

## Web Scraping

[futuresbydesign.net](https://futuresbydesign.net)

### Eine zugängliche Möglichkeit der Datenerfassung für jedes kleine und mittlere Unternehmen

#### Grundlagen des Web Scraping und warum es einen Mehrwert für Ihr Unternehmen generiert

Heutzutage, da sich die Wirtschaft in einem beeindruckenden Tempo digitalisiert, wird datengestützte Entscheidungsfindung für jedes Unternehmen immer wichtiger, um seinen Wettbewerbsvorteil zu wahren. Die wichtigste Zutat hierfür sind natürlich Daten. Dies bezieht sich sowohl auf Daten über die internen Unternehmensprozesse als auch auf Daten außerhalb des Unternehmens.

Die Organisation der internen Datenprozesse wurde bereits an anderer Stelle dokumentiert, in diesem Artikel konzentrieren wir uns daher auf die Erfassung von externen Daten.

Die Beschaffung von Daten kann ein kostspieliges Unterfangen sein, sowohl in Bezug auf Zeit als auch auf Geld, insbesondere für kleine und mittlere Unternehmen (KMU). In diesem Artikel befassen wir uns mit **Web Scraping** als Mittel zur Erfassung wertvoller externer Daten.

#### Das World Wide Web als Datenquelle

Web Scraping ist die systematische und automatisierte Extraktion von Daten aus Websites. Mit Web Scraping können Unternehmen eine der bedeutendsten Datenquellen anzapfen, die es gibt: das Internet. Im Internet hinterlassen Kunden und Konkurrenten Daten über ihr Verhalten, Erkenntnisse und Interaktionen, die für bestimmte Entscheidungen, die ein Unternehmen treffen muss, nützlich sein können. Es ist eine perfekte Datenquelle für KMU da es eine effiziente und kostengünstige Datenerfassung ermöglicht.

Vor allem für KMU, die sich für die Arbeit mit Daten interessieren oder erforschen wollen, wie Data Science ihrem Unternehmen helfen kann, bietet Web Scraping - wenn es auf spezifische organisatorische Herausforderungen und Entscheidungen angewandt wird - eine niedrige Hürde, um mit dem Sammeln externer Daten zu beginnen und die internen Data-Science-Fähigkeiten zu entwickeln.



## Web Scraping – die Grundlagen

Praktisch jeder Text, jede Zahlenreihe oder jedes Bild, das Sie auf einer beliebigen Webseite sehen, kann gescraped werden und somit als Daten eingestuft werden, die gesammelt, analysiert und in wertvolle Erkenntnisse umgewandelt werden können, die bei der Entscheidungsfindung helfen. Die einzige Bedingung ist, dass die Daten über mehrere Seiten einer bestimmten Website hinweg konsistent dargestellt werden sollten.

Zum Beispiel wird der Preis eines Amazon.com-Produkts immer an derselben Stelle für verschiedene Produkte angezeigt und kann daher abgefragt werden:

- 1) Crawling – d. h. "Durchsuchen der verschiedenen Seiten, die Sie interessieren"; und
- 2) Scraping – d. h. "Auswählen und Speichern der Daten, an denen Sie interessiert sind".

Diese Prozesse werden oft verwechselt, aber sie sind sehr unterschiedlich:

Das Web-Crawling sorgt für das Laden der verschiedenen Seiten, von denen die Daten ausgelesen werden müssen. Zum Beispiel alle verschiedenen Brettspiel-Produktseiten auf Amazon.com. Anschließend extrahiert Web Scraping die gewünschten Informationen von all diesen Seiten, wie z. B. Preise, Namen, Bewertungen usw., und fasst sie in einem nützlicheren Format zusammen (z. B. Excel-Arbeitsmappen). Danach sind die Daten bereit für die Analyse.

