

## Googleが推奨するHTTPキャッシュはETag (Last-Modified/If-Modified-Sinceもサポートするけど)

Google は、クローラーが HTTP キャッシュをどのように処理するかを説明するセクションをクローラーの技術ドキュメントに追加した。また、HTTP キャッシュの仕組みをさらに詳しく解説する記事を検索セントラルブログでも公開した。

Google は、クローラーが HTTP キャッシュをどのように処理するかを説明するセクションをクローラーの技術ドキュメントに追加しました。また、HTTP キャッシュの仕組みをさらに[詳しく解説する記事](#)を検索セントラルブログでも公開しました。

### HTTP キャッシュの利点

キャッシングは、サーバーとクライアント(ブラウザとクローラー)の両方にいくつかの利点をもたらします。

#### サーバー側の利点

- **コンピューティングリソースの削減:** クライアントのリクエストがキャッシュされたバージョンと一致する場合、サーバーはコンテンツを生成するためのコンピューティングリソースを消費する必要がない。これはサーバーのコスト削減につながる
- **データ転送量の削減:** キャッシュされたバージョンが利用可能な場合、サーバーは HTTP 本体を転送する必要がないため、データ転送量を節約できる。これもサーバーのコスト削減につながる

特に、個々の URL でコンテンツがめったに変わらない大規模なサイトでは、ローカルでのキャッシングを許可すると、サイトのクロール効率が向上する場合があります。

#### クライアント側の利点

- **読み込み時間の高速化:** キャッシングにより、再訪問時のページ読み込みが非常に高速になる——コンテンツがクライアントの内部キャッシュから取得される場合、データ転送が発生しないため、読み込みが速くなる
  - **リソース消費量の削減:** クライアントは、コンテンツがすでにキャッシュされている場合、再度ダウンロードする必要がないため、リソースを節約できる
  - **帯域幅の節約:** キャッシングは、クライアントとサーバーの両方にとって貴重な帯域幅を大幅に節約する
- キャッシングは、クライアントの読み込み時間の高速化にもつながり、結果としてユーザー体験を向上させます。

### Googlebot がサポートする HTTP キャッシュ

Googlebot は、次の 2 種類の HTTP キャッシュの技術をサポートします。

- ETag と If-None-Match の組み合わせ
- Last-Modified と If-Modified-Since の組み合わせ

Google としては、ETag / If-None-Match の構成を強く推奨しています。


こちらの方がエラーや設定ミスを起こしにくいからです。

## Googleが推奨するHTTPキャッシュはETag（Last-Modified/If-Modified-Sinceもサポートするけど）

Google は、クローラーが HTTP キャッシュをどのように処理するかを説明するセクションをクローラーの技術ドキュメントに追加した。また、HTTP キャッシュの仕組みをさらに詳しく解説する記事を検索セントラルブログでも公開した。

しかしながら、可能な限り Last-Modified / If-Modified-Since との併用が望まれます。  
こちらしかサポートしないシステムがウェブに存在します。

それぞれのキャッシュ方式の詳細な構成方法は技術ドキュメントとブログ記事を参照してください。

 すぎ注:この記事を書いている時点では、どちらも英語のみ

- [HTTP Caching – Google Search Central Documentation](#)
- [Crawling December: HTTP caching – Google Search Central Blog](#)

## robots.txtが503エラーを返すとどうなる？→最初の12時間はクローリング完全停止、その後は？

robots.txt の取得にエラーが発生した際の Googlebot の挙動を説明する技術ドキュメントを Google は更新した。

robots.txt の取得にエラーが発生した際の Googlebot の挙動を説明する[技術ドキュメント](#)を Google は更新しました。

### robots.txt が 5xx エラーを返す場合

具体的には、Googlebot による robots.txt の取得時に 5xx エラーが返されたときの処理の説明が変わりました。

とはいえ、従来の処理から変更があったわけではありません。

説明が複数箇所に散らばっていてわかりにくかったため、一箇所にまとめました。

robots.txt が 5xx のサーバーエラーを返した場合、Googlebot は次のように対応します。

1. For the first 12 hours, Google stops crawling the site but keeps trying to fetch the robots.txt file.
2. If Google can't fetch a new version, for the next 30 days Google will use the last good version, while still trying to fetch a new version. A 503 (service unavailable) error results in fairly frequent retrying. If there's no cached version available, Google assumes there's no crawl restrictions.
3. If the errors are still not fixed after 30 days:
  - If the site is generally available to Google, Google will behave as if there is no robots.txt file (but still keep checking for a new version).
  - If the site has general availability problems, Google will stop crawling the site, while still periodically requesting a robots.txt file.

この記事を書いている時点で日本語ドキュメントは未更新のため、僕が訳します。

1. 最初の 12 時間は、サイトのクローリングを停止しますが、robots.txt ファイルの取得は継続して試みます。
2. 新しいバージョンを取得できない場合、その後 30 日間は、最後の有効なバージョンを使用し、引き続き新しいバージョンの取得を試みます。503(サービス利用不可)エラーが発生した場合、かなり頻繁に再試行されます。キャッシュされたバージョンがない場合、Google はクローリング制限がないものと見なします。
3. 30 日経ってもエラーが修正されない場合:
  - サイトが Google で一般的に利用可能な場合、Google は robots.txt ファイルがないかのように動作します(ただし、引き続き新しいバージョンを確認します)。
  - サイトに一般的な利用状況の問題がある場合、Google はサイトのクローリングを停止しますが、robots.txt ファイルは定期的にリクエストします。

