

# インターネット検索エンジンの技術革新と そのインターネット経済に与える影響

佐 薙 充・大 島 章

## 目次

- I. はじめに
- II. 検索エンジンの技術革新とインターネット広告のビジネスモデルの変化
- III. 検索エンジンのビジネスの拡大とその公正さに関する課題
- IV. 検索エンジンの技術的課題
- V. 検索エンジンの将来
- VI. 今後必要な取り組み
- VII. おわりに

## I. はじめに

インターネットは低価格で普遍的な通信インフラとして、企業活動および個人の生活にとって欠くことができない道具となった。米国の企業および消費者のインターネット利用に関する米国商務省の統計データ<sup>1)</sup>によれば、

- ・企業間の取引でのインターネットの利用は製造業・卸売業で普及度が高く、製造業の2001年の企業間取引額の18.3% (7250億ドル)がインターネットを使った取引である。
- ・消費者向けの商品の販売に関してはインターネットを通じた取引は、2001年において全売上額の1.1%に過ぎない。
- ・インターネットを通じたサービスの売上額の統計では、旅行予約、出版、証券取引、システムの設計開発の四分野が大きな比率を占めている。

日本でのインターネットの浸透は、最近2～3年で個人を対象とするブロードバンドインターネット契約が急増したことに特徴がある。DSLの加入契約数は2003年12月末に1000万を越えた<sup>2)</sup>。DSL、光ファイバなどのブロー

ドバンド網の利用者の急増により日本国内のインターネットの基幹網のトラフィックが急増しており、これを受けて基幹網のリニューアルの検討が総務省の委員会では始まった<sup>3)</sup>。

個人のインターネット利用の内容を見ると、米国政府による2001年9月の調査<sup>4)</sup>ではメールの利用が第1位で利用者は全人口の約45%を占め、次いで製品・サービスの検索が第2位で36%を占める。日本でもこの傾向は同様である。また、インターネットを利用した個人向けの商取引が急成長しており、米国ではAmazon.com, eBay, Yahoo!などがそれぞれ年間数千億円の売上をあげ、日本ではヤフー、楽天などが年間数百億円の売上をあげている。旅行予約、チケット販売などの業種でもインターネット経由の販売への依存度が増加している。

以上の事実から日米とも、

- ・インターネットの通信インフラは企業、個人に広く普及した。
- ・商品やサービスの販売のインターネット化に関しては企業間の取引が先行している。

の推移(速報)」、2004年1月30日、<http://www.soumu.co.jp/>

<sup>1)</sup> United States Department of Commerce, "E-Stats", March 19, 2003, <http://www.census.gov/estats/>

<sup>2)</sup> 総務省「インターネット接続サービスの利用者等

<sup>3)</sup> 総務省報道資料「次世代IPインフラ研究会の開催」、2004年2月2日、<http://www.soumu.co.jp/>

<sup>4)</sup> United States Department of Commerce, "A Nation Online", Feb. 2002

- ・消費者向けの販売でのインターネットの利用は、全取引額に占める割合はまだ低いが、着実に増加しており、特定分野では売上数千億円規模のビジネスとなっている。
- ・コミュニケーションおよび情報を探す手段としてのインターネットの利用は企業、個人を通じ広く浸透している。

ことがわかる。また日本では携帯電話を使ったインターネットの利用の普及率が高く、NTTドコモのiモードなどのインターネットにアクセス可能な携帯電話加入者数は2003年12月に6780万人に達しており<sup>5)</sup>、インターネット利用者の幅を大きく広げている。

パソコン、携帯電話いずれを使う場合も、インターネット利用の主要な用途はメールとWebに関連した情報の検索である。インターネットで何かを探す時に道案内役をつとめるのが「インターネット検索サービス」あるいは「インターネット検索エンジン (Internet search engines)」と呼ばれるサービスであり、通常無料で利用可能であり、雑多な情報の集合であるインターネットを有効に利用するためには不可欠なサービスである。本論文では、技術革新により利用者を増やし影響力を増したインターネット検索エンジンの現状を分析し、インターネットを利用する個人および企業にとってそれが何を意味するかを考察し、産業、研究、行政の場でどんな取り組みが必要かを論ずる。なお、以下の記述では、インターネット検索サービスのことを「検索サービス」、インターネット検索エンジンのことを「検索エンジン」と略記する。

## II. 検索エンジンの技術革新とインターネット広告のビジネスモデルの変化

### (1) ディレクトリ方式の検索サービスと Yahoo!

最初の本格的な検索サービスである Yahoo!<sup>6)</sup>はスタンフォード大学の学生が自分たちのために作った Web のリンク集 (有用な Web サイトのインターネット上のアド

レスを集めたもの) が発展し、1995年に企業化された。Yahoo! は無料で Web サイトに関する情報を利用者に提供し、ホームページに掲載する広告により収入を得るビジネスモデルを確立した。その後 Yahoo! は Web サイトの検索サービスだけでなく、地図情報、ニュース、金融市場の情報などの各種情報の提供、Web 上での商品の販売の場の提供などへ事業内容を拡大し、2003年には年間売上16億ドル、純利益2億3800万ドルを挙げ、インターネット上の巨大なポータルサイト<sup>7)</sup>企業に成長した。

