Webサーチエンジンの基本技術と最新動向（下） 最新技術

福島俊一¹

著者抄録：本稿は「Webサーチエンジンの基本技術と最新動向」と題した2部構成論文の後編である。前編ではWebサーチエンジンの背後にある基本的な情報検索技術を解説した。この後編ではWebサーチエンジン技術にかかわる5つの新しい話題を解説する。Googleで一躍有名になったリンク解析の改良・高度化が第1の話題、規模を追求してきた汎用Webサーチエンジンに対する目的特化やユーザ適応／状況適応という発展方向が第2・第3の話題、情報提供側の意図を反映した情報提示手法が第4の話題、サーチエンジン最適化（SEO）というコンサルテーション手法に着目したサーチエンジンスパムの問題が第5の話題である。

キーワード：Webサーチエンジン，Web情報検索，全文検索，Webロボット，ランキング，目的特化型Web検索，ユーザ適応型Web検索

Basic technologies and trends of Web search engines: Part 2 The recent topics

FUKUSHIMA Toshikazu¹

Author Abstract: This paper is the latter part of the series. The former part surveys fundamental information retrieval technologies utilized in the Web search engines. The latter part introduces recent topics related to the Web search engines technologies. The following five topics are described in this part; 1) advanced link analysis methods, 2) specialized Web search engine architectures, 3) user adaptive Web search methods, 4) recommendation links and advertisement, 5) search engine spam problem.

Key words: Web search engine, Web information retrieval, full-text search, Web robot, ranking, specialized Web search, user adaptive Web search

原稿受理 (2003-08-13)

(情報管理 46 (7): 436-445)

1. はじめに

本稿は、「Webサーチエンジンの基本技術と最新動向」と題した2部構成の論文の後編である。

前編の基本技術編では、インターネットの道案内システムであるWebサーチエンジンに関して、その背後にある基本的な情報検索技術を解説した。最初に伝統的な情報検索技術として全文インデックス方式や検索結果ランキング方式などを概観した上で、Webサーチエンジンに関して、その仕組

みや、Webロボット、並列全文検索、Web特有のランキング方式などの基本技術を紹介した。

後編であるこの最新技術編では、Webサーチエンジン技術にかかわる新しい話題をいくつかピックアップして解説する。第2節では、Googleで一躍有名になったリンク解析ベースのスコア計算法について、WiseNutやTeomaなどでどのような改良が加えられたかを紹介する。第3節と第4節では、規模を追求する汎用Webサーチエンジンとは異なる方向として、目的特化とユーザ適応／状況適応があることを示し、それらの具体的な事例も紹介する。第5節では、通常のWeb検索結果と連動する形で表示される広告やお薦めリンクなどの情報提供意図を反映した情報提示手法について触れる。第6節では、サーチエンジン最適化（SEO）というコ

¹日本電気株式会社 インターネットシステム研究所
(〒630-0101 奈良県生駒市高山町8916-47)

Tel. 0743-72-3756 E-mail: t-fukushima@cj.jp.nec.com

¹Internet Systems Research Laboratories, NEC Corporation
(8916-47, Takayama-cho, Ikoma-shi, Nara, 630-0101)

ンサルテーションビジネスに着目することで、特にWebサーチエンジンがスパムとみなして排除する類のWebページがあることを紹介する。

なお、前編の内容は教科書的な文献や学会論文に基づいたものであったが、この後編の内容は主にWeb上に公開されているニュース記事、解説記事、サイト紹介情報などを参考にしており、その性質の違いには留意いただきたい。

2. リンク解析の高度化

前編では第3.4節にて、Webサーチエンジンに特有のランキングファクタとして、ページ更新日時(新鮮度)、参照履歴(人気度)、リンク構造(引用度)、ページタイプなどが用いられていることを紹介した。これらのなかでも特にリンク構造の利用は、Google (<http://www.google.com/>)¹⁾がメジャーなWebサーチエンジンとして成長するにつれて、その効果が一般にも知られるようになった。同時に、リンク構造解析に基づくWebページのランキング方式はGoogleのキーテクノロジーであるが、Google以外にも同種の技術を売りにするWebサーチエンジンが現れてきた。なかでもWiseNut (<http://www.wisenut.com/>)とTeoma (<http://www.teoma.com/>)が最近特に注目されている。

