

Keio WIX システム (1) ユーザーインターフェース

林 昌弘[†] 青山 峻[†] 朱 成敏[†] 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: [†]{hayashi,aoyama,joo}@db.ics.keio.ac.jp, ^{††}toyama@ics.keio.ac.jp

あらまし 著者らはユーザ主体の Web 情報資源結合を実現する Web Index(WIX) システムを開発している。WIX システムは、キーワードとそれに対応する URL からなるエントリを持つ WIX ファイルを用いることで、ユーザが Web 文書を閲覧している際、多くのキーワードに対して目的のコンテンツを持つ Web 文書に即座にアクセスすることができる。ユーザは好みの WIX ファイルをブックマークしてウェブブラウザ上のツールバーから WIX システムを利用する。本研究では、WIX システムのクライアントサイド実装と、ユーザが WIX ファイルをブックマークする際の支援システムの提案、実装を行った。

キーワード WIX, Web 情報システム

Keio WIX system(1) User Interface

Masahiro HAYASHI[†], Shun AOYAMA[†], Sungmin JOO[†], and Motomichi TOYAMA^{††}

^{††}Department of Information and Computer Science,
Keio University

Hiyoshi 3-14-1, Kouhoku-ku, Yokohama-shi, Kanagawa, 223-8522 Japan

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

1. はじめに

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、著者らは Web Index (以下 WIX とする) と呼ぶ情報資源表現形式の提案、開発を行っている。WIX ファイルとは、キーワードとそれに対応する URL をペアとするエントリの集合であり、XML 形式で記述される。それを閲覧中の Web 文書に結合 (アタッチ) することで、Web 文書中のキーワードが対応する URL へのハイパーリンクに変換される。

ユーザは目的に適った WIX ファイルを予め探してブックマークし、ウェブブラウザ上のツールバーから WIX システムを利用する。本論文では、WIX システムのクライアントサイドの実装と WIX ファイルをブックマークする支援システムの提案、実装を行った。

本論文の構成は以下の通りである。まず、2 章で WIX の概要について述べる。3 章で提案システムについて説明する。4 章で WIX システムのクライアントサイド実装について説明する。5 章 6 章で評価・まとめを行う。

2. WIX

WIX の概要は次の通りである。

2.1 背景

近年、Web の普及と共に人々は検索エンジンを利用して情報検索を行うようになった。ユーザは情報を得たい単語を検索エンジンに入力し、その結果を出力した Web 文書中から必要な情報を得るのが一般的である。したがって、ユーザが Web 文書中の単語に対して新たな情報を得たいという要求が生じた場合、ユーザは更にその単語を検索エンジンなどにかけなければならない。このような単語が複数存在する場合、ユーザは何回も検索エンジンを使わなければならない、かなりの負担になってしまうと考えられる。

2.2 WIX ファイル

WIX ファイルは文書中のキーワードとそれに対応する URL からなるペアを一つのエントリとしたものの集合である。実際には図 1 のような XML 形式で記述されている。キーワードを keyword 値に、URL を target 値それぞれ格納し、それらを一つとして entry タグで囲う。

2.3 アタッチ

閲覧中の Web 文書中にある英単語について詳細な情報を得たい場合、従来なら英単語 1 つずつ検索エンジンなどで検索しなければならない。これに対し WIX では閲覧中の Web 文書に WIX ファイルを結合することで、英単語からその英単語の情報

```

<WIX>
  <entry>
    <keyword>XXX</keyword>
    <target>http://www.XXX.com</target>
  </entry>
  <entry>
    <keyword>YYY</keyword>
    <target>http://www.YYY.co.jp/YYYY.html</target>
  </entry>
  ...
</WIX>

```

図 1 WIX ファイル記述方式

へのハイパーリンクが生成される。これはある文書から他の文書への動的な結合であるとみなすことができ、この結合操作を”アタッチ”と呼ぶことにする。

また、同じ英単語をキーワードとした WIX ファイルでも、例えば英和辞書の他にも英英辞書、英仏辞書、または同じ英和辞書でも会社が異なる辞書まで WIX ファイルを提供すれば、利用者は必要に応じて自由に WIX を選択し、自分の閲覧している文書にアタッチすることで状況に応じた情報を簡単に切り替えることができる。これにより Web ページ作成者の意思に関係なく、ユーザ主体で動的に自由な結合を行うことができ汎用的であると言える。

