



Web Indexによる情報資源結合

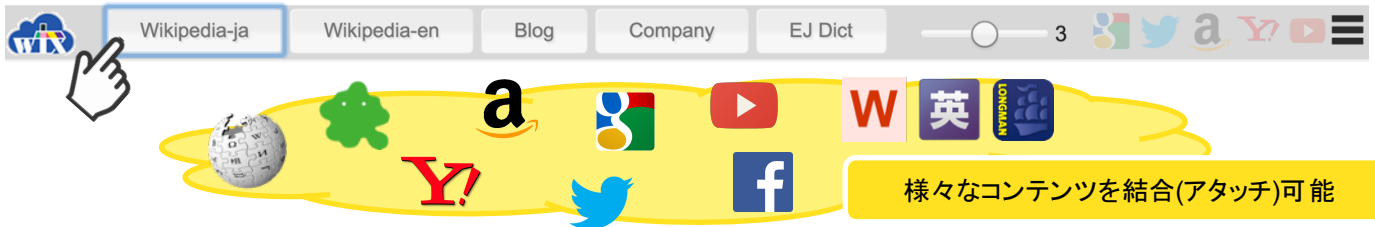
WIXから始まる、新しいWebの利用形態



Web文書内の単語を
ハイパーリンクに変換



WIXツールバーのボタンをクリックするだけ！



様々なコンテンツを結合(アタッチ)可能

```
<?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> 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>
```

単語とURLのペアを
XML形式で記述

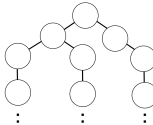
WIXファイル



DBIに格納

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	5	坂本勇人	http://www.giants.jp/G/player/prof_2756.html
:	:	:	:

ACオートマトンを構築



③辞書式マッチング

入力テキスト

フェデラーとの試合は、日本時間11日午後11時以降に始まる。

[ロジャー・フェデラー](http://ja.wikipedia.org/wiki/ロジャー・フェデラー)との試合は、日本時間11日午後11時以降に始まる。

④

②リクエスト

⑤レスポンス

アタッチ



①ブックマークボタンをクリック



サーバーサイド
クライアントサイド

研究者名

慶應義塾大学情報工学科 遠山 元道

お問合せ先

メールアドレス(遠山元道) : toyama@ics.keio.ac.jp
メールアドレス(遠山研究室) : db-lab@db.ics.keio.ac.jp
ホームページ(遠山研究室) : https://www.db.ics.keio.ac.jp/
ホームページ(Web Index) : kwix.jp



Web Indexによる情報資源結合

マッチング対象の拡張

パターンマッチャーの導入

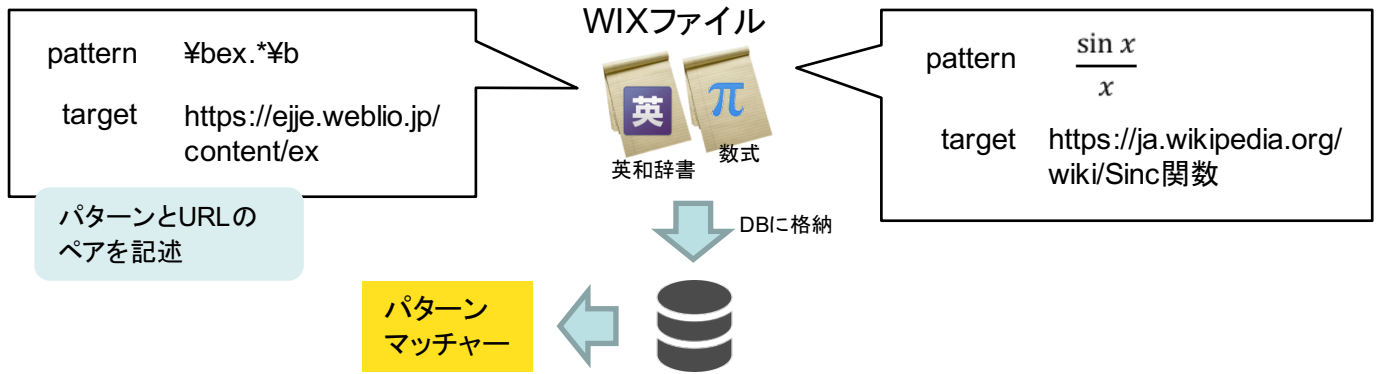
従来のWIX

固定文字列に対して辞書式マッチング



提案

様々なパターンに対してマッチング



パターンマッチング

■ パターン例① 正規表現

He explained to us
what we were expected to do.

入力テキスト

「¥bex.*¥b」にマッチ

接頭辞「ex」を持つ
単語にアタッチ

He explained to us
what we were expected to do.

出力テキスト

■ パターン例② 数式

$$\lim_{\theta \rightarrow 0} \frac{\sin \theta}{\theta} = 1$$

サブツリーと
パターンを
マッチング

$$\lim_{\theta \rightarrow 0} \frac{\sin \theta}{\theta} = 1$$

MathML: 数式をXML形式で表現

```

<munder>
  <mi>lim</mi>
  <mrow>
    <mi>x</mi>
    <mo>&RightArrow;</mo>
    <mn>0</mn>
  </mrow>
</munder>
<mrow>
  <mfrac>
    <mrow>
      <mi>sin</mi>
      <mo>&af;</mo>
      <mi>x</mi>
    </mrow>
    <mi>x</mi>
  </mfrac>
</mrow>
    
```

■ パターン例③ 英熟語

He is used to getting up early.



入力テキストの
品詞解析

「be動詞 used to 動名詞」
パターンにマッチ

He is used to getting up early.



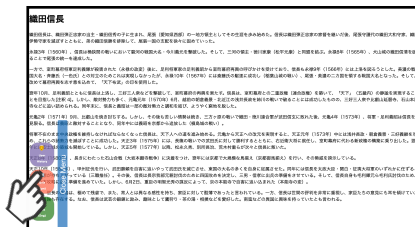
Web Indexによる情報資源結合 教育システムへの応用

選択肢問題の自動生成

課題：任意のWEBページに対する理解度を測りたい

提案：自動的に閲覧ページから選択肢問題を生成

閲覧ドキュメント



問題生成

選択肢問題



①問題生成ボタンをクリック

②HTMLドキュメントを送信

④生成された問題を返信

クライアントサイド

サーバーサイド

Web Index
(キーワードに対する情報資源結合)

③Wikipediaへのリンクが
アタッチされる

カテゴリ：戦国武将

上杉謙信
浅井長政
今川義元

武田信玄



アタッチされたキーワードを
「正解」として選択

同カテゴリ内のキーワード
を「不正解」として選択

Sentence Selection
(問題文になり得る文を選択)

Correct Key Generate
(問題文に対する正解の選択)

Distracter Generate
(問題文に対する不正解の選択)

<問題の生成例>

「三方ヶ原の戦いで織田・徳川連合軍が()に敗れた後、元龜4年(1573年)、将軍・足利義昭は信長を見限る」
1. 上杉謙信 2. 今川義元
3. 武田信玄 4. 浅井長政

研究者名

慶應義塾大学情報工学科 遠山 元道

お問合せ先

メールアドレス(遠山元道): toyama@ics.keio.ac.jp
メールアドレス(遠山研究室): db-lab@db.ics.keio.ac.jp
ホームページ(遠山研究室): https://www.db.ics.keio.ac.jp/
ホームページ(Web Index): kwix.jp