

ユーザーガイド

初心者から学ぶ

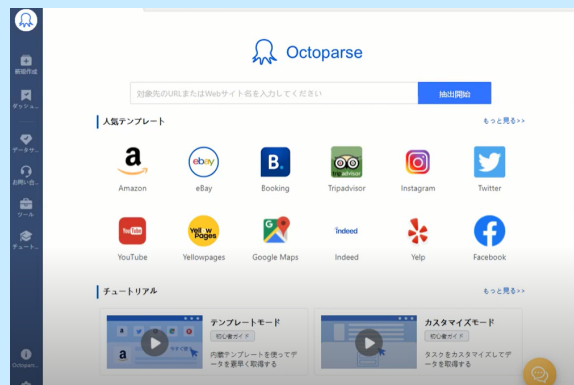


Octoparse は何ができる

① スクレイピングツール「Octoparse」

本ツールでテンプレートをご利用したり、タスクを作ったりをすることで、プログラミング知識がなくても、データをスクレイピングを楽しむこともできます。また、本ツールの基礎機能のご利用は無料でございます。

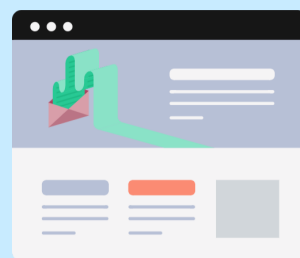
ツールダウンロードは[こちらをクリック!](#)



② テンプレート&クローラー&タスクの代行サービス

本サービスは[以下の方](#)におすすめです：

- サイトの構造により、自作のタスクで抽出できない場合
- タスクの作成は面倒と思う方、
- 効率的にタスクを導入する方へ



弊社のエンジニアが作成したテンプレート、クローラー、タスクをOctoparseツールで導入することで、クリック一つでデータを安定的に正確的に抽出できます。このサービスを利用する場合は対象サイト構造の変更によるデータ抽出できないになったら、こちらが無料でテンプレートやクローラーなどの修復をサポートできます。

もっと詳しくは：[こちらをクリック!](#)

③ データサービス

本サービスは以下の方におすすめです：

- 直接にデータを購入したい
- 膨大なデータ量のある場合

もっと詳しくは：[こちらをクリック!](#)



Octoparseは何かできる

スクレイピングについて法的な問題

Webスクレイピングには、その適用に対処するための明確な法律がありませんが、スクレイピングを行う際に気を付けなければ違法になってしまう場合が3つあります。

- 利用規約に違反する
- サーバに過度の負荷をかける
- 著作権を侵害する

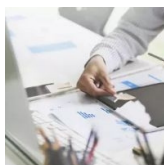
スクレイピングする際に著作権者の利用許諾が必要になりますが、著作権法は、いくつかの例外を認めています。それは「私的使用のための複製」「情報解析のための複製」と「検索エンジンの提供のための利用」です。



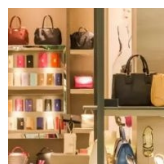
関連記事は[こちらをクリック!](#)

Octoparse 応用例

Octoparseは日本語の他に、英語、ドイツ語、フランス語、スペイン語の独立のWebサイトもサポートしています。2021年12月の時点で全世界で400Wの利用者を持っています。SONY、accenture、Wayfair、Tencent、TOYOTA、NEC、HONDAなどの様々な業界の1000以上の企業、学校、政府機関から信頼をいただいています。



簡単ステップで
Web上のデータを短時間で収集!
消費者の生の声を分析し、クライアントに提案!



高度なカスタマイズで
ECサイト上の細かいデータを取得!
独自のデータ分析方法で最適な仕入れ先と商品を選定!

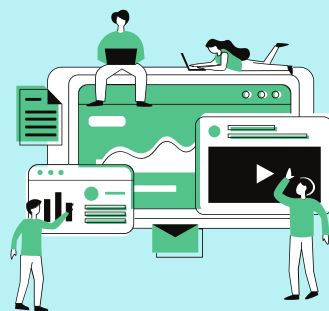


1日3~4万件!
一番鮮度の高い営業リストを取得!
広告代理社のOctoparse活用方法



簡単な設定で
データを自動取得!
自営業ビジネスのために情報面の優位性とデータベースの整備を両立

応用例についてもっと詳しくは:



octoparseの応用例についてもっと詳しくは[こちらをクリック!](#)

