

WinActorにおける XPathによる要素選択の一手法

2022/12/20





- WinActorは、Webページを操作するために、操作対象のWebページ内の部品をXPathで指定する必要がありますが、シナリオ作成時に、このXPathによる指定がうまく行かないとお困りの声をいただきます。
⇒ XPathによる部品選択のやり方について解説します。

XML Path Language (XPath) は、標準化団体W3Cで開発された、XMLやXMLに準拠したHTML中の特定の部分 (Webページ内の部品) を指定することのできる簡単な言語です。

WinActorのライブラリの**23_ブラウザ関連**配下のWebページ内のボタンやリンクをクリックするライブラリ等の多くでは、操作対象の要素を選択するためにこのXPathを用いています。

以下の流れにて、WinActorのシナリオに設定するXPathの指定方法について説明します。

- ① ブラウザによるXPathの取得方法
- ② ブラウザから取得したXPathを用いた場合の問題点
- ③ この問題点に強いXPathの指定方法
- ④ IE操作で用いている対象コントロール指定をXPathで置き換える方法 (参考)

The screenshot shows a configuration window for a 'クリック' (Click) action. It includes a '名前' (Name) field with the value 'クリック' and a 'コメント' (Comment) field. Below these are tabs for '設定' (Settings), 'スクリプト' (Script), '注釈' (Notes), and 'バージョン情報' (Version Information). The '設定' tab is active, showing instructions: 'Webページ内の要素(ボタンやリンク等)を指定しクリックします。' (Specify an element in the web page (button or link, etc.) and click it.) It also includes fields for 'ブラウザ名' (Browser Name) and 'XPath', both with '値⇒' (Value⇒) and a dropdown arrow. Explanatory text for these fields is provided: '[ブラウザ名] : 操作するブラウザのブラウザ名を設定します。 「ブラウザ起動」で設定した[ブラウザ名]と対応します。' and '[XPath] : クリックする要素のXPathを設定します。'



ブラウザの開発ツールからのXPathの取得

- ブラウザ(Edge, Chrome, Firefox)の開発ツール (F12) を利用し、ブラウザに表示しているHTML中の要素のXPathを取得することができます。

① WinActorのウェブページ (https://winactor.biz) を表示し、開発ツール (F12) を開く。

② 要素選択 (Ctrl+Shift+C) を選択し、HTML中の操作対象の要素 (WinActorのアイコン) を選択すると、開発ツール内の該当要素が選択された状態になります。

③ 開発ツール内の該当要素が選択された状態で、右クリックし、コンテキストメニューを開き、コピー → XPathのコピーを選択すると、HTML中の操作対象の要素のXPathが取得できます。

```
<html lang="ja">
<body class="v2-1-body-top">
  <!-- Google Tag Manager (noscript) -->
  <!-- End Google Tag Manager (noscript) -->
  <!-- ヘッダー -->
  <header class="v2-1-header">
    <div class="v2-1-header-inner">
      <h1 class="v2-1-header-logo">
        <a href="/>
```

取得されたXPath: `/html/body/header/div[1]/div/div[1]/a/img`

① ブラウザ (左図はEdge) にて、F12を押すことにより、緑枠の開発ツールを利用できます。

② 要素選択 (Ctrl+Shift+C)を選び
③ HTML中の操作対象の要素 (左図ではWinActorのアイコン) を選択すると
④ 開発ツール内の該当要素が選択された状態になります。

