

[ 2024年最新アルゴリズム対応 ]  
サイト内部修正指示書

Google APIコンテンツ倉庫参照  
×  
独自のアルゴリズム研究

# 最新Googleアルゴリズム対策

©2024年5月末にGoogleの社内文書が漏洩しました。Google CloudのDocument AI Warehouseというサービスから「Content Warehouse」APIというGoogleの内部文書が流出しました。

## Google APIコンテンツ倉庫の情報参考元

<https://hexdocs.pm/google-api-content-warehouse/0.4.0/api-reference.html>

<https://ipullrank.com/google-algo-leak>

<https://paulgraham.com/articles.html>

<https://searchengineland.com/how-seo-moves-forward-google-leak-442749>

<https://searchengineland.com/google-search-document-leak-ranking-442617>

<https://sparktoro.com/blog/an-anonymous-source-shared-thousands-of-leaked-google-search-api-documents-with-me-everyone-in-seo-should-see-them/>

<https://blog.marketmuse.com/google-content-warehouse-api-leak/>

<https://cloudonair.withgoogle.com/events/next-tokyo/watch?talk=d1-db-01>

<https://www.searchenginejournal.com/extending-your-schema-markup-from-rich-results-to-a-knowledge-graph/511958/>

<https://www.justice.gov/atr/us-and-plaintiff-states-v-google-llc-2020-trial-exhibits>

<https://www.blindfiveyearold.com/its-goog-enough>

<https://events.goldcast.io/splash/052be756-fba8-4bfc-aa9d-c6eafe63262e>

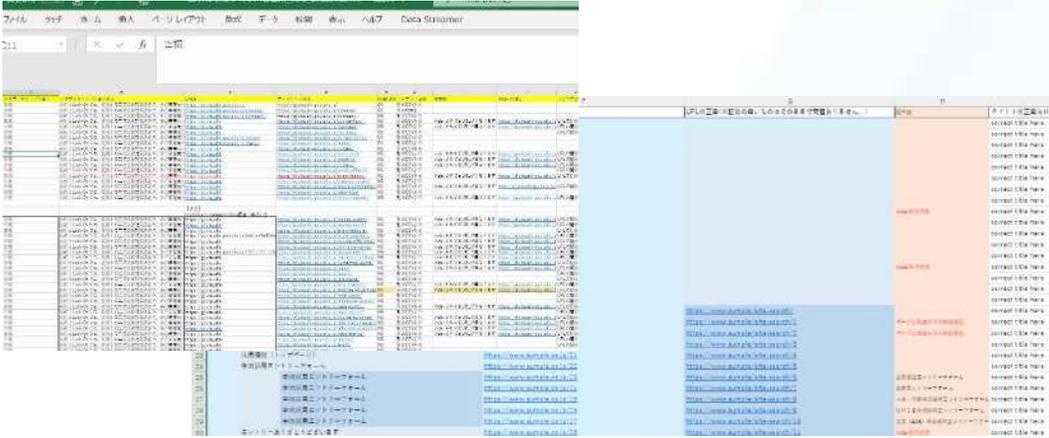
<https://www.semrush.com/blog/semantic-search/>

## Google APIコンテンツ倉庫参照修正指示書

**GOOGLE** リードドキュメントには **14,000** のランキング機能などが記載されています

APIドキュメントには、次のような **14,014** 個の属性(機能)を持つ **2,596** 個のモジュールが表されています。

author属性 (メタデータ)	リンク先コンテンツとのミスマッチ修正
Page speed insightによる修正指示	ショートコンテンツの独創性維持
日付更新	UXの不備
文字フォント	ヘルプフルコンテンツアップデート対応