本節では、Googleの方式を改めて紹介した上で、WiseNutやTeomaの方式がどのような点でGoogleの方式と異なるかを説明する。なお、ここではリンク解析技術に絞って説明し、Google、WiseNut、Teomaのシステム全体機能の比較までは行わない。また、WiseNutやTeomaの方式がGoogleの方式と比較してユーザの望みにマッチしたものであるかは評価が分かれるであろう。

2.1. Google : PageRank

Googleが実用化したリンク構造解析に基づくスコア計算法はPageRankと名付けられている(Googleの創業者であるS. BrinとL. Pageによるオリジナル論文²⁾に加えて、PageRankアルゴリズムに関する解説記事³⁾も発表されている)。PageRankの考え方は、前編の第3.4節で紹介済みであるが、改めて以下に説明する。

PageRankは、多数引用されるページは信頼できる、また、信頼できるページに引用されるページも信頼できる、という考えに基づいたスコアである。Webにおいては、WebページAからWebページBへのリンクが引用関係(AがBを引用)を表すの

図1 PageRankの計算方法

で、リンク構造を解析することで上記のような引用関係に基づくスコアを求めることができる。

図1がPageRankの計算方法を示す概念図である。各WebページのPageRank値は、そのWebページが他のWebページから受けているリンクの重みを累積することで計算する。これによって、多数引用されるほどスコアが高くなる。さらに、各々のリンクの重みとして、リンク元のWebページのPageRank値をリンク数で割った値を用いる。これによって、信頼できるWebページに引用されるとスコアが高くなる。このスコア計算は再帰的なものになるが、行列の固有値問題として扱うことができる(計算方法の詳細は文献^{2,3)}を参照)。

2.2. WiseNut : 文脈依存のリンク解析

WiseNutで用いられているリンク解析ベースのスコア計算法はWiseRankと名付けられている。WiseRankは、PageRankと同じリンク構造解析に基づくスコア計算をベースにしているが、Context-sensitive link analysis(文脈依存のリンク解析)という改良を加えている^{4,5)}。

PageRankは検索語に依存しない値である(Googleでは検索語との適合度スコアを別に計算してPageRankスコアと組み合わせている)。それに対して、各WebページのWiseRank値は検索語に応じて異なる値になる。検索語に関連の深いWebページからのリンクは、スコア計算における重みを大きくして扱われる。例えば、WebページAからWebページBへのリンクXがあり、検索語がKであるとき、ページA中のリンクXの周辺(ページBへのアンカー文字列の周辺)にKと関連の深いキーワードが出現している場合、リンクXの重みは大きくされる。WiseRankにおける文脈依存(Context-sensitive)とは、スコア計算対象Webページのリンク元Webページ中のキーワードを考慮するという意味である。

2.3. Teoma : トピック限定のリンク解析

Teomaで用いられているリンク解析ベースのスコア計算法はSubject-specific popularityと呼ばれる。Subject-specific (トピック限定) でリンク解析に基づくスコアを計算する点がポイントで、同じトピックのWebページ集合 (コミュニティ) のなかで、スコア計算対象のWebページが多数のリンクを受けているとスコアが高くなる^{6,7)}。すなわち、Googleではトピックの分野に依存しない形でPageRankスコアを計算するが、Teomaではトピックに応じて関連の深いコミュニティ内での評価を重んじてスコアを計算する。

3. 目的特化

Googleに代表されるロボット型Webサーチエンジンは、すでに70億ページを超えた世界中のWebページ群をできる限り多く収集するという規模拡大を目指してきた。この規模の拡大は、Webサーチエンジンで検索できるインターネット上の情報のカバー率向上を意味し、ユーザにとっては、さまざまなトピック (情報要求) に対して欲しい情報を見つけられる確率が高まるという大きなメリットが得られる。

