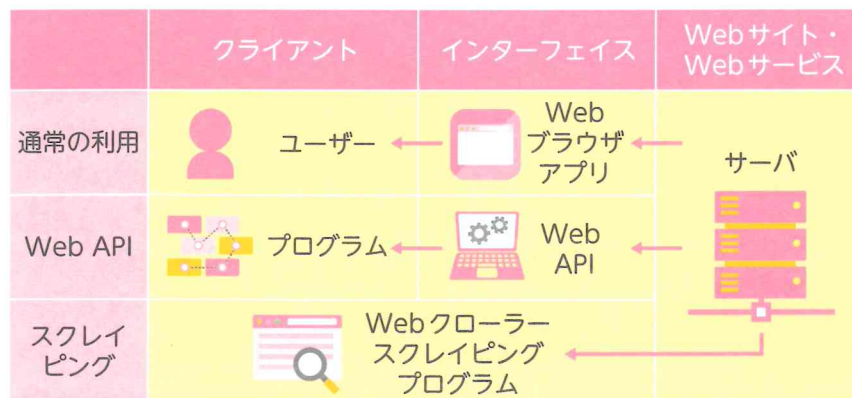


## ○ Web APIやスクレイピングを活用する

インターネットは、画像・動画・音声をはじめとした多種多様なデータが公開されており、もっとも巨大なデータベースと言えます。とはいえ、人間がブラウザやアプリを操作することでインターネットからデータを手に入れようとすると、膨大な時間がかかってしまいます。その点プログラムが自動で行ってくれるのであれば、これほど強力なデータベースはありません。しかしインターネットのデータは、Webページのデザインがすべて異なることなどからわかるように、形式が整っておらず、かんたんにはデータを自動取得できません。したがってインターネット上のデータを利用する場合には、WebページやWebサービスから目的のデータを抽出して機械学習で利用できるような形式にまとめるプログラムが必要となります。ここではその方法として代表的な**Web APIとスクレイピング(クローリング)**について解説します。

### ■ Web APIとスクレイピングのデータフロー



Web APIは、一般的にはWebサービスの提供者が用意したアプリケーションです。Web APIを利用することで、そのWebサービスで扱われているデータを取得できます。取得できるデータはWebサービスによってさまざまです。たとえばFacebookやTwitterなどのSNSのWeb APIでは、SNS内の投稿やユーザー情報、トレンドなどの情報を、また大手インターネット通販サービスのAmazonのWeb APIでは商品情報や売れ筋などの情報を取得できます。

一方でスクレイピングは、一般的にWeb APIを介さずに直接WebサイトやWebサービスからデータを取得することです。スクレイピングは主に、Web APIが存在しないWebサービスからデータを取得したい場合や、Web APIで取得できないデータを必要としている場合に利用されます。スクレイピングには、公開されているWebクローラーか自身で作成したプログラムを利用します。

WebサイトやWebサービスのデータは、Webサーバに存在しています。私たちがそれらを利用するときは、ブラウザやアプリをインターフェイス(窓口)としてデータを取得しています。人間の代わりにコンピュータがWebサーバにアクセスする場合には、Web APIやWebクローラー、スクレイピングプログラムがインターフェイスとなりデータを取得します。

Web APIは提供者によって利用方法がまとめられている場合が多く、またWebサービスの仕様変更の影響を受けにくい点が特徴です。ただし、そもそもWeb APIが用意されていない場合には利用できません。それに対してスクレイピングは、基本的にどのようなWebサイトやサービスであってもデータを取得することができますが、その仕様変更に合わせてWebクローラーやスクレイピングプログラムを対応させる必要があります。また、スクレイピングが禁止されているWebサーバに実施したり、サーバに大きな負荷を与えたりした場合などには法的責任を問われる可能性があります。

### ■ Web APIとスクレイピングのメリット・デメリット

	メリット	デメリット
Web API	<ul style="list-style-type: none"> <li>・ 利用方法がまとめられており利用しやすい</li> <li>・ 仕様変更の影響を受けづらい</li> </ul>	<ul style="list-style-type: none"> <li>・ Web APIがないと利用できない</li> <li>・ データの取得に制約が存在することがある</li> </ul>
スクレイピング	<ul style="list-style-type: none"> <li>・ Web APIが用意されていないWebサービスのデータが取得できる</li> <li>・ Web APIに設定されていないデータを取得できる</li> </ul>	<ul style="list-style-type: none"> <li>・ 仕様変更の影響を受ける</li> <li>・ 法的責任を問われる可能性がある</li> <li>・ 一定の知識とスキルが必要</li> </ul>

## まとめ

- データの収集方法は、「自分で収集」「公開されているデータベースの利用」「Web APIやスクレイピングの活用」