Octoparseの特徴

ノーコードのスクレイピングを楽しむ

💡 octoparseはコードを書かなくても簡単にデータを抽出できるノーコードスクレイピングツールです。



- octoparseで100以上のテンプレートと自動識別機能を備え、素人にもクリックするだけで、Webスクレイピングを実現！

メリット：簡単！

デメリット：限界がある

- カスタマイズタスクの場合はワークフローを作って、Webスクレイピングもノーコードで実現！

メリット：高度なカスタマイズ性

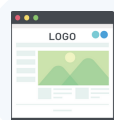
デメリット：octoparseを使い方を学ぶ必要がある

使い方が簡単！



自動識別機能

自動的に抽出したいデータを見分ける。



テンプレート

100以上のWebサイトのテンプレート！
クリッカー一つでデータを抽出！



操作ヒント

タスクを作成する際に、次のステップを提示します。

データ抽出の高速化



加速モード

加速モードを利用して、ローカル抽出の速度を3倍以上アップ！



クラウド抽出

クラウド型プラットフォームでは、6~20倍の高速抽出が可能です。

クラウドサービス



自動 IPローテーション

IPが自動的に巡回できるので、アクセス拒否を効果的に避けられる。



管理やすい

いつでもどこでもデータの利用、保存、共有化を実現！

スケジュール設定

定時抽出



毎日、毎週など指定したスケジュールにあわせて自動取得が可能！

多種のデータ出力



CSV、Excelなどの形式で出力でき、APIに接続、データベースにも保存できる。

ダウンロード & インストール

Octoparseをインストール

PCのシステム必要条件：

Win7 / Win8 / Win8.1 / Win10 (x64)

注意点：

- Windows XPまたは32ビットオペレーティングシステムの場合は、Octoparse 7.3.0を利用してください。
- システムが64ビットか32ビットかわからない場合は、スタートメニューの「設定」アイコンをクリックし、「システム」→「バージョン情報」を選択して確認できます。

Windows



ツールダウンロードは
こちらをクリック！

Mac



ツールダウンロードは
こちらをクリック！

アカウントを登録

下記の手順からノーコードのスクレイピングの旅を始めましょう！

①アカウントを登録する

>初めてOctoparseを利用する方は下記の画像で示したように、新規アカウントを作成しましょう！
アカウントを作成は：
[こちらをクリック！](#)

>Octoparseのアカウントをお持ちしている方はユーザー名/登録したメールアドレスとパスワードを入力して、ログインします。

②スクレイピングを楽しむ

Octoparseでの基礎機能のご利用は**無料**でございます。クラウド抽出、API整備、テンプレートなどのご利用は有料機能でございます。もっと強力な機能と作業効率化を図るお客様はOctoparseの有料機能を是非お試ししてください。

octoparseの支援制度

割引&トライアル&学術研究への支援制度

トライアル：

Octoparseでは、新規ユーザー向けに14日間無料トライアルを提供しております。14日間の無料トライアルは正式プレミアムと同じ機能を使うことができます。例えば、APIアクセス、IPローテーション、クラウド抽出などです。
もっと詳しくは：[こちらをクリック！](#)

トライアルを申し込みたい場合はツール内で申し込みます。



割引についてのご相談・お申し込みは下記のURLからサポートにお問い合わせください。

<https://helpcenter.octoparse.jp/hc/ja/articles/360037277374>

学割：

①Octoparseの学割を申請する方は下記の資料を用意して、カスタマーサポートまでお問い合わせください。

- 身分証明書（学生証、教員証、または学校用のメールアドレスなど）
- 学校の名前と専攻
- データスクレイピングプロジェクトに関する簡単な説明

②学校用のメールアドレスで新規アカウントを登録してください。

もっと詳しくは：[こちらをクリック！](#)



学術研究への支援制度：

Octoparseは教育機関、ジャーナリズム、非営利団体向けに特別割引料金をご用意しております。申請に必要な資料を用意して、カスタマーサポートまでお問い合わせください。