## robots.txtが503エラーを返すとどうなる？→最初の12時間はクローリング完全停止、その後は？

robots.txt の取得にエラーが発生した際の Googlebot の挙動を説明する技術ドキュメントを Google は更新した。

簡潔にまとめるとこうなります。

1. 最初の 12 時間はサイト全体のクローリングを停止
2. (robots.txt を継続して取得できず)その後、30 日間:
  - 最新のキャッシュの robots.txt があればそれを使用
  - キャッシュした robots.txt がない場合は、制限なくクローリング
3. 30 日経過後:
  - サイトの大部分がアクセス可能 → 制限なくクローリング
  - サイトの大部分がアクセス不可 → クローリング停止

どの段階でも robots.txt の再取得は繰り返し試行されます。

「最初の 12 時間はサイト全体のクローリングを停止」という処理は、どのドキュメントに記載されてたでしょうか？

僕は知らなかったです。

robots.txt が長期間エラーを返す場合は、通常のページの利用状況によってクローリングするかしないかが分かります。

“generally available”を「サイトの大部分」と僕は解釈しましたが、[ゲイリーは「重要なページ \(some other important pages\)」](#)と言っていました。

ここは、ちょっと曖昧です。

いずれにしても、robots.txt の取得がエラーを返すのは悪い状況です。

一刻も早く解消しなければなりません。

## Google、会話型検索を提供するAIモードを開発中か

Googleは、より会話的でインタラクティブな検索体験を提供することを目指し、新たな検索機能として AI Mode (AI モード)を開発中らしい。

Googleは、より会話的でインタラクティブな検索体験を提供することを目指し、新たな検索機能として AI Mode (AI モード)を開発中であると [The Information](#) が報じました。

### 対話型検索へのアプローチ

AI モードと呼ばれるこの新機能は、検索されたトピックの簡単な要約を提供する既存の AI Overview をさらに進化させたものです。

AI モードは、Google の Gemini チャットボットのウェブバージョンに似た新しいインターフェースを提供します。

### AI モードへのアクセス方法


AI モードは、検索結果ページの上、検索ボックスの下にある、「すべて」、「画像」、「動画」などのタブの左側に表示されると言われています。

ユーザーが AI モードを選択すると、新しいインターフェースに移動します。

このインターフェースは、Gemini ウェブクライアントに似た外観になるようです。

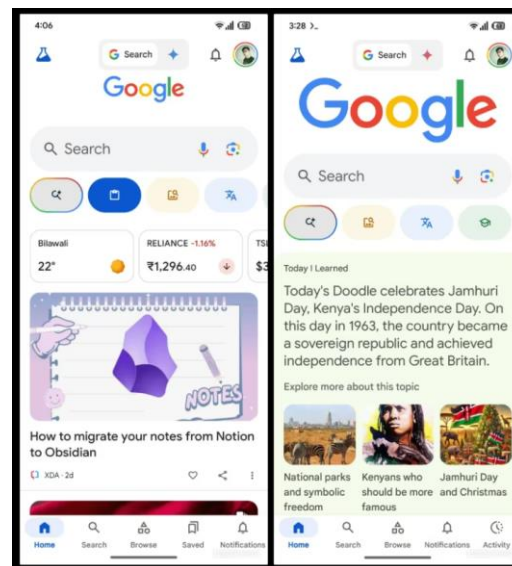
[Android Authority](#) が、Google アプリの APK (アプリケーションファイル) を分解して、

AI モードを有効にしています。

 **すずき注:** インターフェイス上で有効になっただけで、実際には機能しなかったとのこと

検索ボックスの下にあるアイコン群のいちばん左、星マークが付いた虫めがねアイコンを選択すると

AI モードに移行すると思われる。



## Google、会話型検索を提供するAIモードを開発中か

Googleは、より会話的でインタラクティブな検索体験を提供することを目指し、新たな検索機能として AI Mode (AI モード)を開発中らしい。

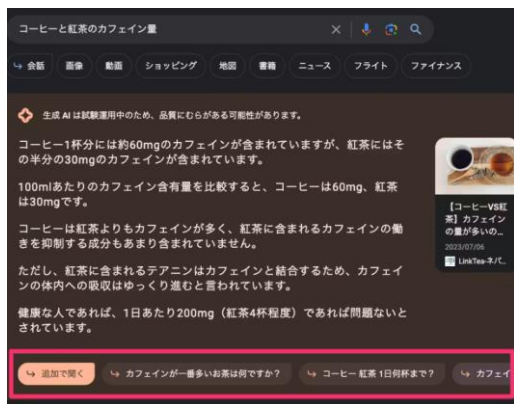
### 会話型検索とフォローアップ質問

AI モードでは、ユーザーが検索クエリを入力すると、回答およびそれに関連するウェブページとその URL を提供します。

このモードの重要な特徴は、フォローアップの質問をユーザーができるように設計されていることです。

トピックをより深く、より詳細に探求できます。

AI Overview の前身である SGE ではフォローアップ質問できましたが、AI Overview ではこの機能はなくなりました。



また、AI モードは、テキストだけではなく音声によるプロンプトをサポートし、ユーザーにハンズフリーのエクスペリエンスを提供します。

SGE では、追加質問できた

### Gemini との統合と幅広いアクセス

AI モードの開発は、Google が Gemini をより広範囲に統合するための戦略の一環であると考えられます。

アクティブユーザー数において OpenAI の ChatGPT との競争に Google は苦戦していると報じられています。

検索プラットフォームを通じて Gemini を導入することで、何十億人ものユーザーが AI チャットボットにアクセスできるようにすることを Google は目指しています。

また、この新しい AI モードが検索エンジンのパフォーマンスも向上させるとも Google は考えているそうです。

## Google、会話型検索を提供するAIモードを開発中か

Googleは、より会話的でインタラクティブな検索体験を提供することを目指し、新たな検索機能として AI Mode (AI モード)を開発中らしい。

### 社内テスト中

現在、Google 検索の AI モードはベータ版で、社内でテスト中です。

そのため、正式な公開時期や AI モードの実際の動作については不明です。

テストだけで終了し、一般公開しない可能性もあります。

OpenAI 開発の ChatGPT の登場に始まり、Google が対抗馬としてリリースした Gemini など、AI チャットボットの加熱ぶりも一段落しました。