初期の検索サービスは Yahoo! に代表されるように、定評のある Web サイトのアドレスを分野別に分類して提示するディレクトリと呼ばれる方式をとっていた。ディレクトリ方式では最新の情報を利用者に提供するために、ディレクトリの情報を常に人手で更新する必要があり、その作業が運営上の負担となる。また、ディレクトリには多くの人に関心を持つトピックが掲載されるため、情報の網羅性は不十分であり、特殊なトピックについて情報を探す時には役に立たないという欠点がある。

### (2) ロボット型検索エンジンと Google による技術革新

ディレクトリ方式の欠点を克服する方式としてロボット方式と呼ばれる検索エンジンが現れる。ロボット方式では、インターネット上の Web サイトの情報をコンピュータにより自動的に収集しデータベース化し、利用者が指定する検索キーワードを使ってこのデータベースを検索し、そのキーワードに関連する Web サイトを利用者に回答する。現在の検索サービスの多くは、ロボット方式であり、Yahoo! などのディレクトリ方式の検索サービス企業はロボット方式による検索サービスを併用している。「検索エンジン」ということばは通常このロボット方式の検索サービスのことを意味する。

<sup>5)</sup> 総務省「インターネット接続サービスの利用者等の推移 (速報)」、2004年1月30日

<sup>6)</sup> Yahoo! Inc. <http://www.yahoo.com/>

<sup>7)</sup> 利用者がインターネットを使う時に最初に接続する Web サイト。検索、メール、ニュースなどのさまざまなサービスまたはサービスへのリンクを提供する。

ロボット方式の代表的な検索サービスは Google<sup>8)</sup>であり、Google が先行して進めた検索技術の技術革新により、検索サービスの精度が向上し、検索エンジンの影響力が増え、インターネット上でのビジネスの構図が変わりつつある。

Google の技術革新の内容は、

- Web ページの重要度を自動的に計算する手法を開発し、それをもとに検索エンジンの回答の質を向上したこと。
- パソコンを使った低コストのシステムにより、Web ページ情報を収容する巨大なデータベースと検索要求に対して即座に回答を返す高性能な検索システムを実現したこと。

の二点に要約できる。

インターネットは情報の質という点では何の基準もなく、さまざまな質の情報の集まりであるため、ロボット方式の検索エンジンが Web ページの情報を自動収集するだけでは、大部分の情報は、利用者から見て役に立たない情報となる。Google の創業者たちは、この問題を解決するために、Web ページの重要度（情報の質のレベル）を自動的に計算する方法を考案した。これは PageRank<sup>9)</sup>と呼ばれ、Web ページ間のリンク（一つの Web ページから別の Web ページへのポイント）の分析により Web ページの重要度を評価する方法である。Google はこのアイディアを中核にして精度の高い検索サービスを実現した。これが前述の技術革新の一番目である。

Google の二番目の技術革新は、多数台のパソコンを使った徹底した並列処理により、検索システムの能力の拡張性を安いコストで実現したことである。利用者増にもかかわらず、Google が安定した検索サービスを提供し続けていることは、Google のシステム構築に関する試みが成功したことを実証している。Google のシステムは前例のない巨大な規模

となっている。Google の発表<sup>10)</sup>によれば、Google のデータベースは43億ページの Web ページを含む60億件のデータを収容している。また、Google は一日に2億件以上の利用者の検索要求を処理しており<sup>11)</sup>、大量の検索問い合わせに対して即座に回答を返すために、Google のシステムは世界各地の十箇所以上のデータセンターに分散設置された合計5万台以上のパソコンを使って実現されていると言われている<sup>12)</sup>。

これらの技術革新により、検索エンジンの回答の質が向上し、インターネット利用者は頻繁に検索エンジンを使うようになる。Google は1998年の創業以来利用者を増やし続け、第一位の検索サービスの地位を確立した。独立のコンサルタントが検索サービスに関する情報を発表している Web サイトである Search Engine Watch<sup>13)</sup>が会員の投票をもとに毎年選考している「優れた検索サービス (Outstanding Search Service)」のランキングで Google は2000年から4年連続して第一位に選ばれている。

Google は未上場の企業であり、その財務情報は公開されていないが、ニューヨークタイムズは Google の関係者の話として2001年にはゼロに近かった売上が2003年には10億ドル近い額になり、およそ3億5000万ドルの利益を挙げたと伝えている<sup>14)</sup>。米国の証券業界は Google がこの急成長を背景に IPO（株式新規公開）による巨額の資金調達を行うことを予想している。

### (3) キーワード広告によるインターネット広告のビジネスモデルの革新

検索の技術革新と並んで、検索エンジンのビジネスの観点から重要なのは、検索と直接関連づけた広告である。利用者が指定した検

<sup>8)</sup> Google Inc. <http://www.google.com/>

<sup>9)</sup> Page, L., Brin, S., Motwani, R. and Winograd, T., "The PageRank Citation Ranking: Bringing Order to the Web" (1998), <http://dbpubs.stanford.edu/pub/1999-66>