## Web Scraping in der Praxis

Da mittels Web Scraping fast alle Daten gesammelt werden können, die online zu finden sind, besteht der wichtigste Aspekt darin, sich Gedanken darüber zu machen, wie man eine bestimmte Art von Daten nutzen kann. Die wichtigste Erkenntnis dabei ist: Alles, was Sie online sehen, kann gescraped werden. Vergewissern Sie sich stets, dass Sie Daten sammeln, die Ihnen dabei helfen, bestimmte Entscheidungen zu treffen. Ein Beispiel für ein Unternehmen, das Web Scraping für sein Kerngeschäft nutzt, ist HiQ. HiQ sammelt öffentlich zugängliche Mitarbeiterdaten von LinkedIn (ein Gericht hat entschieden, dass dies im Jahr 2019 legal ist). Das Unternehmen nutzt diese Daten, um die Personalabteilungen seiner Kunden bei der Verwaltung ihrer Belegschaft zu unterstützen. Sie entwickeln Skill-Maps für Mitarbeiter und erstellen Fluktuationsmodelle, um vorherzusagen, wann Mitarbeiter wahrscheinlich ihren Arbeitsplatz verlassen werden. Auf diese Weise kann HiQ voraussagen, ob ein Qualifikationsdefizit entstehen könnte, und das Fluktuationsrisiko schon Monate im Voraus abschätzen, wodurch die Kunden viel Geld und Zeit sparen.

Ein weiteres interessantes Beispiel ist Proven, ein Hautpflegeunternehmen, das Kundenrezensionen auswertet, um sein Produktangebot zu personalisieren. Web Scraping hilft ihnen dabei, herauszufinden, was sie tatsächlich über die Erfahrungen ihrer Kunden mit ihren Produkten wissen. Da es sich dabei um eine große Datenmenge handelt, werden die gesammelten Bewertungen in einen Algorithmus für maschinelles Lernen eingespeist, um interessante Korrelationen zu entdecken, die wiederum dazu beitragen, das Kundenerlebnis und die Rentabilität zu steigern.



## Was kann ich mit Web Scraping machen?

Einige Beispiele für auswertbare Daten und ihre möglichen Anwendungen sind:

- *Aktienkurse:* Aktienkurse können von Finanz-Websites wie [finance.yahoo.com](https://finance.yahoo.com) gesammelt werden, um bessere Anlageentscheidungen zu treffen.
- *E-Mail-Adressen/Telefonnummern:* E-Mail-Adressen und Telefonnummern können von Gelbe-Seiten-Websites wie [Yellowpages.com](https://yellowpages.com) gesammelt werden, um bessere Verkaufserfolge zu erzielen.
- *Adressen:* Geschäftsadressen können aus Unternehmensregistern wie [e-justice.europa.eu](https://e-justice.europa.eu) entnommen werden, um eine Liste von Geschäftsstandorten zu erstellen und so die geografische Landschaft der Konkurrenz zu ermitteln.
- *Produktinformationen:* Produktinformationen wie Preise, Namen, Bewertungen und Lieferzeiten können von E-Commerce-Websites wie [Amazon.com](https://amazon.com) ausgelesen werden, um eine Analyse der Wettbewerber durchzuführen.
- *Beiträge in sozialen Medien:* Beiträge in sozialen Medien können von Websites wie [Twitter.com](https://twitter.com) gesammelt werden, um eine Stimmungsanalyse durchzuführen und die öffentliche Meinung zu bestimmten Themen zu verstehen.
- *Menüinformationen:* Restaurant-Menüinformationen können von Websites abgerufen werden, um dynamische Preise für Ihre eigenen Gerichte zu erstellen.

## Wie kann ich mit Web Scraping beginnen?

Die obigen praktischen Beispiele zeigen, dass Web Scraping in einer Vielzahl von Bereichen eingesetzt werden kann, die für ein Unternehmen von Interesse sind. Letztendlich ist es ein wichtiges Werkzeug, das jedes KMU nutzen kann, um an wertvolle, öffentlich verfügbare Daten zu gelangen. Womit sollte man beginnen?

- *Schritt 1:* Bestimmen Sie einen Schwerpunkt - was ist eine spezifische Herausforderung innerhalb Ihres Unternehmens, die mit externen Daten gelöst werden könnte?
- *Schritt 2:* Erstellen Sie eine Übersicht darüber, welche Daten Sie konkret benötigen.
- *Schritt 3:* Prüfen Sie, ob die Website das Scrapen der betreffenden Daten erlaubt - einige Websites verfügen über einen Kopierschutz bzw. ein Datenbankrecht.
- *Schritt 4:* Verwenden Sie ein bestimmtes Scraping-Tool, um die Daten wie in Schritt 2 beschrieben zu sammeln. Beispiele für benutzerfreundliche Tools sind ParseHub und ScrapingBee.
- *Schritt 5:* Benötigen Sie Unterstützung beim Web Scraping? Holen Sie sich Hilfe von Daten-Experten und kontaktieren Sie die JADS unter [info@jadsmkdbdatalab.nl](mailto:info@jadsmkdbdatalab.nl)!