Crawler User Agent List: Expertenübersicht für Profis

Category: SEO & SEM
geschrieben von Tobias Hager | 5. Januar 2026

Crawler User Agent List: Expertenübersicht für Profis

Wenn du glaubst, Suchmaschinen-Crawler seien nur freundliche Bots, die brav deine Website durchforsten, dann hast du noch nicht mit den echten Tech-Profis gesprochen. Hinter den Kulissen tummeln sich unzählige User Agents – manche freundlich, manche aggressiv, und einige völlig unverständlich. Ohne eine klare Übersicht über diese Crawler-User-Agents bist du im Blindflug

unterwegs. Dabei handelt es sich um das digitale Äquivalent zum Türsteher, der entscheidet, wer rein darf und wer nicht. Und wenn du nicht weißt, wer vor deiner Tür steht, kannst du deine Security, SEO-Strategie und Server-Performance gleich vergessen. Willkommen zur ultimativen Expertenübersicht für Crawler User Agents – weil nur die Kenntnis der richtigen User Agents dich vor bösen Überraschungen schützt.

- Was sind Crawler User Agents und warum sie für SEO & Security entscheidend sind
- Die wichtigsten User Agents im Überblick – für Google, Bing, Baidu & Co.
- Wie du eine vollständige Crawler-User Agent List erstellst
- Unterscheidung zwischen freundlichen Bots, Spidern und bösartigem Traffic
- Tools und Techniken zum Monitoring und Blockieren unerwünschter User Agents
- Best Practices für die Handhabung von User Agents in Server- und Firewall-Konfigurationen
- Automatisierte Update-Strategien für deine User Agent Liste
- Risiken und Fallstricke: Was passiert, wenn du User Agents falsch handhabst
- Was viele Agenturen verschweigen – und warum du es wissen solltest
- Fazit: Warum eine solide User Agent Übersicht dein bester Freund im digitalen Dschungel ist

In der Welt des Webs ist der User Agent mehr als nur eine technische Fußnote. Er ist dein erster Anlaufpunkt, um zu verstehen, wer genau gerade deine Server belagert: Googlebot, Bingbot, Baiduspider oder gar fiese Scraper, die nur auf Schaden aus sind. Ohne eine klare, aktuelle Liste der User Agents kannst du keine gezielten Maßnahmen ergreifen – sei es beim Crawling-Management, beim Schutz vor Overload oder bei der Optimierung deiner SEO-Strategie. Denn wer nicht weiß, wer vor der Tür steht, kann auch keine intelligenten Entscheidungen treffen.

Viele Betreiber und Agenturen glauben, die wichtigsten User Agents stünden in der Google Search Console oder in den Server-Logs. Das stimmt nur halb. Denn die Listen der Crawlers verändern sich ständig: Neue Bots kommen, alte verschwinden, und manchmal tauchen sogar gefälschte User Agents auf, die nur dazu dienen, dich zu täuschen oder Server-Ressourcen zu klauen. Für Profis ist es daher essenziell, eine kontinuierliche Übersicht zu haben, um Bedrohungen frühzeitig zu erkennen und das Crawl-Verhalten zu steuern.

Ein weiterer Punkt: Nicht alle User Agents sind gleich. Während Googlebot und Bingbot in der Regel vertrauenswürdig sind, gibt es eine Vielzahl von Bots, die nur so tun, als seien sie legitime Crawler. Manche sind Scraper, die Inhalte klauen, andere sind Angreifer, die versuchen, Sicherheitslücken auszunutzen. Deshalb ist das Wissen um die genaue User Agent Liste die Grundvoraussetzung, um zwischen gut und böse zu unterscheiden und entsprechend zu reagieren.

Was sind Crawler User Agents und warum sie für SEO & Security entscheidend sind

Jede Website wird von einer Vielzahl von Crawler-User-Agents besucht. Diese sind in der Regel im HTTP-Header unter dem Feld „User-Agent“ enthalten. Sie geben an, welcher Bot oder welche Anwendung gerade deine Seite anfragt. Für SEO bedeutet das: Mit der richtigen Erkennung kannst du den Traffic filtern, Crawling-Brennpunkte identifizieren und die Indexierung steuern. Für die Sicherheit ist es noch wichtiger: Du kannst unerwünschte Bots blockieren, bevor sie deine Server lahmlegen oder Daten klauen.