2.4 タグ情報

WIX ファイルには各々特徴を表すため、タグ情報を含んでいる。表 1 は WIX ファイルとそれに付随しているタグの例である。

表 1 WIX ファイルとタグの例

WIX ファイル	タグ
サッカー.wix	スポーツ, サッカー, Jリーグ, なでしこ
芸能人公式.wix	芸能人, 公式
ブログ.wix	芸能人, 政治家, スポーツ選手, ブログ

本研究ではあるブックマークにおいて、このようなタグそれぞれの特徴量を求めることによってブックマークの特徴付けを行って WIX ファイルを推薦していく。

3. 提案システム

3.1 概要

WIX システムでは、ユーザ主導でアタッチを行うために予め、ユーザの目的に適った WIX ファイルをブックマークしておく必要がある。WIX システムはアタッチを行う文書に独立して存在しているため、誰でも自由に WIX ファイルを生成・公開することができる。そのため WIX ファイルの数は膨大なものとなり、ユーザの目的に適った WIX ファイルを探すだけでも困難になると考えられる。そのため WIX ファイルをブックマークしやすいようなシステムでなければならない。

ユーザに最適な WIX ファイルをブックマークするために

- (1) 分類された WIX ファイルからユーザがブックマーク
 - (2) ブックマークからユーザプロフィールを生成
 - (3) 類似したブックマークから WIX ファイルを推薦
- の 3 つの工程からなる。

3.2 WIX ファイルの分類

WIX ファイルは予め図 2 のような木構造となった形で分類されている。

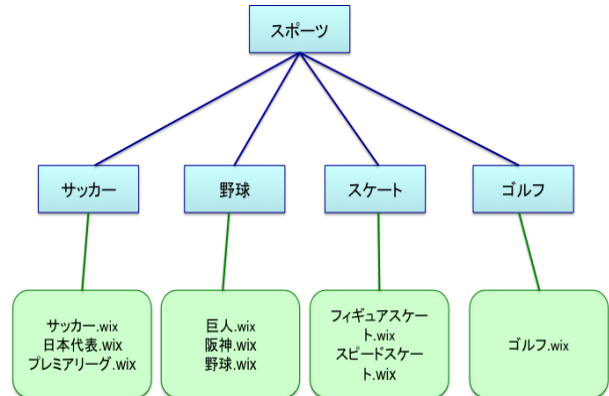


図 2 WIX ファイルの分類例

図 2 では葉ノードのみに WIX ファイルが置かれているが、ルートノードはもちろん全てのノードに対して WIX ファイルが属する可能性がある。

WIX ファイルのブックマークの流れは以下ようになる。まずユーザが WIX システムに登録し、図 2 の木構造から好きなノードを選ぶ。ここで、図 2 を例にとると、スポーツカテゴリを選択した場合とサッカー、野球、スケート、ゴルフカテゴリ全てを選択した場合は厳密には意味合いが異なる。具体的に前者はスポーツ全般の大雑把な情報を得たいと考えているのに対し、後者は各種目の細かい情報を得たいと考えていることになる。このような観点から図 2 を例にすると、根ノードの選択と葉ノードの選択に関係性は持たせないものとする。

こうして選択したノードに属する WIX ファイルの一覧からユーザは複数の WIX ファイルをブックマークをする。



図 3 ノード選択例

3.3 ユーザプロフィール生成

3.3.1 ブックマークを利用することの利点

ブックマークは作成したユーザが「今後この WIX ファイルを使ってアタッチしたい」という強い興味を表すものである。このようにブックマークはユーザの嗜好を明確に反映したものであると言える。また、ユーザの嗜好が変わった場合、ブックマークの内容を削除するなどすることにより、常にユーザの嗜好を反映することができる。以上の点からユーザプロフィールを生成するためにブックマークを利用することは大きな可能性がある。

