

検索エンジン

検索エンジン(Search Engine)とは、インターネット上に存在する Web ページや画像ファイルなどの情報を探するためのシステムのことです。

代表的なものに Google、Yahoo!、Baidu(百度)、Bing などがあります。

調べ物をするとき、パソコンやスマートフォンで、Google あるいは Yahoo!などを介して、とりあえずキーワード検索を試みる方が多いと思います。このような行為を「ググる」と呼んでいます。Google を使って検索をすることから生まれた造語です。

とりあえずググってみたときに、検索結果に表示されているページの内容が自分の意図通りでなかったり、上位には EC サイトなどばかりが並んでしまって、欲しい情報になかなか辿りつけなかったり、という経験をされた方も居られることでしょう。

検索エンジンにより正確に自分の意図を伝えるには、Google などの検索エンジンがどのような仕組みで動いているのかを理解して、検索キーワードを選ぶ必要があります。

検索エンジンの仕組み

検索エンジンには、ディレクトリ型とロボット型とがあります。

ディレクトリ型は、ディレクトリ(要覧)を Web 上に構築しておいて、検索要求時にはそのディレクトリ内を検索するものです。人力で、カテゴリに分けて、サイトの概要を記入したディレクトリが構築されているので、特定分野や地区などに限定したサイトを探しやすく、検索結果の一覧には質の高いウェブサイトが並ぶという特長があります。一方、ディレクトリの構築には人手がかかるので、検索対象となるサイト数を多くできず、結果として目的のサイトが出てこないなどの欠点も抱えています。

インターネットが一般に使われ始めた初期(1990 年代)の検索エンジンの多くはディレクトリ型でしたが、Web サイトが爆発的に増え始めた1990年代の終わり頃からは、ディレクトリ型に代わり、Google に代表されるロボット型が主流になりました。

ロボット型の検索エンジン(以下、単に検索エンジンと記述したものはロボット型の検索エンジンを指します)は、インターネット上に存在する無数ともいえる Web サイトの各ページを、リンクを辿って情報収集し、収集した情報をデータベースに分類・登録します。

Web ページから情報収集する仕組みをクローラ(Crawler)、データベースに分類・登録することをインデクシング(Indexing)といいます。

クローラは巡回する仕組みとご理解ください。世界中の Web サイトを巡回し終わるには時間がかかりますので、生まれたばかりの Web サイトがインデックスされるには、数日程度の時間がかかります。

クローラによって集められた Web ページの情報は、キーワード検索の処理がしやすいように、文字情報が分類されインデックス(索引)されます。また、「title(タイトル)」「h(見出し)」「meta description(ページ概要)」などの HTML タグが付されている文字列、画像、更新日時などの情報もデータベース化されます。

ユーザーが指定した複数のキーワード(文字列)に基づいて、検索処理を行うことをクエリといいます。クエリ(Query)とは問い合わせという意味です。

検索エンジンは、インデックスされた諸情報を全文検索した上で、指定されたキーワードに該当しそうな Web ページを、最も適切だと判断した順番に並べて表示します。どのような順番で Web ページを並べるかを決定するロジックのことを検索アルゴリズムといいます。

検索アルゴリズム

Google は「ユーザーの利便性を高める」ことを理念に掲げて、検索エンジンのアップデートを続けています。Google 検索の基本的な考え方は、「ユーザーにとって有益」と考えられる情報が、見やすく、読み手に理解されるように提供されているページを高く評価し、より上位に表示するというものです。

その考え方にに基づき、Google の検索アルゴリズムでは、ランキングシグナルと呼ぶ順位を決定するための多数の要素を分析しています。このシグナルは数百個もあるとされており、設定されている様々なシグナルがそれぞれ複雑に作用して、順位は決定されています。どのシグナルを重視するかといった細かい調整は日々行われていて、その内部がどうなっているのかは明かにされていません。

「ランキングシグナル」と呼ばれている要素には、例えば

- title 要素にどのようなキーワードが含まれているか？
- Web ページ内に検索キーワードに合致する文字列がどの程度存在するか？
- コンテンツの内容がコピーされたものとか品質の低いものではないか？
- モバイル端末(スマートフォン)に最適化されているか？
- Web ページがいつ更新されたものか？
- Web ページが良く見られているものか？

などというようなものが含まれています。

これらは、Google からアルゴリズムのアップデートとして追加されたことが発表されたり、公式のガイドラインに記載されていたりすることもあります。ほとんどは公表されていません。

主な検索エンジン

Yahoo!と Yahoo! JAPAN

Yahoo!は、1994年にスタンフォード大学の学生であった David Filo と Jerry Yang により開発されたディレクトリ型の検索エンジンで、インターネットの普及に大きな役割を果たしました。当時の Yahoo!の検索エンジンは掲載申請のあったものが審査を経てカテゴリ別に掲載されるような仕組みでした。

1995年にソフトバンクが米 Yahoo!株を一部買い取り、1996年4月から日本版にローカライズした Yahoo! JAPAN(以下、ヤフーと記述)の検索サービスを開始しました。ヤフーでは、ツリー構造のカテゴリから探す「ディレクトリ検索」と「キーワード検索」が提供されています。