Der User Agent ist im Grunde eine Art digitale Visitenkarte. Er enthält oft den Namen des Bots, die Version, manchmal auch eine Kontaktadresse oder Hinweise auf den Zweck. Googlebot zum Beispiel trägt den User Agent „Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)“. Bingbot ist ähnlich, mit „Mozilla/5.0 (compatible; Bingbot/2.0; +http://www.bing.com/bingbot.htm)“. Doch in der Praxis gibt es zahlreiche Variationen, die sich je nach Bot-Version, Plattform oder sogar gefälschten User Agents unterscheiden.

Für SEO-Experten ist es essenziell, diese User Agents zu kennen, um Crawling-Fehler zu identifizieren, Crawl-Budget effizient zu nutzen und Duplicate Content zu vermeiden. Für IT- und Security-Profis ist es eine Waffe gegen Bots, die nicht nur unnötig Ressourcen verschwenden, sondern auch Angriffe starten. Mit einer präzisen User Agent List kannst du deine Server gegen unerwünschten Traffic absichern und deine Website gegen Missbrauch schützen.

Die wichtigsten User Agents im Überblick – für Google, Bing, Baidu & Co.

Der Markt der Crawler ist groß, aber einige User Agents dominieren die Szene. Im Folgenden eine Übersicht der wichtigsten, die du kennen musst – inklusive ihrer typischen Merkmale und Einsatzbereiche.

- Googlebot: Der unangefochtene Spitzenreiter. Er crawlt regelmäßig, erkennt neue Inhalte sofort und ist maßgeblich für dein Ranking verantwortlich. User-Agent: „Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)”
- Bingbot: Der Schreck der SEO-Profis, die auf Microsoft setzen. Auch er ist zuverlässig, aber weniger aggressiv. User-Agent: „Mozilla/5.0 (compatible; Bingbot/2.0; +http://www.bing.com/bingbot.htm)”

- Baiduspider: Der chinesische Gigant – wächst rasant, vor allem bei internationalen Seiten. User-Agent: „Mozilla/5.0 (compatible; Baiduspider/2.0; +http://www.baidu.com/search/spider.html)“
- YandexBot: Der russische Player, zunehmend relevant für europäische Seiten. User-Agent: „Mozilla/5.0 (compatible; YandexBot/3.0; +http://yandex.com/bots)“
- DuckDuckGo Bot: Der Datenschutzfreund, der auf Privatsphäre setzt. User-Agent: „DuckDuckGo/7.0; (+https://duckduckgo.com/duckduckgo-help-pages/duckduckgo-bots/)“
- Google Ads Bot: Für das Crawling von Anzeigen und Shopping-Feeds. User-Agent: „Mozilla/5.0 (Linux; Android 10; SM-G960U) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/89.0.4389.105 Mobile Safari/537.36 (compatible; Google AdWords; +https://www.google.com/adsbot.html)“

Neben diesen bekannten Vertretern existiert eine Vielzahl von spezialisierten, manchmal versteckten User Agents. Manche sind reine Scraper, andere versuchen, sich als legitime Bots auszugeben. Hier gilt: Je besser du deine Liste pflegst, desto mehr Kontrolle hast du über dein digitales Terrain.

Wie du eine vollständige Crawler User Agent List erstellst

Eine aktuelle und vollständige User Agent Liste ist das Herzstück der professionellen Server- und SEO-Strategie. Doch wie kommt man an eine solche Liste? Die Kunst besteht darin, kontinuierlich aktuelle Daten zu sammeln, zu filtern und zu pflegen.

Der erste Schritt: Logfile-Analyse. Server-Logs enthalten alle HTTP-Anfragen, inklusive User Agents. Durch das Parsen dieser Logs kannst du alle Bots identifizieren, die deine Seite besucht haben. Wichtig ist, diese Logfiles regelmäßig zu exportieren und automatisiert auszuwerten. Tools wie GoAccess oder ELK-Stacks helfen, die Daten verständlich zu visualisieren.

Zweitens: Crawling-Tools. Nutze spezialisierte Software wie Screaming Frog, Sitebulb oder DeepCrawl, die dir eine Übersicht der aktuellen User Agents liefern. Viele dieser Tools erlauben auch das Blockieren oder Testen spezifischer User Agents direkt in der Plattform.

Drittens: externe Datenquellen. Es gibt regelmäßig gepflegte Listen, beispielsweise auf GitHub oder in SEO-Communities. Diese Listen sind eine gute Basis, sollten aber stets durch deine eigenen Daten ergänzt werden, um Lücken zu schließen.