教育機関：

- 学校の名前
- 身分証明書（学生証、教員証、学校用のメールアドレスなど）
- スクレイピングの目的

ジャーナリズム：

- レビューのリンクとスクリーンショット
- スクレイピングの目的

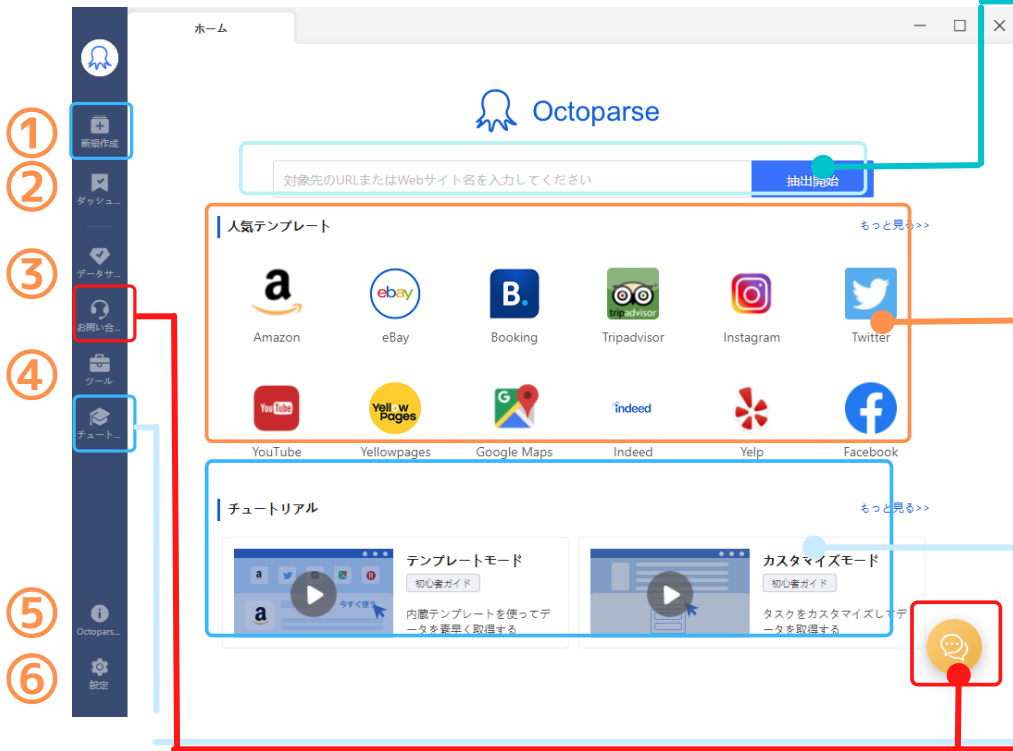
非営利団・NPO：

- 所属機関の証明書（公式サイト）
- スクレイピングの目的

もっと詳しくは：[こちらをクリック！](#)



インターフェース



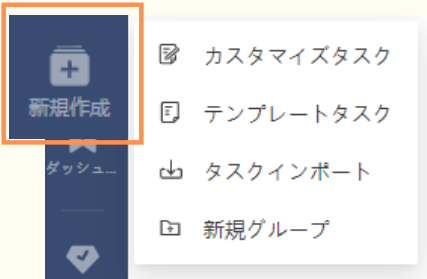
抽出開始欄：
対象先のURLまたはWebサイト名を入力し、「抽出開始」をクリックして、データ抽出のタスクの作成が始まります。

テンプレート：
octoparseで100以上のテンプレートを用意しています。テンプレートを利用して、キーワード、URLなどのパラメータを入力することで対象サイトのデータを抽出できます。

チュートリアル：
バージョン別のチュートリアルを動画の形で用意しています。(最新のチュートリアルとは更新中！最新の情報を見逃すために、チャンネルを登録してください。)

サポート：
何が困ったことやご不明な点やご意見やがございましたら、こちらからサポートに連絡できます。

① 新規作成



- **カスタマイズタスク：**
自作のワークフローでタスクを作成。
- **テンプレートタスク**
内蔵テンプレートでクリック一つでデータを抽出。
- **タスクインポート**
作成したタスクをインポートして直接にご利用。
- **新規タスクグループ**
タスクをグループ化することで、複数のタスクを同時に管理。

② ダッシュボード

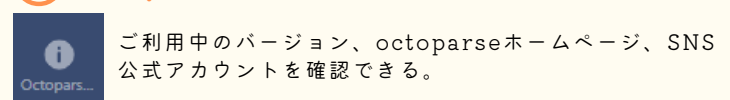


- **タスク管理：**
タスクの検索、実行、削除、複製、エクスポート。
- **タスクチェック：**
タスクの実行状態、データのチェック
- **スケジュールの設定**
- **など…**