<sup>10)</sup> Google Inc. Press Release, "Google Achieves Search Milestone With Immediate Access To More Than 6 Billion Items", Feb. 17, 2004

<sup>11)</sup> Sullivan, D. "Searches Per Day", Feb. 25, 2003, <http://www.searchenginewatch.com/>

<sup>12)</sup> The New York Times, "The coming search wars", Feb. 1, 2004

<sup>13)</sup> <http://www.searchenginewatch.com/>

<sup>14)</sup> The New York Times, "The coming search wars", Feb. 1, 2004

索キーワードに直接関連した商品やサービスの広告を表示するキーワード広告は、利用者の意図と無関係なバナー広告等に比べてより効果的であることが認識されている。キーワード広告は、米国の検索サービス企業のGoTo<sup>15)</sup>が1998年に始め、その有効性を実証し、ビジネスモデルを確立した。現在ではGoogleを含む検索サービス各社が実施している<sup>16)</sup>。

急成長したGoogleの主要な収入源はキーワード広告である。さらにGoogleはキーワード広告を拡張したAdSenseと呼ぶ広告プログラムでビジネスを拡大しようとしている。AdSenseはWebサイトの内容に合った広告をGoogleが配信するサービスであり、検索エンジンの中核技術であるWebページの内容評価技術を使って、Googleが広告依頼主と広告表示Webサイトの間を仲介する新しい形のビジネスである。

米国のインタラクティブメディアの標準化とビジネス推進を目的とする業界団体であるIAB (Interactive Advertising Bureau) の発表によれば、米国でのインターネット上の広告の売上は、インターネットバブル崩壊後の低迷を脱し、2003年第4四半期(10月～12月)に前年比38%増の22億ドルになった<sup>17)</sup>。とくに検索サービスによるキーワード広告の売上の増加がめざましい。2003年第2四半期(4月～6月)のインターネット広告売上中のキーワード広告の比率は31%であり、この比率は前年に比べると3倍以上である<sup>18)</sup>。

<sup>15)</sup> GoTo は広告掲載を主目的とした検索サービスを1998年に開始し、2001年に社名を Overture Services に変更した。その後2003年に Yahoo! が Overture Services を買収した。

<sup>16)</sup> キーワード広告の形態は Google とそれ以外の検索サービスとで違いがある。他の検索サービスが検索結果の中にスポンサー企業の Web サイトを含めて表示するのに対して、Google は広告を検索結果と別の場所に表示する。Google の表示方法は広告の有無により検索結果を変更すべきでないという考え方にもとづく。

<sup>17)</sup> Interactive Advertising Bureau (IAB) Press Release, "Interactive Ad Revenue Set Record With Best Quarter Ever", Feb. 12, 2004

<sup>18)</sup> Interactive Advertising Bureau (IAB) Press Release, Nov. 11, 2003

Google の急成長とキーワード広告の急増とを考え合せると、検索エンジンの性能を上げ、利用者を増やした Google に代表される検索サービス企業が、キーワード広告という効果的な媒体を使ってインターネット上の広告収入を他のポータルサイトから奪っていると解釈することができる。

#### (4) インターネット大手ポータル企業の誤算と挽回策

Yahoo!, Microsoft (MSN), AOL などの大手のインターネットポータルサイト企業は、多種類の魅力的なサービスを自社の Web ページで提供することにより、多数のインターネット利用者を自社の Web ページに引きつけ、その利用者をターゲットにした広告の配信を収入源としてきた。これらの大手ポータルサイト企業は、これまで、検索エンジンは地味な機能でビジネス的な価値が小さいと見て、力を入れておらず、検索専門の企業にアウトソースするのが通例だった。しかし、前述の Google の躍進とキーワード広告の急増により、大手ポータルサイト企業も検索エンジンの重要性を認識し、企業買収、新たな開発プロジェクトなどにより Google に対抗する検索エンジンの技術を自社で用意しようとしている。

この動きは、

- ・検索エンジンの性能が見劣りすれば利用者を他のポータルサイトに奪われてしまう。
- ・キーワード広告に関するきめ細かいサービス (例えば、利用者のクリック回数に応じて広告掲載料金を決める) で差別化するためには検索エンジンの技術を自社で保有することが必要である。

という認識に基づくものと思われる。

#### (5) 米国および日本での検索サービスの状況

表1に米国の主要な検索サービスを示す。Yahoo! は2000年7月から検索エンジンとしてGoogleを使ったサービスを提供してきたが、2003年に有力な検索サービス専門の企業であるInktomiとOverture (旧称GoTo) を買収し、自社製の検索エンジン技術への切り換えを進めている。MicrosoftはGoogleに対抗してロボット型の検索技術の開発を進め