しかしながら、検索とは異なり、誰もが Gemini を利用しているとは言い難いのが現状です。

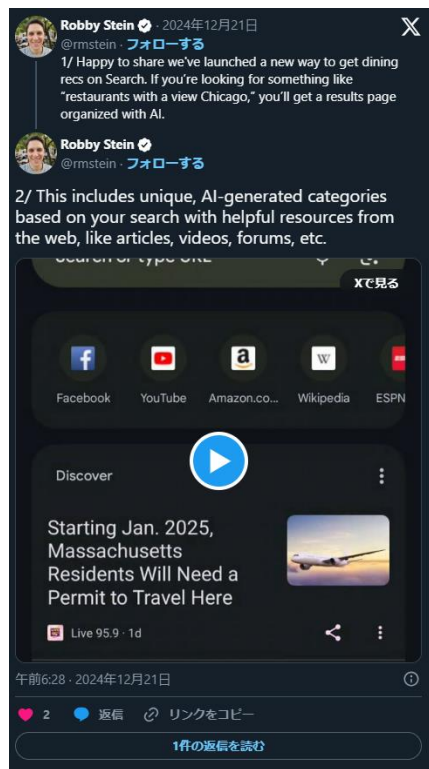
AI Overview は概要を生成する機能であって、インタラクティブに対話できる AI チャットボットではありません。

Google 検索の標準機能とし AI モードが搭載されれば、Gemini が一般ユーザーに浸透する進展はおおいにありえそうです。

## レストランをAIで分類する検索結果が米Googleモバイル検索で導入される

レストランを探すクエリで、AI がカテゴリに応じて分類する検索結果を米 Google のモバイル検索で利用できるようになった。

レストランを探すクエリで、AI がカテゴリに応じて分類する検索結果を米 Google のモバイル検索で利用できるようになりました。



AI が分類する検索結果は、[2024年10月にレシピ関連のクエリで提供](#)されていました。  
レストラン系クエリにも対象を拡大しました



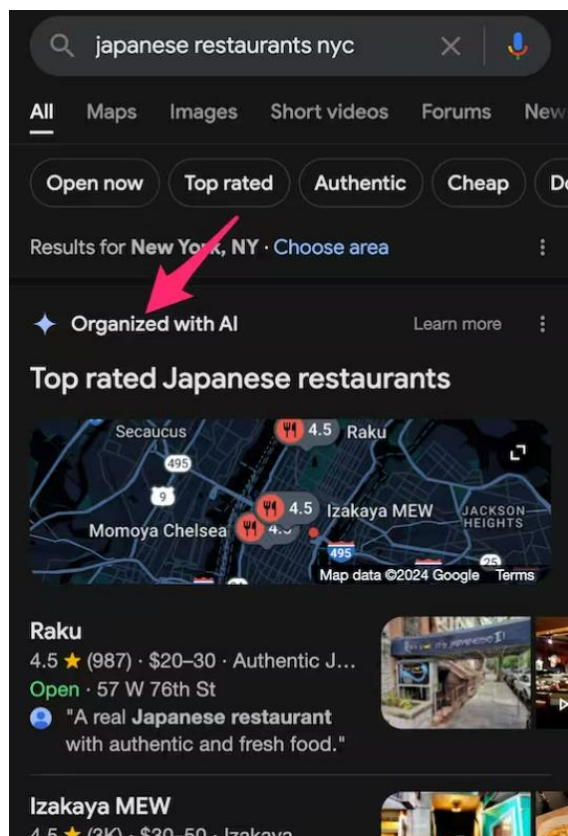
## レストランをAIで分類する検索結果が米Googleモバイル検索で導入される

レストランを探すクエリで、AI がカテゴリに応じて分類する検索結果を米 Google のモバイル検索で利用できるようになった。

### ランキング勝負ではなくなる？

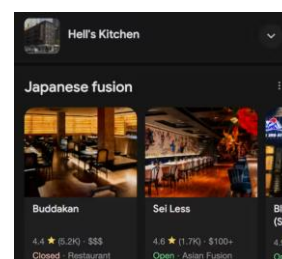
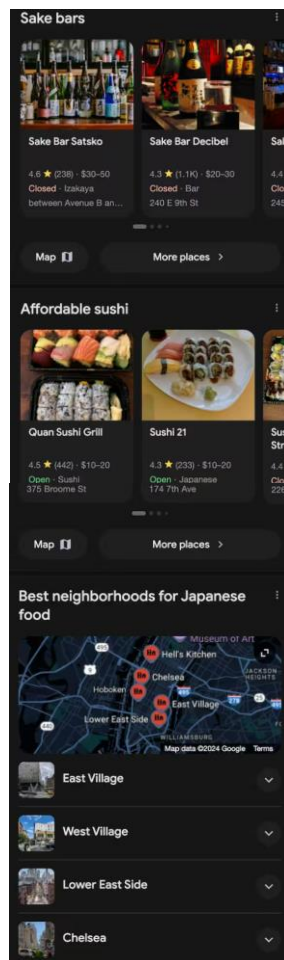
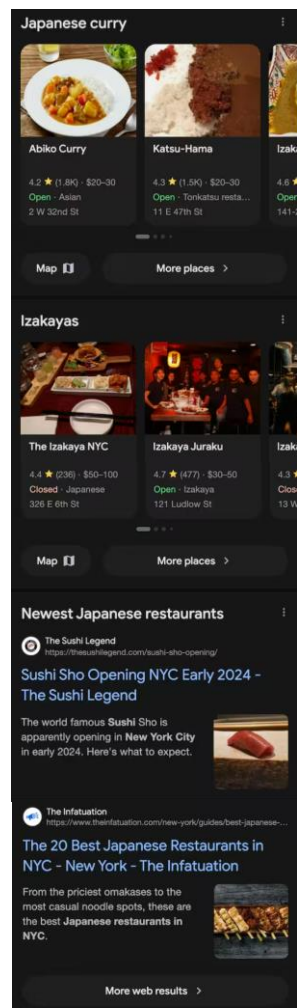
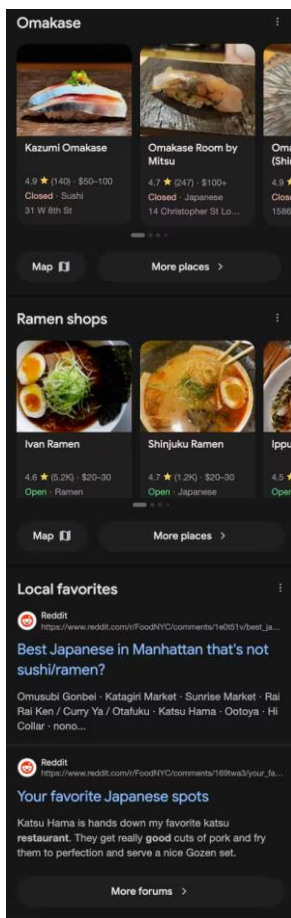
こちらは [japanese restaurants nyc](ニューヨークの日本食レストラン)のモバイル検索結果です。

