

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

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します。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。

以上、補足いたします。