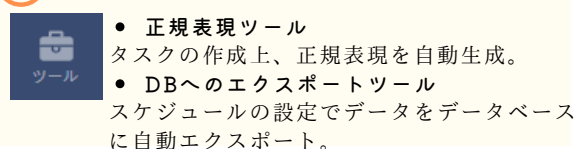
③ データサービス



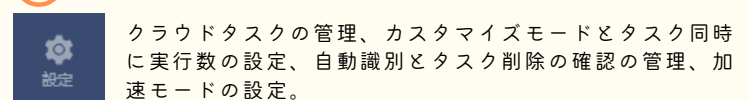
⑤ octoparseについて



④ ツール



⑥ 設定



インターフェース

The screenshot displays the Octoparse web interface. At the top, there's a navigation bar with 'ホーム' (Home) and 'ダッシュボード' (Dashboard). The main content area is divided into several sections:

- Header:** A banner for Scientific American with a 'Subscribe' button and a '内蔵ブラウザ' (Built-in Browser) button.
- Workflow Diagram:** A flowchart showing the process from 'Webページを開く' (Open Web Page) to 'ページネーション' (Pagination), 'ループアイテム' (Loop Item), 'データを抽出' (Extract Data), and 'ワークフロー' (Workflow).
- Data Preview Table:** A table with columns: 'データフィールド' (Data Field), '番号' (Number), 'Title', 'Name', 'Time', and 'オプション' (Options). It shows 5 rows of data extracted from the website.
- Settings Panel:** A 'セッティング' (Settings) panel for the current step, including a URL field, checkboxes for 'ループ内のURLを開く' (Open URL in loop) and 'JSON抽出' (JSON Extract), and a 'タイムアウト' (Timeout) dropdown set to 20 seconds.

項目	解説
内蔵ブラウザ	Chrome、Firefox、Safariなどの内蔵ブラウザを備えています。最大限に各種のニーズに対応します。
操作ヒント	ワークフローの作成に、次のステップをどのように実行しますか、を提示します。実行したい選択肢がない場合は、ワークフローの「+」をクリックし、カスタマイズで追加してください。
ワークフロー	タスクの作成は「操作ヒント」と「ワークフロー」の組み合わせで、データ抽出のロジックを作ります。ツールはそのロジックに従って、重複な作業を自動化にします。 Octoparseは、人間の行動をシミュレートしてデータを抽出します。言い換えれば、octoparseはロボットのように、重複な作業を一つ固定されたプロセス（ロジック）で自動に実行します。それで、タスクの作成の第一歩は「ロジック」を規定します。そのワークフローはある意味での「ロジック」です。
データプレビュー	[データプレビュー] 画面では、タスクの作成中、抽出できるデータ項目に含まれるデータ（項目値）を確認したり、フォーマットをしたりをすることができます。また、日付や対象サイトのURLやデータソースなどのデータ項目を追加することもできます。
セッティング	ワークフローの各ステップがどのように実行するのかを設定することができます。（例えば：Xpathの変更、タイムアウトの設定、トリガー、Ajaxの読み込みなど）

初心者向けの7つチュートリアル



7つ緑枠でのレッスンの学習↓↓↓

学習効果:

- タスクの作成する方法を分かる
- 自作のタスクで50%のWebサイトからデータを抽出できる

*タスクの作成についてもっと勉強しようとしたら、右のチュートリアルをぜひご覧ください。



7つ初心者向けのチュートリアル

レッスン1： 自動識別機能でデータを抽出する方法

もっと詳しくは：[こちらをクリック!](#)

データを抽出できるタスクとは？

重複な作業を自動化にする「ループアイテム」とは？

レッスン2： タスクを最適化する

Octoparseを使ってデータを自動検出するのは便利ですが、必要なデータが正確に検出されない場合があります。このレッスンでは、スクレイピングタスクを最適化するために適用できる簡単な修正について説明します。

もっと詳しくは：[こちらをクリック!](#)

複数のページからデータを抽出する

「もっと見る」ボタンのあるサイトのデータ抽出

詳細ページからデータを抽出

検索結果リストページと詳細ページを同時にデータ抽出方法

キーワードのリストを入力して検索結果をスクレイピングする方法

URLリストからデータを一括的に抽出

テーブル、表組からを抽出する

既存データ項目から新たなデータ項目を追加する方法

7つ初心者向けのチュートリアル

レッスン3：データフィールドを調整する

きれいなデータを抽出するために、データフィールドを調整する方法

もっと詳しくは：[こちらをクリック！](#)

[正規表現の使い方と利用タイミング](#)