ており、Windows オペレーティングシステムとからめた戦略により Google に対抗することが予想されている。

表1 米国の主要なインターネット検索サービス

	利用率 (注1)	検索機能 (注2)	年間売上 (注3)
Google	39%	自社	非公開
Yahoo!	30%	Googleから自社に移行中(注4)	\$1.6B
Microsoft (MSN)	30%	Inktomi	\$2.1B
AOL	16%	Google	\$8.6B

(筆者作成)

(注1) インターネット利用者のうち2004年1月(1か月間)にそのサイトの検索機能を利用した人の比率。出典: Nielsen//NetRatings のプレスリリース(2004年2月23日)。http://www.nielsen-netratings.com/

(注2) 検索機能の実現形態。アウトソースしている場合はアウトソース先。

(注3) Microsoft は MSN 部門の売上、AOL は Time Warner の AOL 部門の売上を示す。

(注4) 2004年2月18日に自社技術による検索サービスを開始したことを発表した。

表2 日本語の主要なインターネット検索サービス

	利用率 (注1)	検索機能 (注2)	経営主体
Yahoo! Japan	85%	Google	ヤフー(注3)
Google	46%	自社	Googleジャパン
goo	37%	Google	NTT-X
Microsoft (MSN)	29%	Inktomi	Microsoft
infoseek	28%	Google	楽天

(筆者作成)

(注1) 自宅からのインターネット利用者の中、その検索サービスを利用している人の比率(複数回答可)。出典: インターネット白書2003。

(注2) 検索機能の実現形態。アウトソースしている場合はアウトソース先。

(注3) ヤフーはソフトバンクが42%、Yahoo Inc. が33%の株式を保有する上場企業である。

表2に日本語で利用できる主要な検索サービスを示す。インターネット白書2003<sup>19)</sup>によると日本での検索エンジンの利用率は、ヤフーが第一位(85%)、次いでGoogle(46%)、goo(37%)、MSN(29%)、infoseek(28%)の順となっている。ヤフー、goo(NTT-X)、infoseek(楽天)の経営主体は日本企業であり、それぞれ独自のサービスを提供しているが、検索エンジンはGoogleを使っている。従って、日本語の検索サービスの利用率上位5社すべてが、米国企業の検索エンジンを使っていることになる。またiモードなどの日本国内の携帯電話のインターネットサービスではGoogleの検索サービスが提供されている。

### III. 検索エンジンのビジネスの拡大とその公正さに関する課題

#### (1) 検索エンジンのビジネスの拡大

前述したように検索エンジンの主な収入源はキーワード広告だが、検索エンジンの利用者数が多いことをてこに、ビジネスを広げようとする動きが見られる。

GoogleはFroogleと呼ぶ、オンラインで販売されている商品の検索サービスを英語圏でテスト運用している。これは、利用者が商品の種類を指定すると該当する商品の一覧が販売元、価格などと一緒に表示される機能であり、消費者が商品購入前に各社の製品を比較検討する用途に特化した検索サービスである。またGoogleはGoogleの検索機能をソフトウェアから呼び出すインタフェース(Google APIと呼んでいる)を試行の形で提供を始めている。この機能を使うことによりサービス提供者がGoogleの機能に自らのサービスを追加したサービスを提供することが容易にできるようになる。

検索エンジンはWebページの情報収集を常時実施し、Webページの情報に関する巨大なデータベースを保持している。検索エンジン企業が新たなビジネスを展開する時に、このデータベースが大きな強みとなると思われる。

<sup>19)</sup> 財団法人インターネット協会監修『インターネット白書2003』(株)インプレス、2003年7月2日発行

## (2) 検索エンジン最適化

多数の利用者が使う検索エンジンが他の企業のビジネスに与える影響力も大きい。検索エンジンはインターネットで何かを探そうとする利用者がまず使うツールであり、売り出し中の企業で顧客への商品・サービスの販売の手段として Web を使っている企業にとっては、自社の Web サイトが検索エンジンの検索結果の上位に表示されるかどうかは、企業の業績を左右する死活問題である。このためこのような企業は自社の Web サイトのアドレスを検索エンジンに表示してもらうため、検索サービス企業と契約を結び、キーワード広告、有料の検索結果表示などを実施している。他方このような掲載料を払わず、自社の Web サイトの作り方を工夫して、検索エンジンに取り上げてもらうよう努力することを「検索エンジン最適化 (Search Engine Optimization; SEO)」と呼び、Web サイトの作成者の間では広く知られている<sup>20)</sup>。検索エンジンはインターネット上の Web サイトを定期的にスキャンして、Web ページに含まれている情報をもとにそのページの重要度の評価を行うので、SEO はその評価アルゴリズムを推測して、例えば、自社の Web ページの中に特定のキーワードを多数入れて検索エンジンの評価を上げようとする。しかし、このような行為は Web ページの重要度評価を歪めることになるため、検索エンジン側はその影響を排除しようとしている。

## (3) 検索エンジンの公正さに関する問題

検索エンジンの性能の向上により、利用者が増え、検索エンジンは大きな影響力を持つようになったが、このことが新たな問題を生む可能性がある。米国の FTC (Federal Trade Commission) は米国の消費者団体 Commercial Alert からのクレーム<sup>21)</sup>に応え、2002年6月に検索エンジン各社に対して、Web サイトのアドレスを企業から掲載料の