しかし、規模を追求していくことは、いくつかの点で限界があり、また、ユーザにとってのデメリットを生む面もある。第1の問題は設備投資の限界である。年々倍増する勢いに追従してWebページを収集・インデックス化し、かつ、ユーザに対する検索レスポンスを維持していくためには、多大な設備投資 (コンピュータリソースの増強) が必要になる。第2の問題は情報鮮度の低下である。大量のWebページを収集するにはWebロボットによる巡回に長い時間が必要である。大手のWebサーチエンジンでは、現状、対象とするWebページのすべてを一巡するのに1か月近くを要していると言われる。したがって、Webサーチエンジンのインデックスに登録されたWebページは、場合によると1か月以上前の情報であるかもしれない。WebサーチエンジンでヒットしたURLにアクセスしてみると、その内容が変わってしまっていたり、そのWebページがなくなっていたりすることも多々ある。

このような問題に対する1つのアプローチが、目的特化型Webサーチエンジンである。目的・用途を絞り、検索対象とするWebページ群を限定することで、規模を抑え、更新サイクルも短くできる。このような目的特化型Webサーチエンジンは現状、

Googleなどの大規模な汎用Webサーチエンジンに取って代わるようなものではないが、ユーザの目的・用途に応じては効果的に活用できる。

3.1. 目的特化型Webサーチエンジンの事例

本節では、目的特化型Webサーチエンジンの事例を示す。汎用Webサーチエンジンと同様に目的特化型Webサーチエンジンについても、人手収集型とロボット型とがあるが、以下では、ロボット型の代表例をあげる。

(1) 論文情報検索

ResearchIndex (<http://citeseer.nj.nec.com/>)⁸⁾がよく知られている (図2)。インターネット上に公開されているコンピュータサイエンス分野の論文 (現時点で50万件、世界最大規模) を自動収集し、全文検索に加えて、論文間の引用関係の分析機能も提供している。

(2) 商品価格比較

ユーザが指定した商品について、インターネット上のどのショッピングサイトでいくらで売られているかという価格情報を一覧比較する機能を提供するサーチエンジンである。ShopBotやショッピングエージェントとも呼ばれる^{9,10)}。mySimon (<http://www.mysimon.com/>) やDealTime (<http://www.dealtime.com/>) がよく知られている。

(3) 評判情報検索

商品名、ブランド名、会社名、人名などを対象とし、インターネット上のWebページ (個人サイト、レビューサイト、掲示板など) に書かれたさまざまな人の意見 (評判) を自動収集するサーチエンジンである¹¹⁾。収集された意見は肯定意見であるか否定意見であるかも自動的に判別される (図3)。なお、このエンジンは、通常のWebサーチエンジンのような一般ユーザ向けサービスとしては公開しておらず、エンジンソフトウェアの販売という形態で実用化を進めている。

(4) 求人情報検索

インターネット上に発信されている企業のWebサイトなどから求人情報のみを自動収集するサーチエンジンである。FlipDog (<http://www.flipdog.com/>) では、25万件を超える求人情報を常時データベース化して公開しており、ユーザの希望する勤務地や職種などを条件に指定して、求人情報を絞り込むことが可能である。

(5) モバイル情報検索

iモードなどの携帯電話向けコンテンツのみを対象としたロボット型サーチエンジンである。Attayo

図2 論文情報検索エンジンResearchIndex

図3 評判情報検索エンジン

図4 目的特化型Webサーチエンジンの処理ステップ

ケイタイ（携帯電話からhttp://attayo.jp/で利用可能）では、iモードページの自動判別技術（精度99%）によって100万ページを超えるiモードページが自動収集されており、国内最大規模の携帯電話向けサーチエンジンとなっている。また、携帯電話からの使用における利便性を高めるため、コンテンツを地域別に自動分類して選別できる機能や、第5節(3)で述べる先読みナビゲーション機能なども提供されている^{12, 13)}。