3.3.2 ブックマークを用いたプロフィール生成

本論文ではブックマークからユーザの特徴を抽出するために、 $TF \cdot IDF$ 法を用いる。WIX システムではブックマークに複数の WIX ファイルを入れることができ、1人のユーザが複数のブックマークを生成することができる。そして、そのブックマーク1つが後述のアタッチボタンとなる。そこで、ブックマーク単位で特徴抽出を行い、各ブックマークで WIX ファイルの推薦を行うことにする。

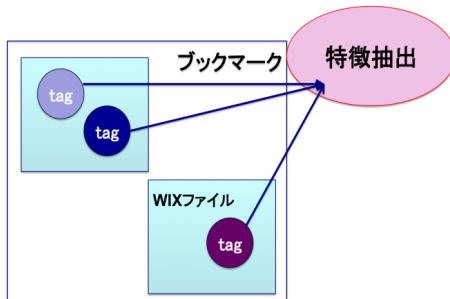


図4 ブックマークを用いた特徴抽出

3.3.3 タグからの特徴抽出

提案するシステムは図4のように、予めブックマーク内に複数の WIX ファイルが登録されていて、その各々の WIX ファイルには複数のタグ情報が含まれているものとする。またユーザは複数のブックマークを持っていることとする。これから $TF \cdot IDF$ 値をもとに、ユーザの特徴を示すタグ情報の抽出を行う。まず、ブックマーク i のにおけるタグ t の $TF \cdot IDF$ 値 $f_i(t)$ は次のように定義される。

$$f_i(t) = tf_U(t) \cdot idf_i(t) \quad (1)$$

$tf_U(t) = U$ 内におけるすべてのタグ t の出現回数

$$idf_i(t) = \frac{U}{\text{ブックマーク } i \text{ 内でタグ } t \text{ が出現する WIX ファイル数}}$$

U : 全ブックマークに登録されている WIX ファイル総数

$f_i(t)$ 値が高いタグほど、そのブックマークにおいて特徴となるタグであると判断し、 $f_i(t)$ 値が高い上位3件を抽出する。次項でこの抽出したタグを用いた特徴ベクトルを使い、WIX ファイルを推薦する。

3.3.4 特徴ベクトルからの WIX ファイル推薦

ユーザのブックマークと他ユーザとのブックマークとの類似度を測り、WIX ファイルを推薦するために、前章で抽出した特徴をベクトルで表す。 $tf \cdot idf$ 値の高かった上位3件の $f_i(t)$ を用いて特徴ベクトル f を生成する。この f は次式のように示す。

$$f = (f_i(t_1), f_i(t_2), f_i(t_3)) \quad (2)$$

t_k : 抽出されたタグ ($k = 1, 2, 3$)

このようにして生成された特徴ベクトルは、他ユーザのブックマーク j から生成された特徴ベクトルを用いて類似度を測る。次式のベクトル g はユーザのブックマーク i と比較する他ユーザのブックマーク j の特徴ベクトルである。

$$g = (f_j(t_1), f_j(t_2), f_j(t_3)) \quad (3)$$

$j = 1, 2, 3, \dots, 100$

ここで大きく異なるのは前章のようにブックマーク内から特徴となるタグを取り出すのではなく、予め決められたタグの $f_j(t)$ 値を計算し、特徴ベクトル生成に利用している点である。また最大100もの他ユーザのブックマークと比較し、後述のように特徴ベクトル f と比較していく。

類似度計算にはコサイン類似度を用いる。2つのベクトル間のなす角度をが小さいほど、ユーザのブックマークは比較しているブックマークと類似度が高い。そしてそのブックマークに登録されていて、ユーザのブックマークには登録されていない WIX ファイルは推薦できると判断した。特徴ベクトル f と g の類似度は次式のように示される。

$$\cos\theta = \frac{f \cdot g}{|f||g|} \quad (4)$$

$\cos\theta$ の値が1に近いほど、両ベクトルは類似度が高いことになる。このようにしてユーザのブックマークと他ユーザのブックマークの類似度を次々に測っていき、上位3件のブックマークに登録されている WIX ファイルを推薦する。

4. WIX システム (クライアント)

WIX システムのクライアントサイドは、Firefox のアドオンや Google Chrome のエクステンションで実装されている。以下の図は Firefox のアドオンの例である。