支払いを受けて表示する場合は広告であることが明確にわかるように表示するよう勧告した<sup>22)</sup>。また米国の雑誌ビジネスウィークは2003年10月の記事で、検索エンジンが掲載料の支払いを受けた Web サイトを表示順位上優遇している疑いがあり、公正さという点で問題だと指摘している<sup>23)</sup>。しかし検索エンジンが Web サイトをどのようなアルゴリズムで順位づけしているかは非公開であり、順位づけの妥当性を客観的に評価するのは難しい。また検索エンジンが何らかの意図をもって、提供する情報を歪めて利用者を誘導することも可能性がないとは言えず、検索エンジンは報道機関と類似した影響力を持っていることを認識すべきである。

## IV. 検索エンジンの技術的課題

Google を例にとり、現在の検索エンジンの技術的な課題について考察する。Google の検索システムの内容については Google の創業者の論文<sup>24)</sup>、検索システムの開発を担当している Google の従業員の論文<sup>25)</sup>および和文の解説<sup>26)</sup>を参考にした。検索エンジンは図1に示すように、以下の三つの構成要素からなる。

- ・インターネット上の Web ページを自動的にダウンロードするロボット (クローラ)。
- ・クローラ (Crawler) が集めた Web ページ (データ圧縮されている) とそれをキーワードから検索するためのインデックスとから

<sup>22)</sup> United States of America Federal Trade Commission, "Re Commercial Alert Complaint Requesting Investigation of Various Internet Search Engine Companies for Paid Placement and Paid Inclusion Programs", June 27, 2002, <http://www.ftc.gov/>

<sup>23)</sup> BusinessWeek, "Web Searches: The Fix Is In", Oct. 6, 2003, pp. 89-90

<sup>24)</sup> Brin, S. and Page, L. "The Anatomy of a Large-scale Hypertextual Web Search Engine", *Proc. of the 7th World Wide Web Conference* (1998), pp. 107-117

<sup>25)</sup> Barroso, L. A., Dean, J. and Hölze, U., "Web Search for a Planet: The Google Cluster Architecture", *IEEE Micro*, Vol. 23, No. 2 (2003), pp. 22-28

<sup>26)</sup> 山名早人・近藤秀和「サーチエンジン Google」『情報処理』情報処理学会、Vol. 42, No. 8 (2001年)、pp. 775-780

<sup>20)</sup> SEO については国内にも多数の Web サイトがある。例えば <http://www.sem-research.jp/sem/>

<sup>21)</sup> Commercial Alert News Release, "Commercial Alert Files Complaint Against Search Engines for Deceptive Ads", July 16, 2001, <http://www.commercialalert.org/>

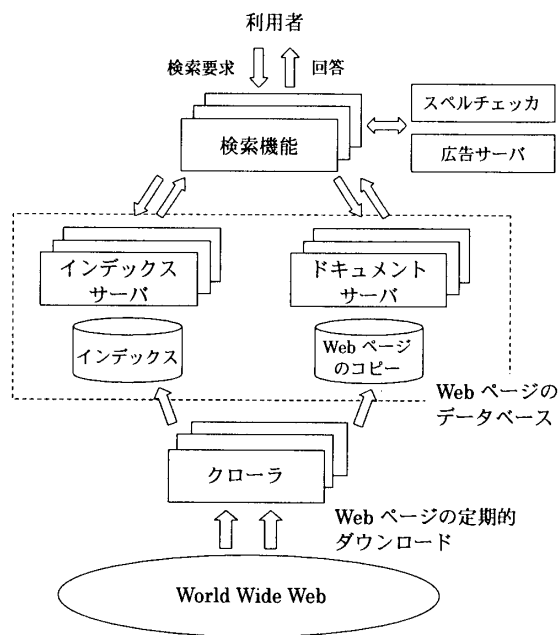


図1 検索エンジンの構成

(脚注24、25の文献の内容にもとづき筆者が作成した)

なる Web ページのデータベース (Web ページの重要度ランキングは Web ページからインデックスを作成する時に同時に計算される)。

- ・利用者の検索要求に応じて上記のデータベースを検索する検索機能。

Google に代表される現在の検索エンジンは多数台のコンピュータを使った並列処理と、Web ページの重要度評価技術により、膨大なデータに対して高い性能の検索を実現しているが、以下の点に課題が残っている。

(1) Web ページの更新を検索結果に反映するまでの時間の遅れ

インターネット上の Web サイトの数は膨大であるため、クローラが一回巡回するのに数週間かかる<sup>27)</sup>。このため、検索エンジンのデータは最新のデータではなく、Web サイトが更新されてからそれが検索エンジンのデータベースに反映されるまでに時間遅れがある。ただし、この点に関して Google は2003年の途中から更新頻度の高いページは頻りにダウンロードするよう方式を変更しており、時間遅れを少なくする努力が見