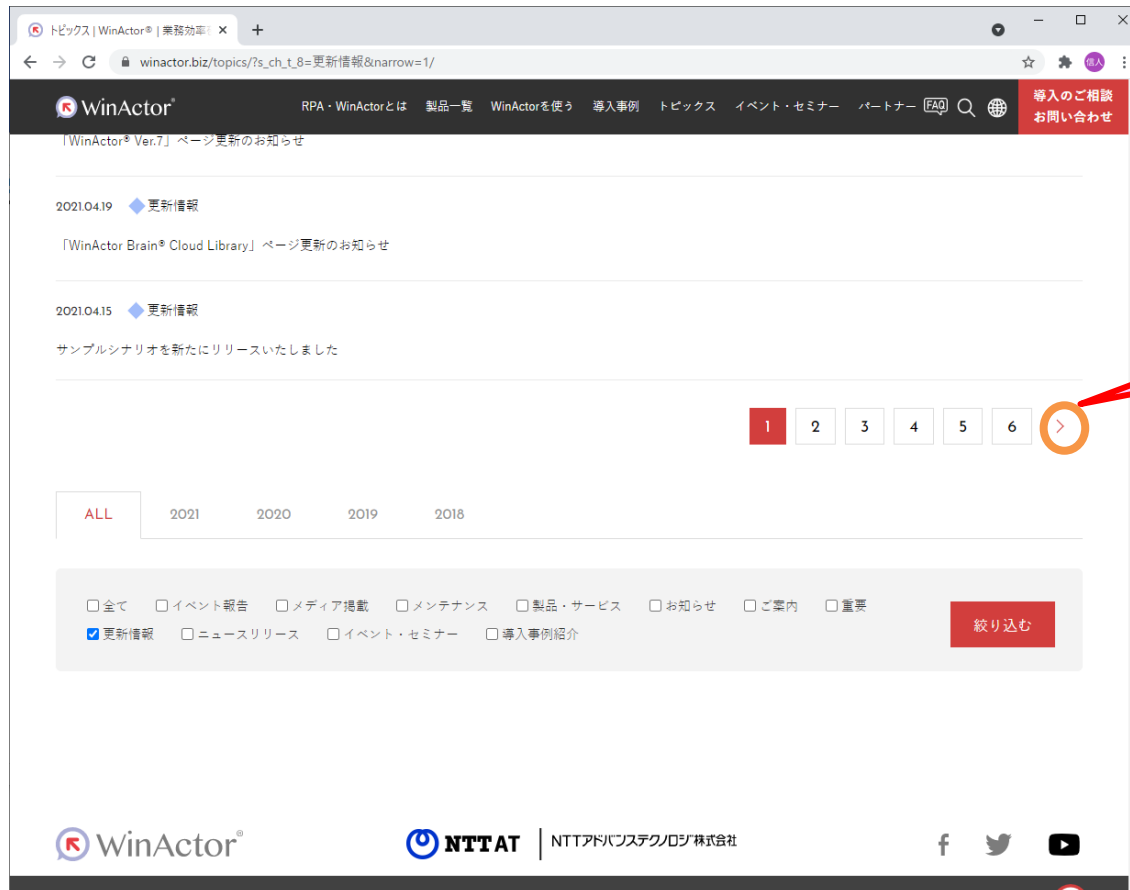
⑤ ④で選択された行を右クリックし、コンテキストメニューを開き、コピー → XPathのコピーにて、HTML中の操作対象の要素のXPathが取得できます。

`/html/body/header/div[1]/div/div[1]/a/img`

開発ツールから取得されるXPathの問題点



- 開発ツールでは、現在ブラウザが表示している要素を指定するXPathが取得されますが、シナリオ実行時に、お知らせのメッセージが増えていたり、検索結果のページが異なっていたりし、HTMLの構造がシナリオ作成時とシナリオ実行時で異なっていると、そのままでは正しく要素を指定することができないことがあります。



昨日まで5ページしかなかったのに、今日は6ページあり、次ページのリンクをクリックできず、シナリオの実行に失敗した！

■ 昨日（シナリオ作成時）のXPath

```
/html/body/div[2]/div[1]/div/div/ul/div/div/a[6]
```



■ 今日（シナリオ実行時）のXPath

```
/html/body/div[2]/div[1]/div/div/ul/div/div/a[7]
```



- Webページの構造の変化に強いXPathを選択します。

○ ※1 ■「XPathをコピー」から得られたXPath

```
//*[@id="js_fs_paginate"]/div/a[6]
```

`//*[@属性="値"]`は、指定された値の属性が設定されている要素を選択します。

これにより、`<div id="js_fs_paginate">`要素が選択されます。

``を指定するために

```
//*[@class="label"]
```

といった指定も可能。

id属性はページ内でその要素を特定するために利用されます。

class属性はページ内で同じ（デザインの）グループを表すために利用されます。

状況に応じ使い分けます。

✗ ※2 ■「完全なXPathをコピー」から得られたXPath

```
/html/body/div[2]/div[1]/div/div/ul/div/div/a[6]
```

左に比べこの部分の変化に弱い



id属性にて、ページ内の要素を特定することができるため、操作したい対象の要素に近いid属性で特定できる要素を選択し、XPathで利用します。

ブラウザの開発ツールの「XPathをコピー」では、この選択は自動的に行ってもらえるので、そのまま利用できます。

id属性で特定できる要素が無い場合は、「XPahをコピー」と「完全なXPathをコピー」で同じXPathが取得されます。

id属性で指定される要素より親側の変化が強くなったが、前のページの子供側の変化に対応できていない！

※1 Edgeでは「XPathをコピー」、Chromeでは「Copy XPath」、Firefoxでは該当メニューなし

※2 Edgeでは「完全なXPathをコピー」、Chromeでは「Copy full XPath」、Firefoxでは「XPath」



ページの変化に強いXPathとは（その2）

● id属性以外の方法による要素の特定方法について、いくつか紹介します。

■ a要素内のテキストが「次へ」である場合 (次へ)

```
//*[ @id="js_fs_paginate" ]/div/a[ text()="次へ" ]
```