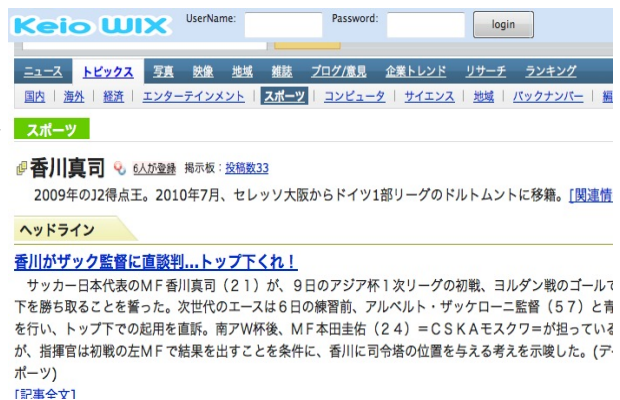


図5 WIX ツールバー (1)



図 6 WIX ツールバー (2)



図 7 WIX ツールバー (3)

図 5, 6 のようにログインすると、ツールバーにアタッチボタンが生成される。このアタッチボタンは前述のブックマークに対応し、ボタン 1 つに対して複数の WIX ファイルを登録できる。アタッチボタンをクリックすることで、図 6 のニュース記事には存在しなかったハイパーリンクが図 7 の記事で生成される。これで WIX ファイル内の target タグに記述されている URL への Web ページと結合されたことになる。このように Web 文書中の単語にハイパーリンクが生成されている例として Wikipedia 等が挙げられる。しかしそのようなサービスのほとんどは同一 Web サービス内への内部リンクである。WIX システムはアタッチボタン 1 つで、様々な URL へのハイパーリンクを生成することができるので、ユーザの様々な要求を満たすことができる。

5. 実験・評価

本システムの有用性を評価するため、評価実験を行う。客観的な評価を行うため、以下のような実験を行った。

5.1 実験：アンケート

本システムを利用者側の視点から WIX ファイルのブックマーク、推薦された WIX ファイルを評価するために、一般の学生 14 名に対しアンケートを行った。まず被験者に対し WIX ファイルシステムの説明を行った後、テストユーザでデモを行って実際に WIX システムを動かして理解させる。次に被験者自らシステムに登録し、ブックマークを複数生成させる。さらに生成したブックマークに対して推薦された WIX ファイルの評価をさせる。4 項目について、そう思う・どちらかと言うとそう思う・どちらとも言えない・どちらかと言うとそう思わない・そう思

わない、の 5 段階で答える。また回答理由についても答える。また、この実験はユーザのブックマーク総数が 50 の時と 130 の時で評価がどう変わるかについても評価を行った。

5.2 結果

5.2.1 WIX ファイルの推薦

WIX ファイルの推薦に関して次のようなアンケートを行った。

Q1. 推薦された WIX ファイルの精度は良かったか (ブックマーク 50)

Q2. 推薦された WIX ファイルの精度はあがったか (ブックマーク 130)

Q1, Q2 のアンケート結果を図 8, 図 9 に表す。

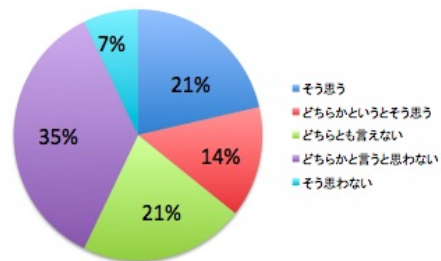


図 8 推薦された WIX ファイルの精度は良かったか (ブックマーク 50)

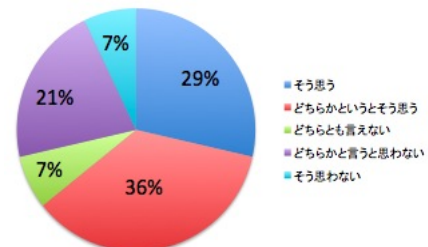


図 9 推薦された WIX ファイルの精度はあがったか (ブックマーク 130)