「Organized with AI」(AIによって分類)というラベルが表示されています。



## レストランをAIで分類する検索結果が米Googleモバイル検索で導入される

レストランを探すクエリで、AI がカテゴリに応じて分類する検索結果を米 Google のモバイル検索で利用できるようになった。



下にスクロールすると、次のようなカテゴリで分類された日本食レストランが検索結果として提供されます。

- Omakase(おまかせ)
- Ramen shops(ラーメン店)
- Local favorites(地元の人のお気に入り)
- Japanese curry(日本風カレー)
- Izakayas(居酒屋)
- Newest Japanese restaurants(新規開店 日本食レストラン)
- Sake bars(日本酒バー)
- Affordable sushi(お手頃な寿司)
- Japanese fusion(日本食フュージョン)
- Hand rolls(手巻き寿司)
- Trending Japanese food(人気上昇中)

👉 すぎ補足: すべての分類ではなく一部

## レストランをAIで分類する検索結果が米Googleモバイル検索で導入される

レストランを探すクエリで、AI がカテゴリに応じて分類する検索結果を米 Google のモバイル検索で利用できるようになった。

ニューヨークの日本食レストランを単純に関連度順に並べた伝統的な検索結果ではなくなっています。

AI が分類したジャンルのなかで選ばれれば、たとえ今までは上位表示していなかったとしても検索結果 1 ページ目に出現することもおおいにありえるでしょう。

AI による分類は、現在は、米 Google のモバイル検索でレシピ検索とレストラン検索だけを対象に導入されています。

レシピの場合は、レシピの種類だけではなく動画や画像、フォーラム投稿といったコンテンツのタイプでも分類されていました。

今後改良を重ねて、より多くのクエリにもAI による分類が適用されるのではないのでしょうか。

日本での導入は不明ですが、特定のクエリにおいて関連度という判断基準だけで検索結果 1 位を目指さなくてもよくなれば、SEO に取り組みやすくなる気がします。

### 2024 年のごあいさつ

今年のブログ更新は今日の投稿が最終です(明日、もう 1 本投稿するけれど、[YouTube ミエルカチャンネル](#)の新着動画紹介なので実質的には今日でおしまい)。

2024 年中も訪問していただきありがとうございました。

2025 年は 1 月 6 日から再開する予定です。

来年も記事を読みに来てもらえると嬉しいです。

では、良い年末年始を！

そして、来年も SEO をがんばりましょう！

## Google社員が明かす、GoogleトレンドのSEO活用術

Google 検索リレーションズチームのダニエル・ワイズバーグ氏に、Google トレンドの SEO 活用方法についてインタビューした。

Google 検索リレーションズチームの [Daniel Waisberg](#) (ダニエル・ワイズバーグ) 氏に、Google トレンドの SEO 活用方法についてインタビューしました。

ワイズバーグ氏が YouTube で公開した [Google トレンド チュートリアル シリーズ](#) をまず視聴してください(このチュートリアル動画は超おすすめ！)

📖 すぎ注: 説明は英語だけど、日本語字幕を利用可。プレイリストは順番が逆なので、下から順番に視聴する

その後に、インタビュー動画を観ると Google トレンドをよりいっそう SEO で利用したくなります。  
ご覧ください。



## CDNについてGoogleが解説

Googleは、米国現地時間12月24日、[Search Central Blogを更新](#)し、CDN(Contents Delivery Network:コンテンツ配信ネットワーク)の利便性と、懸念点、さらにクローラーの特性について解説してくれています。CDNとは、ユーザーに地理的に近いところでHTMLや画像、CSS、JavaScriptなどのリソースファイルをキャッシュするサーバー群の仕組みです。ユーザーに迅速にコンテンツを配信することを目的としています。

### Googleからの記事内容

繰り返しになりますが、CDNとは、HTMLや画像、CSS、JavaScriptなどのリソースファイルのキャッシュを保存しておくようなサーバー群のイメージです。サイト本体のサーバーとなるオリジンサーバーがCDNに対してキャッシュデータを提供します。そうすると、ユーザーはCDNからキャッシュデータを取得して閲覧することになります。それによって、ユーザーがアクセスする度、オリジンサーバーからデータを読み込まなくて済むので、表示速度が改善できたり、Botが大量にきてオリジンサーバーへの負荷を軽減したり、逐一データダウンロードしなくて済むのでクロールも速く行えるような仕組みになっています。今回Googleはこういったことを説明しているわけです。