られる。

(2) 検索エンジンでは見えない Web 情報の存在

クローラは基本的には、静的に存在している Web ページを対象に情報収集を行うので、利用者の要求により動的に(要求があった時にそのつど)作成される Web ページは情報収集の対象とならない。例えば総務省は法令の条文をインターネット上で閲覧できるサービスを行っている<sup>28)</sup>が、この Web サイトは法令の条文を表示する Web ページを動的に作成するため、検索エンジンに対して法令の名前や条文中の語句を検索キーワードとして指定してもこの Web サイトのページを検索することはできない。また、書籍や CD の販売のように大量の商品データを保持する Web サイトでも同様なことが起きるケースがある。これらの例のように検索エンジンの検索の対象外の Web 情報は「見えない Web」、「Deep Web」などと呼ばれており、1999年時点の研究<sup>29)</sup>は、検索エンジンがカバーしているのは Web 全体の情報の15%程度であると推定している。またこの比率は1%以下だという主張もある<sup>30)</sup>。このような批判に代えて、最近では、検索エンジン各社は動的に作成されるページも情報収集の対象になると説明しているが、現実には機械的な情報収集だと情報の量が大量になるため、これはごく一部しか実現していない。

(3) 検索の際に指定するキーワードの巧拙により検索結果に大きな差がでる

これはフリーキーワード方式の古典的な情報検索でも共通する問題だが、検索エンジンを使っていて、非常に多数の Web サイトのアドレスが表示されるが大部分は自分

<sup>28)</sup> 法令データ提供システム、<http://law.e-gov.go.jp/cgi-bin/idxsearch.cgi>

<sup>29)</sup> Lawrence, S. and Giles, L., "Accessibility of information on the web", *Nature*, Vol. 400, No. 6740 (1999), pp. 107-109

<sup>30)</sup> Bergman, M., "BrightPlanet White Paper: The Deep Web: Surfacing Hidden Value", (2001), <http://www.press.umich.edu/jep/07-01/bergman.html>

<sup>27)</sup> <http://www.google.com/webmasters/>

が求めている情報とは無関係であることがよく起きる。この問題の解決法については後述する。

#### (4) 日本語の表記の問題

前項とも関連するが、日本語の場合、現在の検索エンジンは漢字の表記とかなの表記が同じことばを指している場合でも別なことばとして扱う。例えば Google で「インターネット広告のひみつ」という Web サイトを検索する時にキーワードとして「インターネット広告の秘密」と指定すると、求める Web サイトは検索結果に表示されない。この例のように、同じことばに対する表記の違いの吸収、語形（活用形など）の正規化など、日本語の情報検索の基本的な課題が検索エンジンでは十分に解決されてない。2003年10月に NTT-X の運営する goo が Google と提携した時の発表<sup>31)</sup>では、今後 goo は基本的な検索機能は Google のものを使い、日本語特有の問題に注力することを表明しており、上記の課題の解決に向けた一つの動きだと思われる。

### V. 検索エンジンの将来

#### (1) 寡占化に向かう動き

検索エンジンの利用者が増えれば、その影響力が増し広告収入が増加する。この力は利用者数の多い検索エンジンをさらに巨大にし、検索エンジンの寡占化を進める方向に働く。また高性能の検索サービスを提供するためには多額の研究開発資金と巨大な計算機システムを維持する費用を必要とし、これは検索サービスへの新規参入をむずかしくする。しかし他方で、利用者と検索エンジンとの関係に注目すると、利用者はほとんど何の負担もなく自分が使う検索エンジンを別なものに切り替えることができる。現に、複数の検索エンジンを併用している人が少なくない。検索エンジンのこの特性は、パソコンのオペレーティングシステムの場合とは大きく違っており、優れたサービスを提供する検索エン

<sup>31)</sup> 株式会社 NTT-X, Google Inc. プレスリリース「インターネット検索分野における戦略提携について」2003年10月2日、<http://www.goo.ne.jp/>

ジンが登場すると短期間に勢力地図が変わる可能性もある。これらを考え合せると、将来トップの検索エンジンが変わる可能性があるが、少数の企業による寡占化が進むことは避けられないと思われる。

#### (2) Web の発展と検索エンジン

企業間の取引でのインターネットの利用の増加に伴い、Web はその性格を変えつつある。従来の Web は人間が画面を見ながらコンピュータと交信する形態を前提としていたが、新しい Web ではこれに加え、コンピュータとコンピュータが一定の規約にもとづき交信する「Web サービス」と呼ばれる形態<sup>32)</sup>が重要な役割を占める。企業は Web サービスを使って他の企業と連携して（一部の作業は他の企業に依存し）自らの業務を実現する。従来の Web ではコンピュータから送られるデータは人間が見て理解できるものだったが、Web サービスでコンピュータ間で交信されるデータはコンピュータが簡単に解釈し、処理できるデータ形式であることが必要である。さらに不特定多数の相手と自由にデータをやりとりするためには、データ形式や用語を企業間で共通にすること（または機械的に変換できるようにすること）が必要であり、この標準化の問題が Web サービスの普及上の課題となっている。