Q1 では被験者の 35% が肯定的な意見で、56% も否定的な意見であることがわかった。回答理由は推薦されないことがあった、関係ない WIX ファイルを推薦された等の理由が挙げられた。一方、Q2 では被験者の 65% が肯定的な意見で、否定的な意見が 28% にとどまった。回答理由は多くの WIX ファイルが推薦された、推薦された WIX ファイルをブックマークした等の理由が挙げられた。Q2 で”どちらかと言うとそう思う”とアンケートに答えた人の中に最初は推薦された WIX ファイルが良くなかったが、更新をしたら WIX ファイルの推薦が変わるのでよかったというような回答も見られた。

これらの回答に関してブックマーク総数が 100 を超える場合、比較するブックマークはランダムで選ばれるため、毎回推薦される WIX ファイルが異なる。そのため推薦される WIX ファイルの精度にばらつきがでるが、更新することで推薦される WIX ファイルが変わるため、更新して自分の嗜好に合った WIX ファ

- [1] 林昌弘, 遠山元道. "SOM による Web Index (WIX) ファイルのクラスタリング", *DEIM2010*, 2010
- [2] 森良介, 遠山元道. "階層型 WIX(Web Index) とエン트리重複に関する一考察", *DEIM2009*, 2009
- [3] Yahoo! カテゴリ: <http://dir.yahoo.co.jp/>
- [4] Yahoo! ニュース: <http://dailynews.yahoo.co.jp/>
- [5] 土方嘉徳, "情報推薦. 情報フィルタリングのためのユーザプロファイリング技術", 人工知能学会, 19 巻 3 号, pp.365-372, 2004. 5.
- [6] 渋谷雄一, 遠山元道. "Web 文書の特徴語を利用した Web Index(WIX) アタッチ支援", *DEIM2009*, 2009

イルを見つけた人が多かったためこのような結果がでたと考えられる。

5.2.2 ブックマーク数による WIX ファイル推薦精度

被験者にはブックマーク数を 2 個から 5 つに増やして実験を行った。ブックマークの数を増やしたことによる WIX ファイルの推薦の評価に関して以下のようなアンケートを行った。なお、この時のブックマークは 100 以上である。

Q3. ブックマーク数を増やしたことによって WIX ファイルの推薦は向上したか

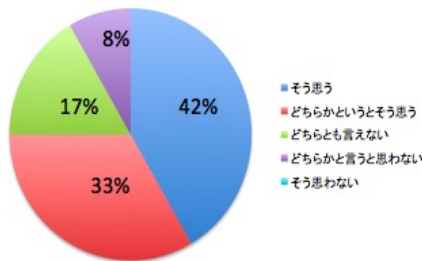


図 10 ブックマーク数を増やしたことによって WIX ファイルの推薦は向上したか

Q3 では被験者の 75%が肯定的な意見で、否定的な意見も 8%にとどまった。回答理由は良い WIX ファイルが推薦されたなどで特に他の事例と変わりはなかった。しかし他の場合と比べて良い結果がでている。

これに関してはユーザのブックマーク数を増やすことで全体のブックマーク数が増えたことに加え、各ユーザのブックマークの増加によりブックマークの特徴がやすかったため、WIX ファイルの推薦結果が向上したのではないかと考えられる。

6. ま と め

本研究では WIX システムにおいて、ユーザが数多く存在する WIX ファイルを簡単に探し、また選択しやすくすることでユーザ主導のよりよいアタッチを実現するため、WIX ファイル選択における支援システムとしてブックマーク機能・WIX ファイルの推薦を行えるように実装した。

ユーザのブックマークからユーザプロファイルを生成することで各ユーザのブックマークで類似度を算出でき、そこから WIX ファイルを推薦することができた。

本研究により、ユーザが自分の嗜好に適った WIX ファイルを探ることができるようになり、ブックマーク機能により簡単に WIX ファイルを選択できるようになることで、WIX システムがよりユーザ主導で利用できるようになったことでシステムの有用性が高まったと言える。

今後は本研究で利用した WIX ファイルへのタグの付加方法や WIX ファイルに含まれる新たな情報の利用、履歴等を利用していくことで WIX システムの更なる充実が望まれる。