また、レンダリングについては、例えばexample.comというサイトがあった場合に、CDNサーバーを

- cdn-example.comのような独自のホスト名のドメインにする
- cdn.example.comのようなサブドメインにする

とした場合、ページの読み込みトラフィックを、example.comとそれ以外で分割できるため、効率的にレンダリングができるようになるという内容です。ただ、別ドメインを読み込むことにもなり、そこが逆に悪影響が出ることもあるので注意しましょう、とも書かれています。CDNサーバーを(リバースプロキシ的なイメージで)example.comに入り込んだ形で提供すると別ドメインを読み込む必要がなくなるのでその部分も解決可能できるとのことで、そこは運用中のホスティングシステムをよく考慮しながらベストプラクティスを選択してください、ということもGoogleは推奨しています。

ちなみに配信にあたって、物理的な距離に関しては、やはり光速とはいえ、遠いより近い方が良いです。ですので、グローバルに展開する場合は、例えば国や地域ごとにCDNを用意して、ユーザーから近いCDNから読み込むように設定することで各地域で表示速度を向上できたりもします(その分コストもかかるので余程大規模でない限りはあまり気にしなくていいのかなとも思います)。

というわけで[Googleからの発表記事](#)を和訳します。

## CDNについてGoogleが解説

Googleは、米国現地時間12月24日、[Search Central Blogを更新](#)し、CDN(Contents Delivery Network:コンテンツ配信ネットワーク)の利便性と、懸念点、さらにクローラーの特性について解説してくれています。CDNとは、ユーザーに地理的に近いところでHTMLや画像、CSS、JavaScriptなどのリソースファイルをキャッシュするサーバー群の仕組みです。ユーザーに迅速にコンテンツを配信することを目的としています。

### クローラーの12月:CDNとクローリング

コンテンツ配信ネットワーク(CDN)は、Webサイトの反応時間を短縮し、Webトラフィック関連の悩みを解消するうえで特に有効な方法です。結局、CDNの主な目的はこれです。つまり、サイトに大量のトラフィックが流入しても、コンテンツを迅速に配信することです。CDNの“D”は、世界中にコンテンツを配信または配布することを意味します。そのため、ユーザーへの転送時間も、どこかの1つのデータセンターでホストするより短時間で済みます。そこで、この記事では、CDNを活用してサイトのクローラーとユーザーエクスペリエンスを向上させる方法を模索し、CDNでバックアップされたサイトのクローラーに関する意図についてもご説明します。

### 要約:CDNとは何か

CDNは基本的に、オリジンサーバー(Webサイトが存在する場所)とエンドユーザーとの仲介役の存在であり、オリジンサーバーに代わって(一部の)ファイルをエンドユーザーに提供します。これまでの長い目で見て、**CDNの最大の特徴はキャッシュ**です。つまり、ユーザーがサイトに対してURLを要求すると、CDNはそのURLの内容を一定期間キャッシュに保存するため、オリジンサーバーはしばらくの間そのファイルを再度提供する必要がなくなります。

CDNは、ユーザーに近い場所からサービスを提供することで、**サイトの表示速度を大幅に向上させる**ことができます。たとえば、オーストラリアのユーザーがドイツでホストされているサイトにアクセスしている場合、CDNはオーストラリアのキャッシュからそのユーザーにサービスを提供して、地球をまたぐほどの往復時間を短縮することができます。光速であろうとなかろうと、**距離というのは依然としてかなり大きい**です。

そして最後に、**CDNはサイトを過負荷やセキュリティの脅威から守る素晴らしいツール**と言えます。CDNが管理するグローバルトラフィックの量に応じて、信頼性の高いトラフィックモデルを構築し、トラフィックの異常を検出し、過剰または悪意のあると思われるアクセスをブロックできます。たとえば、2024年10月21日、[Cloudflareのシステム](#)は、約1分間続いた**4.2Tbps**(編集者注:これはかなり大きい)のDDoS攻撃を自律的に検出し、軽減しました。

## CDNについてGoogleが解説

Googleは、米国現地時間12月24日、[Search Central Blogを更新](#)し、CDN(Contents Delivery Network:コンテンツ配信ネットワーク)の利便性と、懸念点、さらにクローラーの特性について解説してくれています。CDNとは、ユーザーに地理的に近いところでHTMLや画像、CSS、JavaScriptなどのリソースファイルをキャッシュするサーバー群の仕組みです。ユーザーに迅速にコンテンツを配信することを目的としています。

### CDNがサイトにどう役立つか

最速のサーバーと最高のアップリンクを所有しているから速度を上げる必要はない、と考える人もいるかもしれませんが、CDNは長期的にはコストを節約できます。特にサイトが大きい場合はなおのこと発揮します：

- **CDNでのキャッシュ:**メディア、JavaScript、CSS、さらにはHTMLなどのリソースがCDNのキャッシュから提供される場合、サーバーはそれらのリソースの提供にコンピューティングと帯域幅を費やす必要がなくなり、その過程でサーバーの負荷が軽減されます。これは通常、ユーザーのブラウザでのページの読み込みが速くなることも意味し、[コンバージョンの向上にも繋がるでしょう](#)。
- **トラフィックフラッド保護:**CDNは、過剰なトラフィックや悪質なトラフィックを識別してブロックするのに特に優れているため、不正なボットや悪質なユーザーがサーバーに過負荷をかけている場合でも、ユーザーがサイトにアクセスできるようにします。

フラッド保護に加えて、不正なトラフィックをブロックするために使用される同じ制御を使用して、特定のクローラー、特定のパターンに当てはまるクライアント、または同じIPアドレスを使用し続けるトロールなど、単に不要なトラフィックをブロックすることもできます。サーバーやファイアウォールでもこういったことを行うことができますが、通常はCDNのユーザーインターフェイスを使用する方がはるかに簡単です。

- **信頼性:**一部のCDNは、サイトがダウンしている場合でも、ユーザーにサイトを提供できます。もちろん、これは静的コンテンツにしか機能しないかもしれませんが、ユーザーが他のサイトに移らないようにするには、それだけでも十分かもしれません。