3.2. 目的特化型Webサーチエンジンの仕組み

第3.1節で紹介した事例では個別に作り込まれたシステムが多いが、ここでは、目的特化型Webサーチエンジンの仕組みを、できるだけ汎用化・共通化して考えたものを示す。これは図4のように、ページタイプ選別処理と情報抽出・分類処理という2つのステップから成る。ページタイプ選別処理では、個々の目的に適合したページタイプのWebページのみを検索対象として切り出し、情報抽出・分類処理では、それらを目的に応じた分類軸で検索可能にする¹⁴⁾。

(1) ページタイプ選別処理

前編の第3.4節(4)で述べたように、個々の目的・用途に対して、多くの場合、それに適合するWebのページタイプが定められる。このページタイプ選別処理では、種々雑多なWebページ群のなかから、「ショップ」「掲示板」「イベント情報」「求人情報」「iモードページ」など、特定のページタイプに該当するもののみを取り出す。

(2) 情報抽出・分類処理

この段階では、特定のページタイプをもつWebページのみが処理対象になる。目的・用途に応じてページタイプを限定したとき、多くの場合、それらに固有のキー項目が存在する。例えば「イベ

ント情報」タイプでは開催日や開催地、「求人情報」タイプでは勤務地や職種などである。これらのキー項目は、ユーザの目的に即した分類軸となる。このようなキー項目の抽出は、特定ページタイプのWebページのみを処理対象に想定するならば、辞書とパターンマッチルールを用意することで、実用的な精度を得ることが可能である。情報抽出・分類処理では、Webページ中から上記のようなキー項目を抽出することで、イベント情報を開催日別・開催地別に分類して提示したり、求人情報を勤務地別・職種別に分類して提示したりすることを可能にする。

4. ユーザ適応・状況適応

第3節で取り上げた問題のほかに、汎用Webサーチエンジンでは、誰が使用しても同じクエリに対して同じ検索結果しか得られないという問題がある。ユーザの嗜好や状況に応じてランキングを変えることで、ユーザにいっそうマッチした検索結果を提供できる可能性がある。

そこで、本節では、ユーザの嗜好や状況に適応した情報選別の技術とシステム事例を紹介する。

4.1. 情報フィルタリング方式

ユーザの嗜好にマッチした情報を選別する情報フィルタリング方式には、内容ベースフィルタリング（Content-based filtering）と協調フィルタリング（Collaborative filteringまたはSocial information filtering）という2通りのアプローチがある。

(1) 内容ベースフィルタリング

文書（Webページなど）の内容とユーザプロフィールとをマッチングすることで、ユーザの嗜好に適合した文書を選別する方式である。図5の例のよ

図5 内容ベースフィルタリング方式

図6 協調フィルタリング方式

うに、文書とユーザプロフィールの両方をキーワード集合の形で表現し、それらの間の類似度を計算する。前編の第2.2.1節で紹介したベクトル空間モデルが適用できる。

(2) 協調フィルタリング

し好の類似した他ユーザからの推薦（いわゆるクチコミ）を自動化したものである。ユーザ間のし好の類似性は、各ユーザが過去に好んだもの／好まなかったものを履歴として蓄積・比較することで判定する。図6を例にすると、過去に読んだ文書に対する評価の傾向がユーザAとユーザCとで類似している。このとき、ユーザCが高く評価した文書3をユーザAは未評価（まだ知らない）なので、文書3をユーザAに推薦してあげると、ユーザAのし好にマッチする可能性が高い。

この協調フィルタリングでは、内容ベースフィルタリングと異なり、システムが文書の内容を見ることはしない。したがって、画像が中心でテキストがほとんど含まれていない文書に対しても適用できる。ただし、ユーザの誰もまだ評価を与えていない文書は処理対象にできない。このような長所／短所は、内容ベースフィルタリングとちよ

うど正反対であり（内容ベースフィルタリングは画像中心の文書は扱えないが、誰も評価していない文書でも処理できる）、協調フィルタリングと内容ベースフィルタリングを補完的に組み合わせると、より高い効果が得られる¹⁴⁾。