[text()="文字列"]

にて、要素の内部の文字列が一致する要素が選択されます。

■ a要素内のテキストが「次へ」を含む場合(一致はしない場合) (>次へ)

```
//*[ @id="js_fs_paginate" ]/div/a[ contains(text(), "次へ") ]
```

[contains(text(), "文字列")]

にて、要素の内部の文字列が部分一致する要素が選択されます。
[contains(@class, "label")]
といった使い方もできます。

■ 最後に現れるa要素の場合

```
//*[ @id="js_fs_paginate" ]/div/a[ last() ]
```

[last()]

にて、最後の要素を取得することができます。
([0]で先頭の要素)

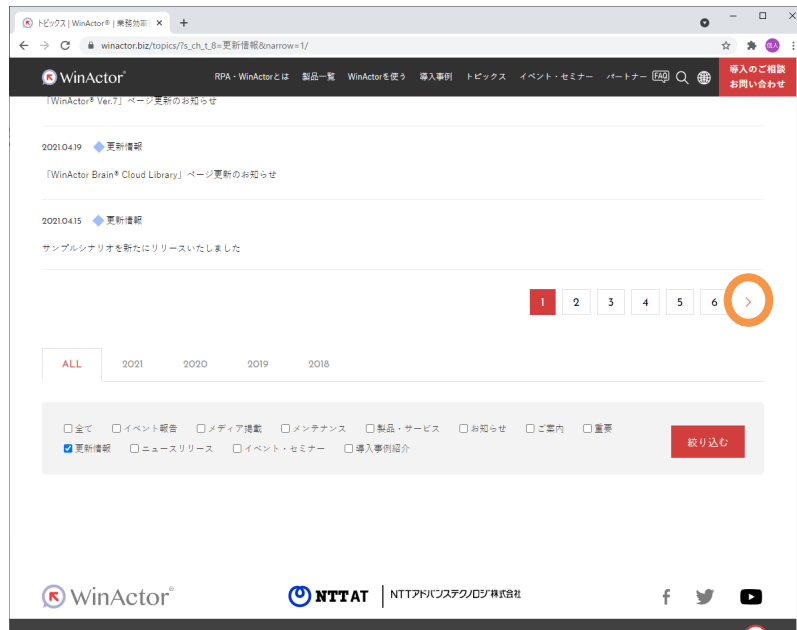
ブラウザの開発ツールから取得されるXPathについて、id属性にて特定できない要素に対しては、テキスト、他の属性、及び、順番といった様々な条件(XPathでは上に記載したもの以外にも多くの条件を用いることができます)を組み合わせることで、変化に強いXPathを設定できます。

23_ブラウザ関連/テーブルスクレイピングライブラリの、ページ遷移要素XPathの選択ボタンではテキストによる条件を用いたXPathを自動的に作成していますので、参考にしてください。

IE操作の対象コントロール指定をXPathで行うには？



- IE操作では対象コントロールの指定にて、XPathではなくtag, frame index, tag index等を利用しています。
- IEのサポート終了に伴い、Edge等の別のブラウザによる操作へ置き換える方法はないか？



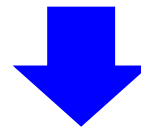
■「対象コントロール指定」で得られる詳細設定

基本設定	詳細設定
<input checked="" type="checkbox"/> 対象特定種別	値
<input checked="" type="checkbox"/> tag	値⇒ a
<input checked="" type="checkbox"/> frame index	値⇒ 0
<input checked="" type="checkbox"/> tag index	値⇒ 130
<input type="checkbox"/> name	値⇒
<input type="checkbox"/> type	値⇒
<input type="checkbox"/> id	値⇒
<input type="checkbox"/> value	値⇒



//a[130]

変化には弱いが、
現シナリオが安定しているならば試したい。
ブラウザの種類により表示を変える
ページもあり、必ずしも一致はしません。



現時点では、動作確認が必要なものとはなりますが、
以下のXPathで概ね代用できますので、お試しください。

`//<tag>[<tag index>][@name=<name>][@type=<type>][@id=<id>][@value=<value>]`

<frame index>が0以外の場合、23_ブラウザ関連/フレーム選択ライブラリを用いたフレームの選択が必要です。

WinActorメーカー 公式ホームページ

<https://winactor.biz/>



未来を拓くチカラと技術。