つまり、CDNはWebサイトの味方であり、サイトが大きい場合や大量のトラフィックが予想される場合(または既に大量のトラフィックを受信している場合)は、価格、パフォーマンス、信頼性、セキュリティ、カスタマーサポート、スケーラビリティ、将来の拡張などの要素に基づいて、ニーズに合ったCDNを使用することをオススメします。ホスティングプロバイダまたはCMSプロバイダに問い合わせ、オプション(または既に使用しているかどうか)をご確認ください。

### クローラーがCDNを使用するサイトに与える影響

クローリングの面では、CDNも役立ちますが、クローリングの問題を引き起こす可能性があります(稀です)。下記ご覧ください。

## CDNについてGoogleが解説

Googleは、米国現地時間12月24日、[Search Central Blogを更新](#)し、CDN(Contents Delivery Network:コンテンツ配信ネットワーク)の利便性と、懸念点、さらにクローラーの特性について解説してくれています。CDNとは、ユーザーに地理的に近いところでHTMLや画像、CSS、JavaScriptなどのリソースファイルをキャッシュするサーバー群の仕組みです。ユーザーに迅速にコンテンツを配信することを目的としています。

### CDNのクロール率への影響

Googleのクロールインフラストラクチャは、CDNによってサポートされているサイトでより高いクロール率を可能にするように設計されています。これは、クローラーがアクセスするURLを提供しているサービスのIPアドレスから推測されます。これによって、ほとんどの場合はうまく機能します。

例えば、今日ストックフォトサイトを立ち上げたとします。ストック写真が1,000,007枚あります。ランディングページ、カテゴリページ、すべてのアイテムの詳細ページを備えたWebサイトを立ち上げると、当然ページ数が多くなります。[クロール容量制限に関するドキュメント](#)では、Google検索はこれらすべてのページをできるだけ早くクロールしたいと考えますが、クロールによってサーバーが過負荷にならないようにする必要があることを説明しています。クロール リクエストの数が増えてサーバーの応答が遅くなると、サーバーの過負荷を防ぐためにGoogle側でスロットリングが適用されます。このスロットリングに関して、サイトがCDNによってバックアップされていることがクローリングインフラストラクチャによって検出されることで、サーバーが処理できる可能性が高いと判断され同時にリクエストを送信しても問題ないと想定された場合、そのしきい値ははるかに高くなります。これにより、Webサイトへのクロールが高速化されます。

ただし、URLへの最初のアクセスでは、CDNのキャッシュは“コールド”です。つまり、まだ誰もそのURLをリクエストしていないため、そのコンテンツはCDNによってまだキャッシュされていません。そのため、オリジンサーバーは、CDNのキャッシュを“ウォームアップ”するために、そのURLを少なくとも1回は提供する必要があります。これは、[HTTPキャッシュの動作](#)と非常によく似ていると言えます。

つまり、WebサイトがCDNによってサポートされている場合でも、サーバーは少なくとも1回は1,000,007個のURLを処理する必要があります。最初の処理が終わって初めて、CDNはキャッシュで対応できるようになります。これは“クロール バジェット”に大きな負担となり、クロール率は数日間高くなる可能性があります。一度に多数の URLを起動する予定がある場合は、この点にご注意ください。

### CDNのレンダリングへの影響

クロールに関する[12月の最初のクロールブログ投稿](#)で説明したように、リソースを独自のホスト名またはCDNホスト名(cdn.example.com)に分割すると、Webレンダリングサービス(WRS)でページをより効率的にレンダリングできるようになります。ただし、注意点があります。この方法では、別のホスト名への接続の負荷過多によりページのパフォーマンスに悪影響が出る可能性があるため、レンダリングパフォーマンスと[ページエクスペリエンス](#)を慎重に検討する必要があります。

これは、メインホストをCDNでバックアップすると、問題を回避できます:クエリするホスト名は1つで、重要なレンダリングリソースはCDNのキャッシュから提供される可能性が高いため、サーバーがそれらを提供する必要がありません(ページエクスペリエンスに影響はありません)。

最終的には、ビジネスに最適なソリューションを選択しましょう:静的リソース用に別のホスト名(cdn.example.com)を用意するか、メインのホスト名をCDNでバックアップするか、またはその両方を実行します。Googleのクロールインフラストラクチャは、どちらのオプションも問題なくサポートします。



## CDNについてGoogleが解説

Googleは、米国現地時間12月24日、[Search Central Blogを更新](#)し、CDN(Contents Delivery Network:コンテンツ配信ネットワーク)の利便性と、懸念点、さらにクローラーの特性について解説してくれています。CDNとは、ユーザーに地理的に近いところでHTMLや画像、CSS、JavaScriptなどのリソースファイルをキャッシュするサーバー群の仕組みです。ユーザーに迅速にコンテンツを配信することを目的としています。

### CDNが過剰保護になってしまうと

CDNのフラッド保護とクローラーのクロール方法により、サイトにアクセスしてほしくないボットがCDNのブロックリストー通常はWebアプリケーションファイアウォール(WAF)、に登録されることがあります。これによりクローラーがサイトにアクセスできなくなり、最終的にはサイトが検索結果に表示されなくなる可能性があります。ブロックはさまざまな方法で発生する可能性があり、Googleの検索結果でのサイトの表示に他の方法よりも悪影響を与えるものもあれば、CDN側で発生するため制御が難しい(または不可能)こともあります。このブログ投稿では、ブロックをハードブロックとソフトブロックの2つのカテゴリに分類しています。

