

本書ではChromeのデベロッパーツールを例に見ていきます。デベロッパーツールを開くと、上側にWebサイトが表示され、下側に開発ツールが表示されています。開発ツールの上のメニューの一番左にある矢印アイコンをクリックすると、「その要素がどのHTMLで表されているか」を表示することができるモードになります。その状態で、上側に表示されているWebサイトからランキング1位のタイトルをクリックしてみてください。すると、タグの階層に挟まれて、商品ページのリンクと商品名を表すHTMLが開発ツールの画面に表示されます。ランキングの商品のタイトルは、次のようなHTMLの構造の中にあることがわかります。

```
…省略
<div id = "zg_item_row">
…省略
<img alt = "ドラゴンクエストXI 過ぎ去りし時を求めて" src= …省略 >
…省略
</div>
…省略
```

HTMLのタグをもとに要素を抜き出すと、次のようなコードになります。



```
>>> import requests ↵
>>> from bs4 import BeautifulSoup ↵
>>> game_ranking_html = requests.get('http://www.amazon.co.jp/gp/
                                     new-releases/videogames/637394') ↵
>>> soup = BeautifulSoup(game_ranking_html.text, "html.parser") ↵
>>> soup.find(class_='zg_itemRow').find('a').img['alt'] ↵

'ドラゴンクエストXI 過ぎ去りし時を求めて'
```

soupを作った後、findメソッドを使っています。findメソッドは引数に渡した条件のタグを探してきて、最初に見つけた場所を取得します。ここではfindメソッドを使って、imgタグのalt属性から目的の商品名を取得できました。確認ができたので、次はfindAllメソッドを使って、ページに表示されているランキングを取得します。



findメソッドの注意点

findメソッドでHTMLのclassの名前で検索するときには、「class='クラス名'」ではなく、「class_='クラス名'」と、_を付けてちょっと名前を変えて指定する必要があります。なぜなら、Pythonのプログラムの中では「class」はPythonでのクラスとして使うために予約語(⇒p.54)に指定されているからです。



```
>>> import requests ↵
>>> from bs4 import BeautifulSoup ↵
>>> game_ranking_html = requests.get('http://www.amazon.co.jp/gp/
                                     new-releases/videogames/637394') ↵
>>> soup = BeautifulSoup(game_ranking_html.text, "html.parser") ↵
>>> for game in soup.findAll(class_='zg_itemRow'): ↵
...     print(game.find(class_='zg_rankNumber').string + game.img['alt']) ↵
```

1. ドラゴンクエストXI 過ぎ去りし時を求めて
2. ドラゴンクエストXI 過ぎ去りし時を求めて (早期購入特典「しあわせのベスト」「なりきんベスト」を先行入手することができるアイテムコード 同梱)
3. PlayStation(R)4版 ドラゴンクエストXI 過ぎ去りし時を求めて 公式ガイドブック (SE-M00K)
- …省略
19. ヒットマン ザ・コンプリート ファーストシーズン
20. 【PS4】クラッシュ・バンディクー ブッとび3段もり!



ランキング通りに商品名を取得することができました。ここでは「zg_rankNumber」の値も取得して、商品の他のデータ順位も表示しました。他にも発売日や値段などを取得したい場合は、同じようにタグの情報を見ながら、追加してみてください。

先ほど説明した通り、HTMLのページはXMLと比べてタグの数が多いのですが、基本的な流れは同じであることがわかりましたね。なお、もしかすると途中で気づかれた方もいるかもしれませんが、実はAmazonのランキングのページはRSSで配信もされています。

▶ Amazon ゲームの新着ランキングRSS

URL <http://www.amazon.co.jp/gp/rss/new-releases/videogames/637394>

もし、HTML版のほうでうまく実行できなかった場合は、こちらでもチャレンジしてみてください。