次世代の Web への動きとしては、コンピュータが Web 情報の「意味」を解釈し、自動的な検索、情報収集、データ処理を可能にすることを目指した、「セマンティック Web」と呼ばれる研究<sup>33)</sup>が Web サービスとは独立に進められている。セマンティック Web でも各 Web ページはコンピュータが簡単に解釈できる形式でデータを提供することが前提であり、データ形式や用語の標準化など Web サービスと同様な問題の解決が必要となる。

現時点では上記のような課題が残っている

<sup>32)</sup> 丸山宏・小坂一也・浦本直彦「Web Services による動的な電子商取引の実現-SOAP/WSDL/UDDI-」『情報処理』情報処理学会、Vol. 42, No. 7 (2001年)、pp. 643-647

<sup>33)</sup> 萩野達也他「セマンティック Web とは」『情報処理』情報処理学会、Vol. 43, No. 7 (2002年)、pp. 709-717



が、今後 Web は Web サービスやセマンティック Web を包含したものに段階的に発展することが予想される。現在の検索エンジンは人間を対象とした Web ページを前提としているため、そのままでは、新しい Web の環境に対応できない。Web の発展に合わせて、検索エンジンの側にも新たな技術や手法が求められる。

## VI. 今後必要な取り組み

検索エンジンについて、ビジネスと技術の両面から分析してきたが、最後に、これらの分析をふまえ、産業、研究、行政の場で今後取り組むべき課題を以下に述べる。

### (1) 検索のエンジンの重要性

検索エンジンはビジネス的にも技術的にもインターネットのかなめとなるサービスであり、他分野への影響・波及効果の点から見ると、マイクロプロセッサやオペレーティングシステムと同等の重みを持つ。Web を使ってビジネスを行う企業にとって検索エンジンが重要であるのはもちろんだが、情報発信、意見の発表、報道などのコミュニケーションの手段としてインターネットを使う場合も検索エンジンの影響力は大きい。また、検索エンジンが収集し、保有する Web ページに関する膨大なデータベースは、さまざまな新しいビジネスにつながる可能性があり、大きな価値がある。

一方、現状のまま推移すれば検索エンジンは技術もビジネスも米国企業に独占される可能性が大きい。検索エンジンの重要性について認識を新たにして、日本の企業、学界、行政府が本格的に取り組むべきだと考える。具体的な取り組み施策については衆知を集めて検討すべきだが、いくつかの案を次項以下に述べる。

### (2) オープンな検索エンジンの追求

検索エンジンの技術が企業秘密のベールの中に隠され、ブラックボックス化することは公正な情報提示をチェックする観点から弊害が大きい。歯止めとして、オペレーティングシステムでの Linux に相当するオープンソース<sup>34)</sup>

の検索エンジンを育てる努力が有効である。米国を中心にその動きがあるが、Linux の場合と違って検索エンジンの場合は、検索エンジンのテスト・評価・運用のために、大規模な Web ページのデータベースが必要になるため、ボランティアベースの共同作業の中でこのような大規模なデータベースを作成し、日々それを更新する方法を考え出す必要がある。問題の完全な解決とはならないが、ひとつの検索エンジンが収集した Web ページのデータを他のプロジェクトと共有しデータを充実させることや、収集した Web ページのデータを公開し、それを使ったさまざまな実験を可能にすることなどが考えられる。

### (3) キーワード検索技術の改善

利用者が指定した検索キーワードに対して、適切な情報を Web から探し出すことが検索エンジンに課された使命である。Google の成功に触発され、Web ページの重要度のランキング評価手法については多くの研究が行われているが、利用者の指定したキーワードに最も合った内容の文書を探す方法の研究もそれに劣らず重要である。現状では、コンピュータは文書の内容を判定することはできないため、指定されたキーワードが文書のどの位置に何回現れるかといったレベルの情報をもとに文書を選択している。改善の方法としては、自然言語で表現されている検索キーワード、文書をコンピュータに解釈させる自然言語処理と呼ばれるアプローチと、Web サービスで進めているようにコンピュータに容易に解釈可能な形に Web ページのデータを整備、標準化するアプローチとがある。いずれの方法も簡単ではないが、実用化すれば多くの分野へ応用可能な技術であり、検索エンジンを実践的な実験場として使い、具体的な研究を進めるべきだと考える。

### (4) 検索エンジンの検索対象外の情報 (見えない Web) の取り込み

価値の高い情報を Web 経由で配信しているにもかかわらず、技術的な理由で検索エンジンの検索にかからないのは不合理である。この問題は Web サイトと検索エンジン双方の努力により改善可能だと考える。Web サイ

<sup>34)</sup> 青山幹雄「オープンソースソフトウェアの現状」『情報処理』情報処理学会、Vol. 43, No. 12 (2002年)、pp. 1319-1324

トの側が検索エンジンの特性を意識して Web ページを作る (例えば、動的に作成されるページに含まれるキーワードを静的なページとして別に用意する) ことにより問題を改善できる。検索エンジンの側は動的に作成されるページを収集する方法 (例えば、動的ページを作成する Web サイトから所定の形式で定期的にデータの提供を受ける) を工夫することにより動的なページを検索対象に含めることができる。

