

Web スクレイピングとは？

Web スクレイピング

- Web クローラー : Web サイトを巡回し情報収集、インデックス作成
- Web スクレイパー : Web ページから必要データを抽出
- マーケティング、業務効率化、サービス開発などの用途



人間のコピーアンドペースト
 テキストパターンマッチング
 HTTP プログラミング
 HTML 解析
 DOM 解析
 垂直集約
 セマンティック注釈認識

Web スクレイピングは、RPA の自動化で多くの業務に必要とされる機能です。

Web スクレイピング (Web scraping) とは、Web サイトから、データを抽出してくれる技術手法のことです。Web harvesting、web extraction と表現されることもあります。人が、Web ページの情報をコピー＆ペーストする行為と同じですが、それをコンピュータの自動処理で行えるため、気の遠くなるような大量のデータを Web 上から取得することができます。

Web スクレイピングは、一般的に Web クローラーと Web スクレイパーの 2 つの部分で機能します。

Web クローラーは、Google 社の「Googlebot」のように、インターネットを巡回して、リンクをたどって Web サイトの情報を収集し、自動的にコンテンツのインデックス作成と検索を行うプログラムのことを指します。

そして、Web スクレイパーは、Web ページから必要なデータを正確かつ迅速に抽出するように設計された専用ツールです。Web スクレイピングを行う為のツールは、オープンソースの無償のものから、有償で販売されているものまで、いろいろなものが入手可能です。

RPA で業務を自動化する場合、この Web スクレイピングは様々な場面で活躍します。事務的な業務の中でも、クラウドアプリケーションが多くなってくると、複数の Web サイトにログインして、情報を入力したり、表示された情報を転記したりする作業が多くなります。それを、いちいち業務担当者が、コピー＆ペーストしたり、検索して情報集めている場合、Web スクレイピングを利用することで、担当の負荷も減り、正確性も増します。

多くの RPA ツールは、Web 関連業務の自動化を推進するためにも、このスクレイピング機能に長けています。有料のツールでなくても、Selenium などのオープンソースのソフトウェアを活用して、Python などの汎用プログラミング言語で、自社でコーディングして Web スクレイピングを実行することも可能です。

また、Web クローラーと Web スクレイパーの専用ツールを使用すれば、その多くはプログラミングしなくても、Web スクレイピングが可能なので、うまく RPA ツールと組み合わせることで業務の自動化が可能になります。

クラウドで行う業務が多くなる傾向の中、業務の改変や、新規業務には、この Web スクレイピングを上手く取り入れて業務の効率化を図っては如何でしょうか？