4.2. TPOベースの情報配信エンジン

第4.1節で説明した技術を用いて、Web上の情報をユーザ別に選別提示するシステムは、Amazon (<http://www.amazon.com/>) に代表されるショッピングサイト、あるいは、社員別にカスタマイズした情報を提供する企業ポータル（EIP: Enterprise information portal）などで実用化されている。さらに、携帯電話や携帯端末からのインターネット利用が普及したことから、ユーザのし好だけでなく位置情報や時間情報までも考慮した状況依存の情報選別のための方式・システムが開発されるようになってきた。

ここでは、その代表例としてTPOCASTを紹介する^{13, 15)}。TPOCASTのユーザは、そのとき、その場で、し好にマッチしたWebページ（レストラン情報、観光情報、店舗情報、広告などのPOI (Point of interest) 情報）を取得することができる。以下、図7に示したTPOCASTの2つの技術的特長を説明する。

(1) TPO別コンテンツ選択

第1の特長は、T（時間）・P（位置）・O（し好）を考慮してユーザとコンテンツをマッチングすることである。このようなTPOベースのマッチングを実現するために、各コンテンツ（Webページ）には、オープン時間（T）、所在地（P）、内容説明文・キーワード（O）などのTPO属性を事前に付与しておく（コンテンツとは別にXML形式で記録しておく）。一方、ユーザに関しても、ユーザの現在時間（T）、現在位置（P）、し好（O）などを追跡記録する。TPOCASTでは、これら2つのTPO属性に基づいて、コンテンツとユーザをT・P・Oの3軸で比較・スコア付けすることで、ユーザのTPOにマッチしたコンテンツを選択する。

(2) TPO別ユーザし好学習

第2の特長は、ユーザの行動からT・Pを考慮したユーザし好（O）を学習することである。第4.1節で説明した情報フィルタリング方式を、OだけでなくT・P・Oに拡張している。ユーザの行動を「いつ」「どこで」「どんなコンテンツに関して」「何をした」という4つ組と考え、「いつ」「どこで」はTとP、「どんなコンテンツに関して」「何をした」はOの条件として学習する。「どんなコンテンツに関

図7 TPOCASTの特長

して」の部分にはコンテンツに対応するキーワード群を抽出し、「何を(O)」の部分には行動の種類（「閲覧」「訪問」「購入」など）に応じてキーワードの重みを変えるという形で学習に反映する。

図8にユーザの行動からし好を学習した結果、ユーザに提示される情報が変化した具体的な事例を示した。同じ時間帯に同じ場所でユーザに提示される情報が、学習の進行によってユーザのし好を反映したものに変化していることがわかる。

5. 情報提供意図の反映

情報検索システムのモデルは、元来、クエリに対する文書の適合度を求めることを基本にしており、情報提供側の意図は特に考慮されてはなかった。しかし、昨今のWebサーチエンジンでは、通常の検索結果と連動する形で、情報提供側あるいは検索サイト運営側の意図を反映した情報提示が行われるようになってきている。具体的には、検索語連動広告、公式サイトリンク、先読みナビゲーションなどがその種の情報提供機能にあたる。このような機能は、単なる情報の押し売りではなく、情報提供意図が盛り込まれた結果がユーザにとってもメリットあるものになり得る。情報提供側が（対価を払っても）積極的に提示しようとする情報、あるいは、サイト運営側が責任をもって提供できる面があるためである。

(1) 検索語連動広告

検索結果ページに広告を表示する場合、検索語と関連の深い広告を表示した方が広告効果（クリック率）は高くなる。バナー広告のほかに、検索結果リストに並べて表示されるテキスト広告（スポンサードサーチ）がある。特に最近、検索語別に検索結果上位ポジションを入札制で広告主に販売

図8 TPOCASTにおける学習事例

し、ユーザのクリックごとに課金するCPC（Cost per click）の広告モデルが成長しており、Overture (<http://www.overture.com/>) やGoogleがCPC広告を販売している（GoogleのCPC広告はAdWordsという）¹⁶⁾。
(2) 公式サイトリンク