### (5) 検索エンジンの評価方法の研究

検索エンジンの検索結果の妥当性を評価する方法があれば、検索エンジンの優劣、特徴がわかり、利用者の側には大きなメリットがある。検索エンジンが公正な検索結果を返しているかどうかをチェックする手段にもなる。古典的な情報検索ではテストコレクション<sup>35)</sup>という手法により検索システムの評価が行われているが、これは検索対象の文書の集合を決め、正解の解っている問い合わせを検索システムに送ることにより可能となっている。これに対し、検索エンジンの評価では、検索対象の Web ページの集合を小規模な実験環境に限定すると実際の利用環境と大きく変わってしまい、意味のある評価にならない。従来行われている検索エンジンの評価<sup>36)</sup>は、インターネットの実環境で、所定の検索キーワードに対して検索エンジンが回答する文書数により評価するものが多い。この方法は機械的な評価が可能で客観的ではあるが、検索エンジンが回答する文書の大部分は利用者から見ると的はずれであるのが実態であることを考えると有効な評価方法とは言えない。評価にある程度主観が入るのはやむを得ないものとし、あらかじめ決めた検索キーワードと評価項目にもとづいて、評価者が検索エンジンからの回答を評価する方法が有効である。

<sup>35)</sup> 神門照子「情報検索システムの評価プロジェクト：NTCIR ワークショップ」『情報処理』情報処理学会、Vol. 41, No. 6 (2000年)、pp. 689-697

<sup>36)</sup> Lawrence, S. and Giles, L., "Accessibility of information on the web", *Nature*, Vol. 400, No. 6740 (1999)。また、国内の Web サイト「検索デスク」が定期的に検索エンジンの評価結果を発表している。http://www.searchdesk.com/

このような評価方法は消費者団体による商品の評価や、論述形式の試験の採点と類似した方法であり、検索エンジンの評価に対しても経験を積むことにより有効な評価ができるようになると思われる。

### (6) 法規制、意見発表の場の設定

検索エンジンが報道機関と似た影響力を持ち、少数の企業による寡占化が進むことが予想されるため、検索サービスの中での広告の取り扱いを含め、検索エンジンが公正な情報を利用者に提供しているかどうかを、法令や公的機関により監視、規制することを考える必要がある。また、法的な監視・規制ではなく、検索エンジンの関係者 (検索サービス企業、検索エンジンに影響を受ける企業、利用者、研究者など) が参加して、さまざまな意見発表や議論を行う場を設けることも有効である。

## Ⅶ. おわりに

米国のインターネット業界では検索エンジンとくに Google が注目の的となっており、インターネットバブル崩壊後停滞しているシリコンバレーに活気を取り戻す救世主として期待されている。これに対して、日本では検索エンジンに自社の名前を表示させる技法についての情報が氾濫しているが、検索エンジンそのもののインターネット経済の中での重要性、発展性については、一部の専門家を除いて関心が薄いように見える。この問題の重要性を訴え、企業、研究、行政の場での取り組みを本格化させるための一助となることを願って本稿を執筆した。

本論文では詳しくとりあげなかった、検索エンジンが収集した情報を使った新しいビジネスの可能性、検索エンジンが他のサービス企業をたばねる形で新しいタイプのポータルサイト企業に発展する可能性などは興味深いテーマであり、今後の研究課題だと考える。

### 参考文献

- Page, L., Brin, S., Motwani, R. and Winograd, T., "The PageRank Citation Ranking: Bringing Order to the Web"

- (1998), <http://dbpubs.stanford.edu/pub/1999-66>
- Brin, S. and Page, L. “The Anatomy of a Large-scale Hypertextual Web Search Engine”, *Proc. of the 7th World Wide Web Conference* (1998), pp. 107-117
  - Barroso, L. A., Dean, J. and Hölze, U., “Web Search for a Planet: The Google Cluster Architecture”, *IEEE Micro*, Vol. 23, No. 2(2003), pp. 22-28
  - 山名早人、近藤秀和「サーチエンジン Google」『情報処理』情報処理学会、Vol. 42, No. 8 (2001年), pp. 775-780
  - Lawrence, S. and Giles, L., “Accessibility of information on the web”, *Nature*, Vol. 400, No. 6740 (1999), pp. 107-109
  - 丸山宏、小坂一也、浦本直彦「Web Services による動的な電子商取引の実現—SOAP/WSDL/UDDI—」『情報処理』情報処理学会、Vol. 42, No.7(2001年), pp.643-647
  - 萩野達也他「セマンティック Web とは」『情報処理』情報処理学会、Vol. 43, No.7 (2002年), pp. 709-717
  - 青山幹雄「オープンソースソフトウェアの現状」『情報処理』情報処理学会、Vol. 43, No. 12 (2002年), pp. 1319-1324
  - 神門照子「情報検索システムの評価プロジェクト：NTCIR ワークショップ」『情報処理』情報処理学会、Vol. 41, No.6(2000年), pp. 689-697