Google APIコンテンツ倉庫を参考に修正指示書を作成します。但し、弊社の研究においてマイナス評価につながりそうなものは外します。

プラス評価に該当するものだけサイト修正指示書に明記します。

**495,000円** (税込)

- ①Navboost・・・ Navboost(ナブブースト)は、Googleの検索ランキングシステムの一部で、ユーザーのクリックデータを基にページのランキングを決定するシグナルの一つです。SEOにUXが重要なのがわかりました。Googleが検索結果からのクリックデータの13ヶ月分を分析してランキングを調整するシステムであり、最も重要なランキングシグナルの一つ。GoogleがChromeのユーザー行動データを使っていることもわかりました。
- ②セマンティック検索・・・ユーザーの検索意図を理解した上で検索結果をかえす。昔は語彙の完全一致検索を返している程度だった。(エンベディング・ベクトル検索が支えている)
- ③リンクの3つの質・・・Googleは3つの評価にリンクを分けている。評価が高いのはTOPページなどの重要なページからのリンク、普通のページは、中評価、更新頻度の低いページは低評価となっている。流入数が多いページからのリンクがよい。
- ④日付の更新・・・Googleは日付を見て更新頻度が高いページを評価しています。実際に、ページを更新するとランキングがあがるパターンが多いことが弊社でも確認が取れています。
- ⑤Trawler・・・WEBのクロールシステム、クロールキュー(クロールすべきURLリスト)を備え、クロール レート(ウェブサイトに対して行うリクエストの頻度)を維持し、ページが変更される頻度を把握します。
- ⑥インデックス・・・ コアインデックスシステムは、Alexandriaとよばれ、大規模データ処理と高い可用性、スケーラビリティを兼ねそろえています。SegIndexerは、階層化されたドキュメントをインデックス内の階層に配置するシステムです。TeraGoogleは、ディスク上に長期保存されるドキュメントのセカンダリインデックスシステムです。このように主要な3つのシステムが稼働しています。

## ⑦レンダリング

HtmlrenderWebkitHeadless - JavaScript ページのレンダリング システムです。これは Chromium ではなく Webkit にちなんで名付けられています。ドキュメントには Chromium についての言及があるので、Google は当初 WebKit を使用していて、Headless Chrome の登場後に切り替えたと考えられます。

## ⑧処理

LinkExtractor - ページからリンクを抽出します。

WebMirror - 正規化と複製を管理するためのシステム。

## ⑨ランキング

ムスタング - 主要なスコアリング、ランキング、サービングシステム

Ascorer - 再ランキング調整の前にページをランク付けする主要なランキングアルゴリズム。

NavBoost - ユーザー行動のクリックログに基づいた再ランキングシステム。

FreshnessTwiddler - 新鮮さに基づいてドキュメントを再ランク付けするシステム。

WebChooserScorer - スニペットのスコアリングで使用される機能名を定義します。

## ⑩サービング

Google Web Server - GWS は、Google のフロントエンドがやり取りするサーバーです。ユーザーに表示するデータのペイロードを受信します。

SuperRoot - これは Google 検索の頭脳であり、Google のサーバーにメッセージを送信し、結果の再ランク付けと表示のための後処理システムを管理します。

SnippetBrain - 結果のスニペットを生成するシステム。

Glue - ユーザーの行動を利用して普遍的な結果をまとめるシステム。

クックブック - シグナルを生成するシステム。実行時に値が作成されるという兆候があります。

- ①Googleのトップ10に位置するサイトの96%が、ユニークドメインから1,000以上のリンクを持っている。
- ②トップ10にランクインしているウェブサイトのうち、100未満のバックリンクを持つサイトは0.3%であり、50リンク以下のサイトはありませんでした。

→

更新された研究: リンクはGoogleのランキングにとって依然として非常に重要  
新しいGoogleリンク研究は、リンクが検索でのランク付けにとって依然として重要であることを示しています。新しいGoogleランキング研究は、リンクがランキングアルゴリズムにとって非常に重要であることを示しています。

③AIで自動翻訳した記事であっても、高品質で有益であればGoogle検索で問題視されることはありません。

④グーグルにコンテンツを正しく認識させる「エンティティ」

⑤ドメイン権威

Googleの公式見解は「我々はドメイン権威のようなものは持っていない」というものだったが、siteAuthorityというfeatureの存在が確認された。

⑥クリック

これもGoogleの公式見解と異なり、GlueやNavBoostというランキングシステムにより、クリック情報が検索ランクに影響を与えていることが明らかになった。NavBoostは2005年から存在していて、過去18ヶ月のクリックデータを元としている。

⑦ホスト年齢

hostAgeというフィーチャーの存在によって、生まれたばかりの新しいサイトはやはりサンドボックスに閉じ込められていることが明らかに。hostAgeという言葉自体がhost + ageの合成語だけでなく、一単語で「人質・担保」という意味を持つので言葉遊びです。

## ⑧ Chromeのデータ

Chromeの利用データはサーチに影響しないという公式見解をよそに、やはりChromeのデータは使われていた。使わないわけがないので否定すること自体が公式見解の信憑性を損ねていたと思います。

## ⑨アーキテクチャ

Googleのランキングシステムは単体の巨大なアルゴリズムではなくマイクロサービスの集合体。Trawler (クロール) / Alexandria (インデックス) / Mustang (ランキング) / SuperRoot (クエリー処理) などからなる。