企業名・商品名・サービス名・人名などの検索語に対して、その公式サイトを表示する機能であり、検索語と公式サイトの対応関係を事前に人手でデータベース化（運営者側で収集または企業側から登録申請）しておくことで実現したものである。BIGLOBEサーチAttayo (<http://attayo.jp/>) では「的中ナビ（JWordリンク）」、FreshEye (<http://www.fresheye.com/>) では「ずばリンク」、goo (<http://www.goo.ne.jp/>) では「ぴったりgoo」という名称で呼ばれ、Webサーチエンジンの検索結果の先頭位置にこの公式サイトリンクが表示される。

また、上記のようなWebサーチエンジンと連動した形態とは別に、Webブラウザと連動した形態での機能提供も行われている。アクセスポート社が提供するプラグイン「日本語キーワードJWord」(<http://www.jword.jp/>) がそれである。JWordプラグ

NEC Business Solution
(<http://www.sw.nec.co.jp/>)に
おける先読みナビゲーション例

検索語「CRM」に対して、関連する
ソリューション情報や製品情報への
リンクを提示

BIGLOBEサーチAttayoケイタイ
(携帯電話から<http://attayo.jp/>)
における先読みナビゲーション例

検索語「モーニング娘。」に対する典型的な
目的「公式サイト」「着メロ」「CD」の各々に
的の中サイトへのリンクを提示

図9 先読みナビゲーションの例

インを使うと、WebブラウザのアドレスバーにURLの代わりに企業名・商品名などを入力することで、その公式サイトに直接ジャンプすることができる。

(3) 先読みナビゲーション

公式サイトリンクをさらに拡張したもので、個々の検索語に対して、検索語から推測される典型的な目的ごと到的中サイトを表示する機能である。各検索語に対してどのような典型的な目的があげられるかは、クエリのログから同時にどのような検索語が使われることが多いかを分析したり、サイト運営側がユーザをどう誘導したいかを整理したりすることによって決める。例えば、検索語が歌手名であれば「公式サイト」「ファンサイト」「着メロ」「CD購入」など、検索語が地名であれば「タウン情報」「グルメ」「ホテル」などが典型的な目的と考えられる。また、目的別的的中サイトのデータベースは、公式サイトリンクの場合と同様、事前に人手で作成する。

図9に先読みナビゲーションの例を示した。1つは企業のWebサイトにおけるサイト内検索機能に応用した例であり、もう1つは第3.1節(5)で紹介した携帯電話向けサーチエンジンに応用した例である。前者は企業にとって顧客へのソリューション提案を強化するのに役立ち、後者は携帯電話ユーザが検索時に、検索語入力や検索結果の絞り込み操作の煩わしさを軽減するのに有効である。

6. サーチエンジンスパムの排除

最近、Webマーケティング分野でSEOが話題になっている。SEOとはSearch engine optimization (サーチエンジン最適化)の略であり、対象とするWebサイトがターゲットとする検索語について主要なWebサーチエンジンで上位にヒットするように、そのWebサイトの作り方を最適化するテクニック、あるいは、そのためのコンサルテーションのことである。多数のSEOコンサルタント¹⁷⁻²⁰⁾が詳しい情報をWeb上に公開しており、また、解説書^{21, 22)}の類も出版されている。

あるSEOコンサルタント会社のWebサイト¹⁹⁾に掲載されている情報によれば、

- ・ 85%以上のインターネットユーザはサーチエンジンを介して商品やサービスを調べている (Internet.com 2001年11月)
- ・ 80%のインターネットユーザはサーチエンジンを介してWebサイトを見つけ出す (Yahoo! Japan サーベイ 2002年4月)

という調査結果があるそうである。Webサーチエンジンの利用される機会がこれだけ多いのであれば、Webサイト運営者・企業にとって、SEOは、自分のサイトにユーザを誘導し、ビジネスチャンスを拡大するためのマーケティング手法として重要なものになる。

SEOを実施するためには、Webサーチエンジン

のランキング方式を深く理解することが不可欠である。そして、SEOのテクニックは（Webサーチエンジンのランキング方式も）、本来、ユーザにとって良質のサイトとはどんなものかを見極めるのが取り組みの本質である。