Viertens: Automatisierung. Richte dir automatisierte Prozesse ein, die regelmäßig Logfiles analysieren und deine User Agent Daten aktualisieren. So bleibst du immer auf dem Laufenden und kannst schnell auf neue Bedrohungen

oder Veränderungen reagieren.

Unterscheidung zwischen freundlichen Bots, Spidern und bösartigem Traffic

Nicht jeder User Agent ist gleich. Während Googlebot und Bingbot in der Regel vertrauenswürdig sind, gibt es unzählige andere, die nur scheinbar harmlos erscheinen. Manche sind freundlich, andere sind reines Spamming oder sogar Cyberkriminelle in Verkleidung.

Legitime Bots haben klare, bekannte User Agents und verhalten sich vorhersehbar. Sie folgen Robots.txt, respectieren Noindex-Flags und identifizieren sich in der Regel eindeutig. Gefälschte Bots hingegen tarnen sich als Googlebot oder andere bekannte Crawler, verwenden manipulierte User Agents und ignorieren Anweisungen.

Auf der anderen Seite steht der bösartige Traffic: Scraper, Bots, die Inhalte kopieren, Brute-Force-Angriffe, SQL-Injections oder Distributed Denial of Service (DDoS). Hier sind User Agents oft gefälscht, aber auch manchmal sehr kreativ verschleiert. Das Erkennen erfordert eine Mischung aus User Agent-Analyse, IP-Checks, Verhaltenserkennung und Traffic-Monitoring.

Das Ziel: eine robuste, dynamische User Agent-Blacklist, die regelmäßig gepflegt wird. So kannst du unerwünschte Besucher identifizieren, blockieren und deine Ressourcen für echte Nutzer und legitime Crawler freihalten.

Tools und Techniken zum Monitoring und Blockieren unerwünschter User Agents

Der technische Schutz deiner Website beginnt bei der richtigen Infrastruktur. Mit den richtigen Tools kannst du unerwünschte User Agents in Echtzeit erkennen und blockieren. Hier einige bewährte Ansätze:

- Firewall-Regeln: Nutze Web Application Firewalls (WAF) oder Server-Firewalls, um bekannte bösartige User Agents direkt zu blockieren. Viele Anbieter wie Cloudflare, Sucuri oder ModSecurity ermöglichen die einfache Konfiguration entsprechender Regeln.
- Server-Konfiguration: In Apache kannst du beispielsweise mit mod_rewrite oder mod_security User Agents filtern. In Nginx erfolgt die Blockierung durch entsprechende Direktiven in der Konfiguration.
- Monitoring-Tools: Nutze Logfile-Analysetools, um ungewöhnliche Traffic-Muster zu erkennen. Tools wie Grafana, Kibana oder Datadog helfen,

verdächtigen Traffic zu visualisieren und Alerts zu setzen.
- Bot-Management-Plattformen: Professionelle Dienste wie Distil Networks, PerimeterX oder Radware bieten intelligente Bot-Detection, Blockierung und Traffic-Analyse in Echtzeit – inklusive Machine Learning-gestützter Erkennung.

Wichtig: Blockieren ist nicht alles. Es sollte eine Balance zwischen Sicherheit und Nutzererfahrung bestehen. Übermäßiges Blockieren kann legitimen Traffic behindern. Daher ist eine kontinuierliche Feinjustierung notwendig.

Best Practices für die Handhabung von User Agents in Server- und Firewall-Konfigurationen

Bei der technischen Umsetzung gilt: Standardisierte, nachvollziehbare Regeln. Hier einige Best Practices:

- Whitelist-Ansatz: Erlaube bekannte, vertrauenswürdige User Agents explizit und blockiere alles andere. Das minimiert Fehlalarme.
- Blacklist-Ansatz: Blockiere bekannte schädliche User Agents, aber immer in Kombination mit anderen Sicherheitsmaßnahmen.
- Regelmäßige Updates: Aktualisiere deine Listen regelmäßig, um neue Bedrohungen zu erkennen. Automatisierte Prozesse helfen dabei.
- Authentifizierung und CAPTCHAs: Für kritische Aktionen solltest du User-Agent-Checks mit anderen Sicherheitsmaßnahmen kombinieren.
- Logging und Audit: Dokumentiere alle Blockaktionen und analysiere regelmäßig, ob legitimer Traffic versehentlich blockiert wurde.

