



研修会特集 公開講座

インターネットをレファレンスに活かす

—検索の達人を目指そう—

関 裕司

サーチエンジンは誰でも簡単に使える便利なツールですが、イメージしたページが見つかるかどうかはその使い方によります。どのようなキーワードで検索するのがベストなのか。それはサーチエンジンの仕組みを正しく理解し、目的のウェブページの内容を具体的に想像することがポイントとなります。現在最も人気のあるサーチエンジン「Google」を例に、その仕組みと、キーワードの考え方を解説しましょう。

I. Google の仕組みを知ろう

現在 Google は全世界で 33 億ページという膨大な量のウェブページのデータを保有しています。このうち日本語で書かれたページは 1 億数千万ページにもなります。Google の情報収集の仕組みはウェブの本質である「リンク」を利用しています。「ロボット」と呼ばれるプログラムが起点のページからリンクを利用してページ間を移動し、各ページのテキスト情報を取得しながら世界中のデータを集めます。このようなロボットの情報収集を「クロール」と呼びます。Google の場合は基本的に毎月このクロールを行って、検索対象のデータを更新しています。

II. Google の検索対象と検索結果

SEKI Yuji

株式会社増進会出版社 情報システム部

yuji@shikencho.com

Google の主な検索対象データは次の 3 つです。

- ・ ページタイトル
- ・ 本文中のテキストデータ
- ・ 画像の ALT 属性に指定された文字

特に Google は「ページタイトル」を非常に重視しています。検索キーワードがページタイトルとマッチしていると、そのページを検索結果の上位にランク付けする傾向があります。検索結果の順位を左右する要因は他にもたくさんあります。ページ内のキーワードの個数、キーワードの位置、複数キーワードの位置関係、キーワードが HTML のどのようなタグで修飾されているか、などですが、最も重要な要因が「ページランク」と呼ばれるものです。Google は収集しているすべてのページに「点数」を付けています。これは各ページ間のリンク構造を利用した複雑な計算結果による点数付けですが、基本的に人気の高いページ、重要とされているページのランクが高くなります。

膨大な量の検索結果が出て目的のページが見つかりやすい、と言われている Google の秘密は、この検索結果の順位付けにあるのです。

III. キーワードの考え方

「〇〇について調べたい」、ならばキーワードは「〇〇」でいいか、というとそう単純ではありません。Google のようなロボット型のサーチエンジンはページの全文検索をします。キーワードの表現が少しでも違えば、まったく違

った結果を返してきます。この特徴を正しく理解して、次のようなアプローチをしてみることをおすすめします。

1. 専門用語、業界用語を使う

ごく一般的なキーワードでは結果が何万件も返ってきます。より詳しい情報のページを求めているのなら、その分野や業界に特有の専門用語、業界用語で検索するようにしましょう。

2. 特有のフレーズはないか

目的のページに特有のフレーズがないか、常に考える習慣をつけましょう。もしそのようなフレーズがあれば、それが検索の特効薬となります。例えば「CRM」という言葉の意味を調べたとしましょう。「CRM」だけで検索をしても CRM の意味を解説したページはなかなか見つかりません。そこで CRM の意味を解説したページに特有のフレーズを想像するのです。例えば「CRM とは」というフレーズはどうでしょうか。このフレーズの続きには CRM の意味がきつと書いてあるはずですが、意味不明の言葉に「とは」を付けるだけで、検索結果がその言葉の解説集になるわけです。

3. タイトルを考えずに本文を想像する

単語をいくつか組み合わせる検索をする。これが最も普通の検索パターンです。その場合のポイントは「タイトルを考えないこと」です。例えば「東京都内の税務署の住所一覧」を探す場合、皆さんが考えるキーワードのセットは次のようなものではないでしょうか。

「都内 税務署 住所 一覧」

これは目的のページのタイトルを想像していることとなりますが、実際には目的のページに「都内」や「住所」や「一覧」という言葉がある必然性はありません。「住所」は「所在地」という表記がされているかもしれません。ページのタイトルは表記の「ゆれ」が激しく、検索用のキーワードとしてはふさわしくないのです。タイトルのかわりに本文を想像してキーワードを決めましょう。

では、住所一覧の本文にあるキーワードは何か。それは住所そのものです。その中から確実にある文字を選び出すのです。都内の税務署の住所ならば、次のようなキーワードのセットが理想的です。

「税務署 千代田区 大田区 世田谷区 町」

4. キーワードを探せ

検索を一回で済ませようと考えてはいけません。最初から理想的なキーワードを設定することは不可能です。まず、検索。その中から自分が思いつかなかった専門用語、業界用語を探すのです。そして見つけたキーワードで再検索。この繰り返しで検索結果のリストは格段に情報の濃いものとなるでしょう。

IV. Google のオプション機能

検索結果の絞り込みの有効な手段としてオプション機能を使う方法があります。数あるオプション機能から便利な二つを紹介しましょう。

1. サイトの種類を限定させる

「キーワード site:ドメイン名」という形式で検索してみましょう。そのドメイン内だけを対象にした検索ができます。例えば日本政府のサイトから「図書館」を検索する場合は次のようになります。

「図書館 site:go.jp」

2. ファイルの種類を限定させる

ウェブ上には HTML 文書以外にも多数のファイルが公開されています。特にエクセルやワード文書は統計情報やビジネス文書のテンプレートなど、使えるファイルが数多く存在します。

「キーワード filetype:拡張子」という形式で検索してみましょう。拡張子はエクセルなら「xls」、ワードなら「doc」です。例えばエクセルのファイルから「人口」を検索する場合は次のようになります。

「人口 filetype:xls」

(本稿は、第 10 回日赤図書館協議会研修会での講演に加筆したものである。)