## ⑩リランカー

これに加えて、NavBoost / QualityBoost / RealTimeBoostなどのブースターでSERPの一等地に何を表示させるかを競い合わせている。中でもNavBoostは出現率が高く強力な様子。

## ⑪Pandaアルゴリズム

Amit SinghalがリードしたPandaは、ユーザー行動や外部リンクなどに基づいてランクのスコアを調整する。特許にもなっている。ドメイン、サブドメイン、パスなどの多レベルで適用される。オブザーバビリティを担保するため、あえて機械学習を使ってないという点が印象的。

## ⑫ author属性

Google公式のE-E-A-T推奨通り、author属性は特徴量になっていた。ほとんどのサイトやフィードでauthor系メタデータはちゃんと入力されていないケースが多いが、これは対応したほうが良いかもしれない。

## ⑬降格

リンク先のコンテンツとのミスマッチ、SERPにおけるユーザー行動、UXの不備、検索ワードとドメイン名が完全一致する場合、製品レビュー系コンテンツ、グローバルページ、ポルノ、などは降格される。

## ⑭リンクグラフはまだ重要

sourceTypeなどのメトリックによりリンクの重みが変わる。インデックス・ティアに分けられ、重要度高・更新頻度高・アクセス頻度高のコンテンツはより高速なフラッシュメモリに保存され、ティアの低いインデックスはハードディスクに保存される。

## ⑮リンクスパムはベロシティ・シグナル重視

スパムは短期間に大量投下されたもののスパイクを検出することを特に重視している。怪しいものを見つけたら現在のベースラインと比較して異常値を検出。

## ⑩ 最新20件のページ更新履歴を保持

インデックスされたページの評価が高まってから変更して別のコンテンツに誘導するなどのハックを防ぐために過去20件の変更履歴を保持していることが判明。逆にいうと最大20件とバれてしまったので狙われそう…

## ⑪ ショートコンテンツの独創性を評価

Twitterのようなショートコンテンツでは、OriginalContentScoreにより、文字数をカウントし、オリジナリティを評価。

## ⑫ 日付は重要

鮮度は重要。bylineDate / syntacticDate / semanticDateなどにより、コンテンツの正確な日時情報を様々な角度から抽出している。

## ⑬ ビデオサイトは別扱い

もし50%以上のページがビデオならビデオサイトとして扱う

## ⑭ Your Money Your Life (YMYL) も特別扱い

公式情報通り、お金や健康など人生にクリティカルな影響を与える情報はより厳格に扱われる。

## ⑮ ゴールドスタンダードはある

goldenというフィーチャーで人間あるいはAIにラベルされたコンテンツは実在する。https://paulgraham.com/articles.html とか

## ⑯ Embeddingによるトピックチェック

ページのコンテンツがサイトの趣旨に沿ったものであるかどうかをembeddingでチェックしている。これで、サイトにあると違和感を感じるような内容かどうかをチェック。

## ⑰ 零細個人サイトは特別扱い？

smallPersonalSiteというフィーチャーで、資金力にもものを言わせることができない個人サイトは公平のためブーストしている様子。というわけで、これまで謎だったGoogle検索エンジンの様々な側面が今回初めて公にリークされたので、可能な範囲で解説してみました。興味がある方は、以下のリンクの原文を当たってみてください。

[ その他お役立ち情報 ]

【Google】日本独自のハッシュタグ検索が始まりました。

公式情報

<https://twitter.com/googlejapan/status/1803352045094683089>

<https://blog.google/intl/ja-jp/products/explore-get-answers/search-japan-nowcast-hashtag/>

東京SEOメーカーは確かな戦略コンサルで「あらゆる課題」を解決するデジタルマーケティングのプロ集団、それが東京SEOメーカーです。  
検索エンジン黎明期から様々な観点でSEOを中心にWebを研究し続けていた代表が2012年に創業。長年にわたるアルゴリズム研究と検証で独自の上位表示ノウハウを蓄積しつつ、無駄を省いた社内システムを構築することでクオリティの高いサービスを実現しました。

現在ではSEOのみならず、技術と知見を常に最先端へとアップデートしながら、デジタルマーケティング全般へと事業を拡大。マーケティングはいわば「物やサービスを売る仕組みづくり」です。この仕組みを「効率よく最適化することが、収益アップの最短ルート」であり、検索エンジンを最適化するSEOと本質は同じだと私たちは考えます。

研究と実践で培った技術を出し惜しみせず、お客様のために「今できる最高で全力のパフォーマンス」を心掛けることで、2024年現在は国内外あわせて2000社以上の支援を実現しました。