

# ページ内の任意情報を一括で取得①

## 知識①

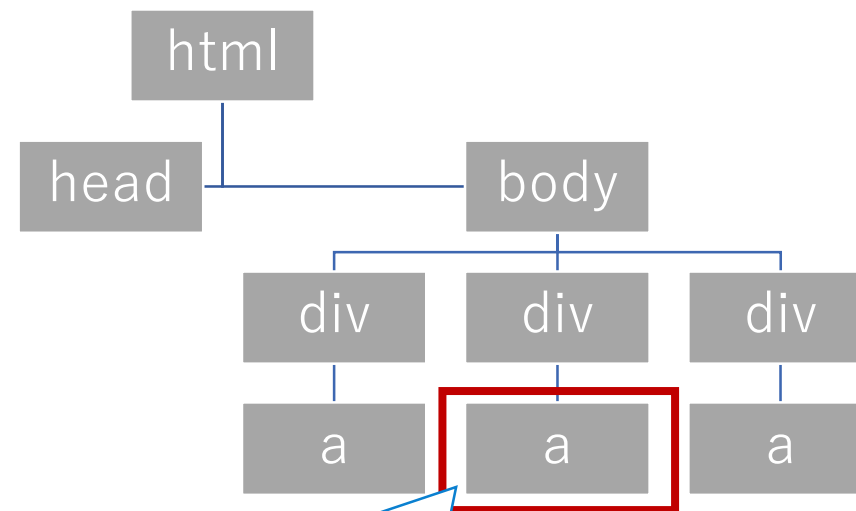
## Xpathとは？

XPathとはページ内の場所を示す住所のようなもの



弊社HPのHTML

HTMLはツリー構造になっている



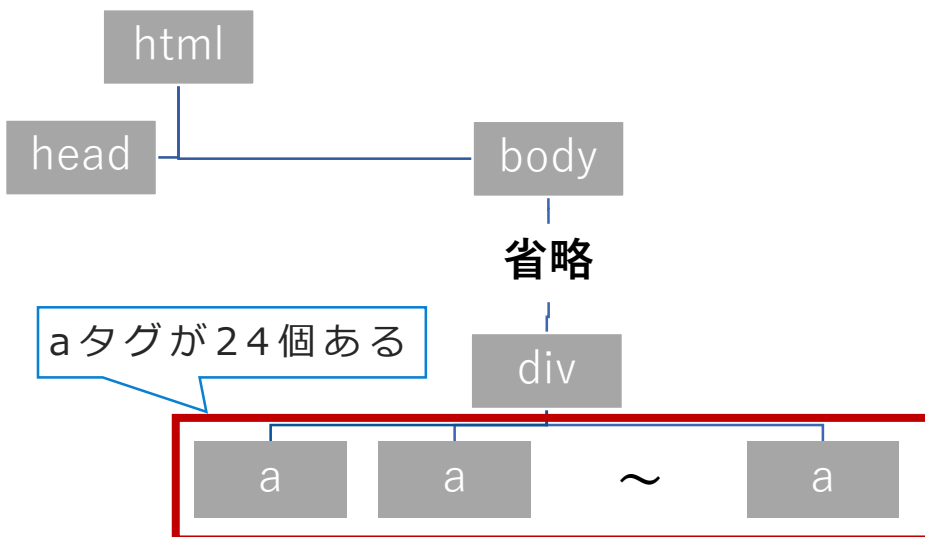
ここのXPathは下記になる  
`/html/body/div[2]/a`

# ページ内の任意情報を一括で取得②

## 知識②

### Xpathを見ると、パターンが見える

ニュースリリースの一覧をツリー構造にすると、  
同じ階層にaタグが24個あり、規則的に作られていることが解る。



各aタグ内にあるpタグにニュースタイトルがある。

最初のニュースタイトルのXPathは、

`/html/body/main/div[2]/div/div/a[1]/p`

最後のニュースタイトルのXPathは、

`/html/body/main/div[2]/div/div/a[24]/p`

赤文字の数字を変えて、ニュースタイトルのXPathを取得する。

# ページ内の任意情報を一括で取得③

## 解決法

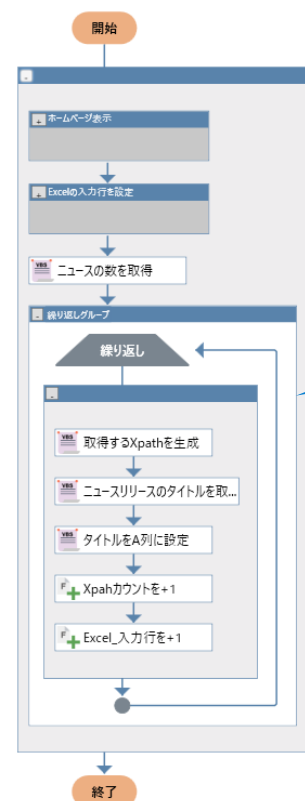
### XPathを文字列連結し、繰り返す

XPathのパターンが理解したら、Xpathを文字列連結で生成し、繰り返しノードを使い、一括取得を実現する。

ブラウザ情報を取得する場合、定番の繰り返しの構成は2点

- ①要素数を取得 ※今回の場合はaタグの数
- ②取得XPathを文字列連結し生成。
- ③1から要素数までをカウントする用の変数 ※XPathカウント

## シナリオ構成



- ① 繰り返しに入る前にニュース数を取得
- ② 繰り返し条件は下記参照
- ③ 繰り返しの中で、XPathを生成
- ④ XPathを使い、ニュースタイトル取得しExcelへ入力。
- ⑤ 繰り返しの最後にカウントアップ

## 繰り返し条件

プロパティ  
プロパティ

条件式

条件 Xpathカウント ≤ Xpath要素数

更新 元に戻す

条件式：XPathカウント ≤ Xpath要素数