスリリースなどのような，時系列で比較的頻繁に情報が記録，更新されていくコンテンツにアクセスすることを可能にするための WIX ファイルの自動生成，自動更新及び管理システムを提案し，実装した．実装にあたっては，Web ページの更新情報の配信に幅広く使われている RSS を利用するシステムを構築し，Web ページの構造を解析して情報抽出するよりも簡単な更新情報の取得を可能にした．本研究で提案するシステムの利用により，WIX ファイルのライブラリの運用者の，WIX ファイルの作成の負担を軽減することが可能となる．

6.2 今後の課題

マッチングに用いるキーワードを，あらかじめ抽出しておいた単語群に限定するのではなく，タイトルや本文などの入力から抽出する方法がある．MeCab [8] のような形態素解析エンジンを利用して，新しい単語をキーワードとして抽出する方法が

考えられる．

Web ページの本文全体を抽出するにあたり，Web ページから広告やナビゲーションなどのノイズとなりうる要素を排除してコンテンツ部分のみを抽出する必要がある．HTML 文書から部分文書検索を行う技術 [9] の本システムへの実装が考えられる．

文 献

- [1] 林昌弘，青山峻，朱成敏，遠山元道. KeioWIX システム (1) ユーザインターフェース. データ工学ワークショップ，DEIM2011. 2011.
- [2] 森良介，藪達也，朱成敏，遠山元道. Keio WIX システム (2) サーバサイド実装. データ工学ワークショップ，DEIM2011. 2011.
- [3] 藤井洋太郎，遠山元道．WIX システムにおけるコンテンツ作成支援システム．DEIM 2012．2012．
- [4] 市東隼，分部亮太，朱成敏，遠山元道．Keio WIX システム (3) コンテンツ作成．DEIM2011．2011．
- [5] 金岡慧，遠山元道．自動更新型 WIX ファイル生成システムおよび Deep Web に対するアタッチ機構の構築．DEIM2014．2014．
- [6] 石崎文規，遠山元道. 大規模 Aho-Corasick オートマトンにおける追加更新手法の提案. データ工学ワークショップ，DEIM2012. 2012.
- [7] RSS Advisory Board, "RSS 2.0 Specification", <http://www.rssboard.org/rss-specification>, 2016 年 1 月 15 日アクセス．
- [8] MeCab <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html?sess=3f6a4f9896295ef2480fa2482de521f6>．
- [9] 樫 惇志，宮崎 純，波多野 賢治，山本 豪志朗，武富 貴史，加藤 博一．XML 部分文書検索技術の Web 文書への適用．DEIM Forum 2014, 2014.
- [10] 吉永 光，古川 孝則，山本 博章，金子 美博．トライ木を用いた XML キーワード検索法，DEIM Forum 2014, 2014.
- [11] 古川 孝則，滝澤 健斗，山本 博章，金子 美博．トライ木を用いた XML キーワード検索法，DEIM Forum 2015, 2015.
- [12] Y. Xu, Y. Papakonstantinou. Efficient Keyword Search for Smallest LCAs in XML Database, Proc. of SIGMOD, pp.537-538, 2005.