しかし、この本質から外れ、ユーザのために良質なサイトを作ることを目指すのではなく、単にWebサーチエンジンの検索結果上位にランクインするためだけの小手先のテクニックに走るケースもあった。このような小手先のテクニックを導入したWebサイトはサーチエンジンスパムとみなされ、Webサーチエンジン側では、該当サイトにペナルティを課し、検索結果から排除する（ブラックリストに記録する）処置をとるようになった。SEOコンサルタントもこのようなサーチエンジンスパムとみなされるような行為は避けるようにとアドバイスしている。

以下には、サーチエンジンスパムの代表的なケースをあげる（SEOサイト17）の解説を参考にして簡単にまとめた。

(1) 隠しテキストスパム

人間には見えない形で、Webロボットによる収集だけに狙いを定めて、HTML本文中にテキストを埋め込む行為である。例えば、背景と同色の文字や判読できないサイズの文字を使って、人気のあるキーワードを埋め込むなどがある。

(2) 詰め込みテキストスパム

キーワードを詰め込むために、メタ属性や非表示要素を悪用する行為である。特にタグのALT属性が悪用されやすい（例えば、画像サイズに比して多数のキーワードが詰め込まれていたり、ページ内容に無関係のキーワードが埋め込まれていたりするなど）。

(3) 複製ページスパム

ほとんど同一内容のページを多数作成して、サーチエンジンの検索結果の上位大半を占めようとするような行為である。

(4) リンクスパム

リンク構造に基づくスコア（GoogleのPageRankなど）を高めるためだけに意図的にリンクを張り合う行為である。

(5) リダイレクトの悪用

人気のある検索語に最適化したページを用意してサーチエンジンに登録されるようにしておき、実際に人間が閲覧する場合には、その検索語と関

係の薄い別なページへジャンプさせる行為である。ただし、このような悪用ケースとは別に、最近ではFlashなどを用いた動的なWebページでサイトを構築するケースが増えてきていることから、サーチエンジン向けにドアページを設けることが必要になってきている。このとき、リダイレクトの悪用（サーチエンジンスパム）だとみなされないように、作り方に配慮が必要である。

7. おわりに

この最新技術編では、Webサーチエンジン技術にかかわる最近の話題を第2節から第6節で5つ取り上げて紹介した。以下、これら5つの話題ごとに、簡単なまとめと今後の展望を述べる。

第2節 リンク解析の高度化：GoogleのPageRankは検索語に依存しないスコアであるが、WiseNutのWiseRank (Context-sensitive link analysis) やTeomaのSubject-specific popularityでは検索語を考慮したリンク解析を行う。PageRankのアイデアはWebサーチエンジン技術における大きなブレイクスルーであったが、WiseNutやTeomaだけでなく当然Google自身によっても、さらなる改良が進められていくであろう。

第3節 目的特化：および第4節 ユーザ適応・状況適応：規模を追求する汎用Webサーチエンジンとは異なる方向として、目的特化とユーザ適応／状況適応があることを示した。今後、これらの2つを組み合わせ、よりカスタマイズを指向したシステムが作られていくであろう。ユーザの状況に応じて、目的にマッチしたWebサーチエンジンがきめ細かに使い分けられるような方向である。

第5節 情報提供意図の反映：情報提供意図を反映した情報提示という側面に目を向けた。ユーザの情報要求に基づいた従来の情報検索技術とは別に、情報提供意図という要素も取り込んだトータルなシステムとしてWebサーチエンジンは発展していく。

第6節 サーチエンジンスパムの排除：サーチエンジン向けにWebサイトを最適化するSEOコンサルティングに着目した。サーチエンジンスパムのような小手先のテクニックに走らずに、ユーザにとって良質なWebサイト作りを追求していくことが、今後ますます重要になっていく。