[ツール内の正規表現ツールで正規表現を生成](#)

[正規表現の使い方と利用タイミング](#)

レッスン4：タスクを実行する前にタスクをチェックする方法：

もっと詳しくは：[こちらをクリック！](#)

レッスン5：データを取得する

もっと詳しくは：[こちらをクリック！](#)

レッスン6：

データの定期抽出、自動データエクスポート、APIなどより効果的かつ効率的なデータ抽出する方法を紹介

もっと詳しくは：[こちらをクリック！](#)

[Octoparseからタスクをエクスポートする](#)

[取得したデータをCSV、Excel、JSON、HTML形式に変更する方法](#)

レッスン7：

ここまで、octoparseの基本機能でタスクを作成方法を紹介しました。いかがでしょうか？

さあ！下記のURLからタスクの作成を[こちらから](#)練習しましょう！



どんどん中級者レベルへ



複雑なサイトにもデータをうまく抽出できる

主な内容:

- Xpathの書き方 (データをより正確的に抽出できる)
- 様々なデータ抽出要望に応じてツールのもっと多くの機能を学び

学習効果:

90%以上のWebサイトのデータをより正確的に抽出できる



ツール利用上の豆知識

データ抽出

それぞれのサイトの構造により、データをうまく抽出するために、octoparseでの対応法(備えた機能)を紹介します。

条件を判断してデータを抽出する

画像スライダーから画像URLを抽出する

ドロップダウンメニューでの特定、あるいは複数のオプションからデータを抽出する方法

マウスオーバーで表示されるデータからデータを抽出する方法

「Webページを開く」のところ、1つのURLを複数のURLに変更する方法

ページ全体を再読み込みせずにWebページの一部情報を更新するサイト(Ajax自動検出)からデータの抽出方法

サイトの検索欄にキーワードを抽出する方法

ログインが必要なサイトからデータ取得

取得したデータを統合する

特定した条件に満たすデータだけを抽出する(トリガー)

条件分岐: 条件を判断してデータを別々に抽出する方法

クラウド抽出

クラウド抽出の実行状況と実行レコード

クラウド「タスク分割」とは何ですか？
(クラウド抽出の高速化)

データ抽出の安定性の改善

広告ブロック&キャッシュクリア

Octoparseブロッキング防止設定

Xpathの書き方

OctoparseはXPathを自動的に生成し、自分で書く必要はありませんがページネーションが突然機能できなくなったり、取得したデータの位置がずれてしまったりする場合があります。その時はXPathを変更して、データをより正しく見つける必要があります。

どんな時にXpathの変更が必要？

- データの位置がよく変わる
- ページ送りボタンが正しくない
- 一覧リストのリンクをクリックできない
- ドロップダウンメニューの特定データが必要

XPathによる特定データの指定

XPath基礎編(1)-XPathの基本概念

XPath基礎編(2)-XPathの書き方

XPath基礎編(3)-XPathによく使う関数

XPath応用編 (1) - ページ送りボタンを指定する方法

XPath応用編 (2) - テーブルで表示されるデータを指定する方法

XPath応用編 (3) - リストで表示されるデータを指定する方法

Xpath

おめでとうございます！

上級者をチャレンジ！

上級者向け

API

APIでoctoparseでの操作をしなくても、抽出されたデータ（クラウドデータのみ）、タスク情報、さらにはタスクを制御できるため、独自のアプリケーションとのシームレスなデータ統合が可能になる。

APIを利用するためにタスクIDを調べる方法

API と接続する方法

Advanced API の使い方

frame：ページ内に別のページをコンテンツとして埋め込む要素をスクレイピングする方法。

差分抽出：ニュースポータルやフォーラムなど常に情報を更新するサイトからデータを抽出する方法。

並行処理：クラウドまたはローカルマシンで同時に実行されている方法。

その他の上級機能



ここまで、上級レベルのチュートリアルを勉強して、大変お疲れ様でした。octoparse使い方の勉強はいかがでしたか？ツールを使いこなしたのでしょうか？もし、自分のタスク作成力に自信を持ちましたら、ぜひoctoparseの資格認証試験をチャレンジしましょう。

申し込み方法：

「資格認定の申し込み」という件名のメールを support@octoparse.com に送って、試験を申し込むことができます。

試験に合格する方は：

- ①「octoparseスクレイピング専門士」資格認定証
- ②octoparseタコちゃんキャラクターをもらえます。