### ハードブロック

ハードブロックとは、CDNがクロール要求に対して何らかのエラーの応答を送信した場合に発生します。これには次のようなものがあります：

- **HTTP503/429ステータスコード**:これらのステータスコードを送信することは、一時的なブロックを通知するための推奨される方法です。これにより、CDNによる意図しないブロックに対処する時間が与えられます。
- **ネットワークタイムアウト**:CDNからのネットワークタイムアウトが発生すると、影響を受けるURLがGoogleの検索インデックスから削除されます。これは、これらの**ネットワークエラーが致命的な“ハード”エラーと見なされるため**です。また、サイトが過負荷になっていることをクローインフラストラクチャに通知するため、サイトのクロール率にも大きな影響を与える可能性があります。
- **200HTTPステータスコード付きのランダムエラーメッセージ:ソフトエラー**とも呼ばれるこのエラーメッセージは、特に悪質です。エラーメッセージがGoogle側で“ハード”エラー(HTTP500等)とみなされた場合、Googleは検索からURLを削除します。Googleがエラーメッセージを“ハード”エラーとして検出できなかった場合、同じエラーメッセージを持つすべてのページがGoogleの検索インデックスから重複として削除される可能性があります。Googleのインデックス作成では重複URLの再クロールを要求する動機がほとんどないため、この状態から回復するには時間がかかる可能性があります。

### ソフトブロック

CDNが“本当に人間ですか?”というインタースティシャルを表示するときにも、同様の問題が発生する可能性があります(これはしゃれのつもりです)。

実際、当社のクローラーは、自分は人間ではないと理解しており人間を装っているわけでもありません。ただクロールしたいだけです。しかし、インタースティシャルが表示されると、クローラーが目にするのはあなたの素晴らしいサイトではなく、インタースティシャルの表示部分だけです。これらのボット検証インタースティシャルの場合、コンテンツが一時的に利用できないことをクローラーなどの自動クライアントに503 HTTPステータスコードの形式で明確に伝えることを強くお勧めします。これにより、コンテンツがGoogleのインデックスから自動的に削除されなくなります。

## CDNについてGoogleが解説

Googleは、米国現地時間12月24日、[Search Central Blogを更新](#)し、CDN(Contents Delivery Network:コンテンツ配信ネットワーク)の利便性と、懸念点、さらにクローラーの特性について解説してくれています。CDNとは、ユーザーに地理的に近いところでHTMLや画像、CSS、JavaScriptなどのリソースファイルをキャッシュするサーバー群の仕組みです。ユーザーに迅速にコンテンツを配信することを目的としています。

### 障害のデバッグ

ハードブロックとソフトブロックの両方において、正しく動作しているかどうかを確認する最も簡単な方法は、[Search ConsoleのURL検査ツール](#)を使用してレンダリングされたイメージを確認することです。ページが表示されれば問題ありません。空のページ、エラー、またはポットチャレンジのあるページが表示される場合は、CDN に相談することをお勧めします。さらに、こうした意図しないブロックに対処するため、Google、その他の検索エンジン、その他のクローラーオペレーターは、[当社のIPアドレス](#)を公開して、クローラーを識別できるようにしています。また、適切と思われる場合は、ブロックされたIPをWAFルールから削除したり、許可リストに追加したりすることもできます。これをどこで実行できるかは、使用しているCDNによって異なります。幸い、ほとんどのCDNと単独型WAFには、すばらしいドキュメントがあります。少し検索すると、次のようなドキュメントが見つかります(この記事の公開時点):

- クラウドフレア:<https://developers.cloudflare.com/bots/get-started/free/#visibility>
- アカマイ:<https://www.akamai.com/products/bot-manager>
- ファストリー:<https://www.fastly.com/products/bot-management>
- F5:[https://clouddocs.f5.com/bigip-next/20-2-0/waf management/waf bot protection.html](https://clouddocs.f5.com/bigip-next/20-2-0/waf%20management/waf_bot_protection.html)
- Googleクラウド:<https://cloud.google.com/armor/docs/bot-management>

サイトを検索エンジンに表示する必要がある場合は、重要なクローラーがサイトにアクセスできるかどうかを確認することを強くお勧めします。知らないうちにIPが自動的にブロックリストに登録される可能性があるため、ブロックリストを定期的に確認することで検索エンジンフレンドリーにしておきましょう。ブロックリストが非常に長い場合(このブログ投稿みたいですが)、IP範囲の最初の数セグメントだけを検索するようにしてください。たとえば、192.168.0.101を検索するなら、192.168だけを検索すれば良いです。この記事は、クローラの12月投稿特集の最後です。私たちが今回のブログを書いたのを楽しんだのと同じくらい、皆様も楽しんでいただければ幸いです。ムニャムニャ…という感じであれば、もうやり方はご存じでしょう。

### 引用)Search Central Blogより和訳

いかがでしたでしょうか。

かなりマニアックな内容ですので、エンジニア向けの内容となっておりますが、CDNに関してはSEOにおいても推奨したり議論したりするケースは大いにありますので、仕組みや意味は理解しておくようにしましょう。

ではでは。

## CDNとクロール:コンテンツデリバリーネットワークがGoogle検索に与える影響

CDN を利用する際の Google 検索における利点と注意点を検索セントラルブログが解説した。

CDN を利用する際の Google 検索における利点と注意点を[検索セントラルブログが解説](#)しました。

クロールをトピックにして昨年 12 月に毎週連載した [Crawling December](#) シリーズの 4 話目(最終)のエピソードです。過去 3 回のエピソードは次のとおりです。

1. [Googlebotはウェブページをどのようにクロールしレンダリングするのか?](#)
2. [HTTP キャッシングの利点と Google がサポートする HTTP キャッシング](#)
3. [ファセットナビゲーションのベストプラクティス](#)

今回の CND の記事はすでに[日本語訳](#)が発行されています。

ここでは要点を紹介します。

### CDN とは?

CDN (Content Delivery Network, コンテンツ デリバリー ネットワーク)は、ウェブサイトのオリジンサーバーとエンドユーザーの間の仲介役として機能します。

