

検索エンジンはどうのようにして 有用なサイトと有用でないサイトを判別しているのか

木下研究室

林 潤 (200502716)

1 はじめに

インターネットの普及に伴い検索エンジンを利用するユーザは日々増えている。そのユーザに企業のホームページや販売サイトに来てもらうには検索エンジンの上位に表示されることが必要である。検索エンジンは、数あるサイトの中から有用なサイトと有用でないサイトを判別して順位付けを行っているが、正式な判別基準の情報は公開されていない。そこで本稿では、既に上位表示されているサイトなどから検索エンジンの判断基準を予想し、最も集客できるとされる、あるキーワードでの検索順位の1ページ目(1位~10位)以内に自分で作成したサイトが表示されることを目標に研究を行う。

2 検索エンジンの仕組み

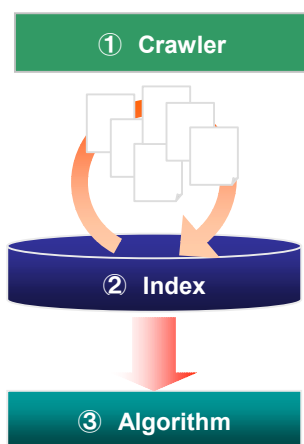


図.1 検索エンジン概略図

(i) Crawler と呼ばれる Robot が常にサイト内のリンクを伝って巡回している。

(ii) Crawler は、その情報をまとめて Index(一時的に情報を保管する保管庫のようなもの)を作成する。

(iii) 検索エンジンは、ユーザが検索したキーワードに対して、AlgorithmにしたがってIndexの情報から「関連性の高い」順に表示する。

検索エンジンは、ユーザの要求(キーワード)に対して適している全てのサイトを関連性の高い順に表示することが目的なので(i)~(iii)の動作を常に繰り返し行っている。

3 有用なサイトの判別方法

検索エンジンの Algorithm が有用なサイトと有用でないサイトを判別するには、内部要因(on the page factors)と外部要因(off the page factors)の2つの要因がある。内部要因と外部要因の概要は以下になると予想される。

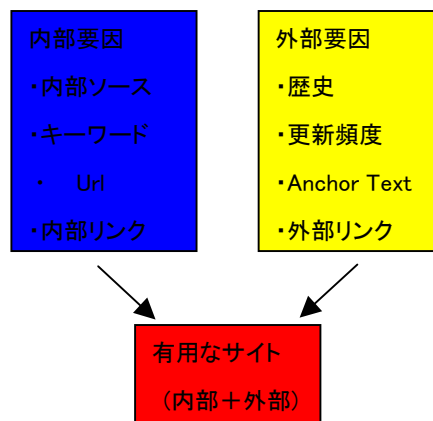


図.2 内部要因と外部要因

内部要因は、HTML タグ(内部ソース)、キーワード含有率(キーワード)、Domain 名(url)、内部リンク構成によって判別している。一方、外部要因は、ページができてからの経過年数(歴史)、更新頻度、被リンクの際の Text(Anchor Text)、外部リンクの質と量によって判別している。外部リンクの質と量というのは、Google の PageRank の「多くの良質なサイトからリンクをもらっているページは、やはり良質なページである。」という基本概念に従っている。

4 実験と考察

上記の要因に沿って「大学偏差値ランキング」というキーワードで上位表示を狙ってサイトを作成(2008/5/21)した結果、2008/1/25時点で、「Google9位、Yahoo33位、Google モバイル1位、Yahoo モバイル2位」に表示された。内部要因対策を施した時点では検索エンジンに Index されていなかったが、外部要因対策を施したことで Index されたので、検索エンジンは内部要因と外部要因の両方の観点から有用であるサイトとそうでないサイトを判別していることが分かった。また、PC とモバイルで検索エンジンの順位の違いが生じたのは、PC とモバイルでは、外部要因と内部要因の判断基準が違うからだと考えられる。本実験では、内部要因対策は十分に行え、外部要因対策の内、歴史などは期間的に短く不十分であった。このことから、PC の検索エンジンは外部要因を重視するのに対し、モバイルの検索エンジンは内部要因を重視することが順位から推測できる。今後検索エンジンは、個人によって異なる検索結果を表示するパーソナライズ検索、検索した場所によって異なる検索結果を表示する GPS 連動検索、動画や地図や画像などさまざまなコンテンツを表示するユニバーサル検索なども用いて有用なサイトかどうかを判別していくと予想されるのでそれらについても今後検討していく必要がある。