# 240ページの「Googleニュースをスクレイピングする」についての補足

2018年10月16日（追記あり）

Googleニュースのページの仕様変更により、この節で紹介している原著のコードではヘッドライン記事のURLを集められなくなりました。

そこで、日経トレンディネットのページから、記事のURLをスクレイピングできるコード（プログラム）を用意いたしました。記事のURLを単純にスクレイピングするscraping1.pyと、重複するURLを取り除いてスクレイピングするscraping2.pyの２種類がございます。

240ページで紹介しているBeautifulSoupのインストールを済ませれば、これらのプログラムを動かせます。プログラムの概要については、プログラム中のコメントをご参照ください。

**【2020 年 6 月 4 日追記】**

上記の日経トレンディネットが日経クロストレンドに移行したことにともない、サイトの構造が変わりました。そこで、コードを以下のように修正いたしました。

1. URL を変更：
   【旧】https://trendy.nikkeibp.co.jp/news/
   【新】https://xtrend.nikkei.com/atcl/contents/new/
2. URL パス条件を変更:
   【旧】if 'atcl/news'
   【新】if 'atcl/contents'
3. (おまけ)リダイレクト対策として:
   【旧】urljoin(self.site, url)
   【新】 urljoin(r.url, url)

以上により、記事の URL をスクレイピングできるコードについては、chap20-crawler-nikkeibp-trendy-news.py と、重複する URL を取り除いてスクレイピングする chap20-crawler-nikkeibp-trendy-news-nodup.py の２種類をご用意しています。

以上、補足いたします。