Automatisierte Update-Strategien für deine User Agent Liste

In der schnelllebigen Welt der Bots reicht eine statische Liste nicht aus. Automatisierte Updates sind Pflicht, um immer auf dem neuesten Stand zu bleiben. Hier einige Strategien:

- Web-Scraping und Feed-Integrationen: Nutze APIs oder RSS-Feeds von bekannten Threat-Intelligence-Anbietern und SEO-Communities, um deine Listen automatisch zu aktualisieren.
- Logfile-Analysen: Automatisiere die Analyse deiner Server-Logs, um neue

User Agents zu identifizieren und in die Blocklisten aufzunehmen.
- Machine Learning: Entwickle Modelle, die Traffic-Muster erkennen und verdächtige User Agents automatisch klassifizieren.
- Regelmäßige Audits: Plane automatisierte Checks, die deine Listen auf Inkonsistenzen oder veraltete Einträge prüfen.

Risiken und Fallstricke: Was passiert, wenn du User Agents falsch handhabst

Falsche Annahmen bei der Handhabung von User Agents können bitteres Erwachen bedeuten. Blockierst du zu viele User Agents, riskierst du, legitimen Googlebot oder andere wichtige Crawler auszuschließen. Das führt zu Indexierungsproblemen, schlechter Sichtbarkeit und Rankings. Umgekehrt: Wenn du nur auf Listen vertraust, die veraltet sind, lässt du schädliche Bots ungehindert rein, was Sicherheitslücken öffnen kann.

Ein weiterer Fallstrick: User Agents sind leicht fälschbar. Manche Angreifer tarnen sich mit gefälschten User Agents, um Sicherheitsmaßnahmen zu umgehen. Daher darf die User-Agent-Analyse niemals alleinstehend, sondern nur ein Teil deiner Sicherheitsstrategie sein. Kombination mit IP-Checks, Verhaltenserkennung und CAPTCHA ist Pflicht.

Schließlich: Übermäßige Blockaden können die Nutzererfahrung beeinträchtigen. Wenn legitime Crawler durch unpräzise Regeln blockiert werden, leidet die SEO-Performance. Es ist daher essenziell, regelmäßig zu prüfen, ob die Blocklisten noch sinnvoll sind.

Was viele Agenturen verschweigen – und warum du es wissen solltest

Viele Agenturen verschweigen, dass sie oft nur mit veralteten oder unvollständigen User Agent Listen arbeiten. Das kann fatale Folgen haben: Sie blockieren wichtige Bots, ignorieren neue Bedrohungen oder setzen auf Standardregeln, die längst überholt sind. Das Ergebnis: Sicherheitslücken, SEO-Verluste und unnötige Serverbelastung.

Hinzu kommt: Das Erstellen und Pflegen einer aktuellen User Agent List ist kein Hexenwerk, sondern eine dauerhafte Aufgabe. Wer das verschweigt, verkauft dir nur halbe Wahrheiten. Ein professioneller Ansatz erfordert kontinuierliches Monitoring, automatisierte Updates und eine tiefgehende Analyse. Alles andere ist nur Augenwischerei.

Wenn du also in der SEO- oder Security-Beratung unterwegs bist, stelle immer die Frage: Wie aktuell ist eure User Agent Liste? Welche Quellen nutzt ihr? Und wie pflegt ihr sie? Nur so kannst du sicherstellen, dass du nicht nur auf dem Papier, sondern auch in der Praxis gut aufgestellt bist.

Fazit: Warum eine solide User Agent Übersicht dein bester Freund im digitalen Dschungel ist

In der Welt der Web-Optimierung ist der User Agent das unsichtbare Tor zur Kontrolle über deinen Traffic. Wer hier nicht mit offenen Augen agiert, läuft Gefahr, von Bots überrollt, von Angriffen ausgeknockt oder in der Indexierung sabotiert zu werden. Eine aktuelle, gepflegte Liste der User Agents ist der Grundpfeiler jeder professionellen Server- und SEO-Strategie.

Ob du nun deine Server gegen bösartige Bots schützen willst, dein Crawl-Management optimieren oder einfach nur einen klaren Überblick behalten möchtest – das Wissen um die User Agents ist dein Schlüssel dazu. Ohne diese Übersicht bist du im digitalen Dschungel verloren, ohne Orientierung und ohne Kontrolle. Also: Mach dich schlau, halte deine Listen aktuell und bleib immer einen Schritt voraus. Denn in der Welt der Bots zählt nur, wer seine Tür kennt – und sie auch kontrolliert.