Yahoo!の知名度、インターネット利用者人口の増加などの様々なプラス要因と、ソフトバンクの経営戦略が見事に当たり、ヤフーは日本における検索サイトの首位の座を固めました。さらに、検索サイトの集客力を武器に、天気予報、ニュース、Yahoo!メール、ショッピング、オークションなど、検索サービス以外のサービスを含めたポータルサイトとしての独走を始めました。

Google によるロボット型の検索サービスに対抗するため、Yahoo!もヤフーも自前でロボット型の検索サービスを始めましたが、やがて、ロボット型の検索エンジンを、Yahoo!ではマイクロソフトの Bing に、ヤフーでは Google に、切り替えています。

ヤフーは 2018 年 3 月にディレクトリ型の検索サービスを終了しました。

Google

1996年1月、スタンフォード大学の学生であった Larry Page と Sergey Brin による研究プロジェクトとして、Google のロボット型の検索エンジンが開発されました。1997年9月に google.com のサービスが開始され、ベンチャー企業への投資ファンドなどから資金を集め、1998年に Google 社が設立されました。2001年に日本法人のグーグル株式会社が設立されています。

「重要なページはそれだけたくさんのところに参照されている」という理論に基づく「ページランク (PageRank)」技術を使った検索システムの優秀さが評価されて、Google は、後発でありながら、検索エンジンの世界でトップに上りつめました。現在では、ヤフーを始め、数多くの検索サイトのロボット型検索エンジンとして採用されています。

Google は、「検索連動型広告」と呼ばれる画期的なビジネスモデルを生み出しました。インターネット上での広告ビジネスの一種なのですが、広告費が実際に発生するのは、エンドユーザーがキーワード検索画面から広告主のページを見た時に限るというものです。

多くのネットユーザーが Google の検索サービスを利用しますが、その検索するキーワードというのはユーザーが興味を持っている対象なので、そのキーワードを検索した人に対してのみ関連する広告を表示すれば非常に効果的です。広告主はどのようなキーワードに広告を出すのかを指定し、そのキーワードが指定されたときにも広告掲載費が発生する仕組みは、ネットユーザーにとって有意義な広告でありつつ、広告主にとっても広告にかかる費用負担が有効的なものになります。

Google 社はこのビジネスモデルの成功をきっかけに巨大な企業になりました。現在は検索エンジンだけでなく、さまざまな Web サービスを提供するほか、携帯電話や OS といった分野にも進出しています。

主な Google の Web サービスには次のようなものがあります。

- Gmail: 電子メール
- Google マップ: 地図、ルート検索、ナビ、ストリートビュー
- Google Earth: 衛星写真
- Google Document: ワードプロ文書、表計算、プレゼンテーションをブラウザで利用
- YouTube: 動画投稿サービス
- Google フォト: 画像投稿サービス
- Google Chrome: ブラウザ
- Android: 携帯端末向けの OS
- Google 翻訳: 翻訳サービス

Bing

Microsoft 社は、1998 年より MSN サーチという名称(その後、Live サーチに改名)で検索サービスを提供してきましたが、2009 年、検索エンジンを刷新し、Bing(ビング)と名称で、Web ページの検索に加え、画像、動画、地図、ニュース、ショッピングサイトの商品に特化した検索を行えるサービスを提供しています。

Yahoo!の検索エンジンは 2009 年より Bing が採用されています。

Windows 10 で提供されている検索ボックスと「Cortana」から利用できる検索エンジンは Bing です。また、Windows 10 のヘルプは Bing の検索結果経由で提供されています。

その他の検索エンジン

Ask.com、goo、Lycos などが細々と利用されています。

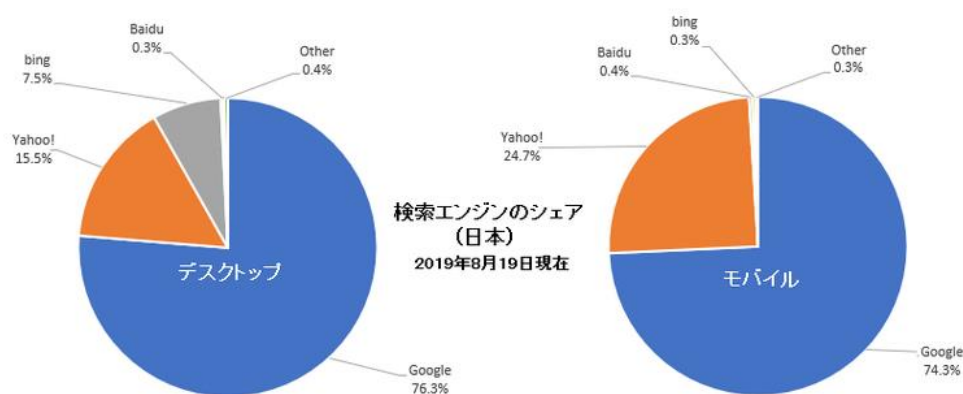
中国の Baidu(百度)、韓国の NAVER は、かつては日本語の検索サービスも提供していましたが、現在は日本語検索ができなくなっています。

既にサービスを終了したものには、初期の頃、日本人が開発した Odin、千里眼や、米国発の Excite、Infoseek などがありません。

(Excite、Infoseek は、ポータルサイトとしてのサービスを継続しています。)

日本での検索エンジンのシェア

下図の通り、日本でも Google が約 3/4 を占めています。ヤフーのロボット型検索エンジンには Google が採用されていますので、9 割強が Google で検索されていると言って過言ではありません。



日本における検索エンジンのシェア