CDN の主要な目的は次の 2 つです。

- コンテンツ配信の遅延を削減する
- ウェブサイトのトラフィックを効果的に管理する

CDN はコンテンツをキャッシュし一時的に保存することで、オリジンサーバーが同じコンテンツを繰り返し提供することを回避します。

このキャッシュメカニズムにより、ユーザーの地理的に近いサーバーロケーションからコンテンツを配信でき、ウェブサイトの速度を向上させます。

たとえば、ウェブサイトがドイツでホストされている場合、オーストラリアのユーザーにオーストラリアのキャッシュから CDN はサービスを提供します。

こうすると、コンテンツ(キャッシュ)を配信するサーバーとユーザーが物理的に近い分だけコンテンツ配信が速くなります。

また、CDN はトラフィックモデルを使用して悪意のあるアクセス試行を検出してブロックし、トラフィックの過負荷や特定のセキュリティ脅威に対する保護を提供します。

4.2Tbps の DDoS 攻撃を Cloudflare(大手 CDN)がわずか 1 分ほどで自律的に軽減したという事例もあります。

## CDNとクローリング:コンテンツデリバリーネットワークがGoogle検索に与える影響

CDN を利用する際の Google 検索における利点と注意点を検索セントラルブログが解説した。

### CDN がウェブサイトに関与する仕組み

次のような利点が CDN にはあります。

- **キャッシング:** メディア、JavaScript、CSS、HTMLなどのリソースをキャッシュすることにより、サーバーの負荷と帯域幅の使用量を削減する。このキャッシングは、ページの読み込み時間を短縮し、コンバージョンを向上させる可能性もある
- **大量のトラフィックからの保護:** CDN は、悪意のあるトラフィックや過剰なトラフィックを識別してブロックすることに優れており、攻撃されてもサイトの可用性を確保する。また、特定のパターンや IP アドレスに基づいて、不要なクローラーやクライアントをブロックすることもできる
- **信頼性:** 一部の CDN は、オリジンサーバーがダウンしている場合でも静的コンテンツを提供できるため、ビジネスの継続性を確保できる

特に、トラフィックが多いサイトに CDN は役立ちます。

CDN を選択する際には、価格やパフォーマンス、信頼性、セキュリティ、サポート、スケーラビリティ、将来の拡張性などの要素を考慮する必要があります。

### CDN がクローリングに与える影響

CDN はクローリングに関与しますが、いくつかの問題を引き起こす可能性もあります。

たとえば、Google のクローリングインフラストラクチャは IP アドレスによって CDN を利用しているサイトを識別し、より多くのリクエストをサーバーが処理できると想定するため、より高いクローリング頻度を CDN は可能にします。

そのため、Google のクローラーは、CDN を利用するサイトに、より多くの同時リクエストを送信します。

ところが、CDN のキャッシングは、URL に初めてアクセスしたときは「コールド」(キャッシングがまだない状態)です。

CDN のキャッシングを「ウォームアップ」(キャッシングが保存された状態)するには、少なくとも 1 回はオリジンサーバーがコンテンツを提供する必要があります。

この初段階のサーバー負荷は、多くの新しい URL を配信するときに大きくなる可能性があります。

別の CDN ホスト名にリソースを分割すると、レンダリング効率が向上する可能性があります。

しかし、追加の接続オーバーヘッドにより、ページのパフォーマンスに悪影響を及ぼす可能性もあります。

メインホスト名に CDN を使用すると、重要なレンダリングリソースが CDN のキャッシュから提供される可能性が高いため、この問題を回避できます。

静的リソースに別のホスト名を使用すること、CDN でメインホスト名を支援すること、またはその両方を構成することを Google のクローリング インフラストラクチャはサポートしていません。

## CDNとクローリング:コンテンツデリバリーネットワークがGoogle検索に与える影響

CDN を利用する際の Google 検索における利点と注意点を検索セントラルブログが解説した。

### CDN が過剰に保護するケース

CDN が提供する大量のトラフィックからの保護は、正当なクローラーまでブロックすることがあります。

これは多くの場合、ウェブアプリケーションファイアウォール (WAF) が原因です。

サイトの検索結果での表示に影響を与える可能性があります。

こうしたブロックは、「ハード」ブロックまたは「ソフト」ブロックとして分類できます。

### ハードブロック

ハードブロックには、次のものが含まれます。

- **HTTP 503/429 ステータスコード:** これらは一時的なブロックを示し、意図しないブロックに Googlebot が反応する時間を与える
- **ネットワーク タイムアウト:** これらは「ハード」エラーと見なされ、影響を受ける URL が Google のインデックスから削除され、クローリング頻度も低下する可能性がある
- **HTTP 200 ステータスコードのランダムなエラーメッセージ:** これらの「ソフトエラー」は、ハードエラーとして解釈された場合、URLが削除されたり重複として扱われたりする可能性がある

### ソフトブロック

ソフトブロックは、多くの場合、ボットでないことを検証するインターstitialが原因です。

Google のクローラーは人間として識別されないため、ブロックされる可能性があります。

Googleのインデックスからコンテンツが自動的に削除されないように、自動クライアントには 503 ステータスコードを送信することが推奨されます。

ブロックをデバッグするには、Search Console の URL 検査ツールを使用して、レンダリングされスクリーンショットを確認します。

空のページ、エラー、またはボットチャレンジが表示される場合、CDN がクローラーをブロックしている可能性があります。

Google やその他のクローラーは、CDN 設定で許可リストに登録できるクローラーを識別するために、IP アドレスを公開しています。

要点を絞って解説するつもりでしたが、意外に長くなってしまいました。

それでも端折ったところがあります。

CDN を利用しているサイト、もしくは利用を検討しているサイトの管理者は元記事も読んでください。

- [Crawling December: CDNs and crawling](#) (英語)
- [CDN でのキャッシュ保存](#) (日本語)