参考文献

- 1) 山名早人, 近藤秀和. サーチエンジン Google. 情報処理. Vol.42, No.8, 2001, p.775-780.
- 2) 馬場 肇. “Googleの秘密－PageRank徹底解説”. (オンライン), 入手先<<http://www.kusastro.kyoto-u.ac.jp/~baba/wais/pagerank.html>>, (参照 2003-08-10).
- 3) Brin, S.; Page, L. The anatomy of a large-scale hypertextual Web search engine. Proceedings of 7th WWW Conference. 1998.
- 4) WiseNut, Inc. “WiseNut search engine white paper”. (online), available from<<http://www.wisenut.com/pdf/WISEnutWhitePaper.pdf>>, (accessed 2003-08-10).
- 5) “Google キラー登場！次世代検索エンジン WiseNut を徹底検証！”. ネットランナー, 2002年3月号. (オンライン), 入手先<<http://www.zdnet.co.jp/internet/runner/0203/sp1/02.html>>, (参照 2003-08-10).
- 6) Ask Jeeves, Inc. “Teoma search technology - search with authority: the Teoma difference”. (online), available from<<http://sp.teoma.com/docs/teoma/about/searchwithauthority.html>>, (accessed 2003-08-10).
- 7) “Google キラーの大本命！？Teoma でディープ情報を集める！！”. ネットランナー, 2002年6月号. (オンライン), 入手先<<http://www.zdnet.co.jp/internet/runner/0206/sp1/part3a.html>>, (参照 2003-08-10).
- 8) Bollacker, K. D.; Lawrence, S.; Giles, C. L. CiteSeer: an Autonomous Web agent for automatic retrieval and identification of interesting publications. Proceedings of the Second International Conference on Autonomous Agents. 1998, p.116-123.
- 9) Doorenbos, R. B.; Etzioni, O.; Weld, D. S. A scalable comparison-shopping agent for the World-Wide Web. Proceedings of the First International Conference on Autonomous Agents. 1997, p.39-48.
- 10) Clark, D. Shopbots become agents for business change. IEEE Computer. Vol.33, No.2, 2000, p.18-21.
- 11) 立石健二, 石黒義英, 福島俊一. インターネットからの評判情報検索. 情報処理学会自然言語処理研究会, 2001, NL-144-11.
- 12) 河合英紀, 赤峯 享, 喜田弘司, 松田勝志, 福島俊一. モバイルサーチエンジンWithAirの試作と評価. 情報処理学会情報学基礎研究会, 2001, FI-64-11.
- 13) 福島俊一. モバイルユーザ向け情報選別配信技術. 情報処理学会高度交通システム研究会, 2002, ITS-9-6.
- 14) 有吉勇介, 福島俊一. 目的および個人に特化したサーチエンジンの開発. 人工知能学会誌. Vol.16, No.4, 2001, p.520-524.
- 15) 小川隆一, 二瓶克己, 伊東紀子, 茶園 篤. “NEC 技報. Vol.54, No.7, 2001, パーソナライズ情報配信プラットフォーム「TPOCAST」”. NEC技報. (オンライン), 入手先<<http://www.nec.co.jp/techrep/ja/journal/g01/n07/g0107mo.html>>, (参照 2003-08-14).
- 16) (株)イノベーション. “オーバーチュア／アドワーズの比較”. (オンライン), 入手先<<http://www.innovation.co.jp/service/net-comp.html>>, (参照 2003-08-10).
- 17) 住太陽. “SEO 検索エンジン最適化”. (オンライン), 入手先<<http://www.searchengineoptimization.jp/>>, (参照 2003-08-10).
- 18) SEO Japan. “検索エンジンマーケティングポータル”. (オンライン), 入手先<<http://www.seojapan.com/>>, (参照 2003-08-10).
- 19) Artisan Crew Engineering, Inc. “SEOによる検索結果上位表示”. SEO-WEB. (オンライン), 入手先<<http://www.seo-web.com/>>, (参照 2003-08-10).
- 20) Jupitermedia Corporation. “SearchEngineWatch.com?: The source for search engine marketing & optimization”. (online), available from<<http://searchenginewatch.com/>>, (accessed 2003-08-10).
- 21) 住太陽, アングラーズネット. アクセスアップのためのSEO：ロボット型検索エンジン最適化. 東京, エーアイ出版, 2002, 199p.
- 22) 渡辺隆広. 検索にガンガンヒットするホームページの作り方－SEO(検索エンジン最適化)テクニックで効果的にPRしよう. 東京, 翔泳社, 2003, 256p.