

サーチエンジン Google

早稲田大学理工学部情報学科

山名 早人 yamana@yama.info.waseda.ac.jp

近藤 秀和 hide@muraoka.info.waseda.ac.jp

Google^{☆1}は、世界最大の情報を持つサーチエンジンとして有名である。Googleは、スタンフォード大学コンピュータサイエンス学科の研究プロジェクトとしてスタートした後、シリコンバレーの2大ベンチャーキャピタルから総額2,500万ドルの投資を受け、博士課程の学生であった当時25歳のLarry (Lawrence) PageとSergey Brinの2人が1998年9月に会社として起業した。2001年3月末時点で8,000台のPCを使い、クラスタコンピューティングにより7,000万件/日の検索を1件あたりほぼ0.5秒で処理する。以下では、Googleのこれまでの軌跡とGoogleの技術をアーキテクチャ面とソフトウェア面の両面から解説すると共に、Googleのビジネスモデルについても紹介する^{☆2}。

Googleの誕生とこれまでの軌跡

Google社は、Webブラウザで有名なNetscape社^{☆3}と同じ米国カリフォルニア州サンノゼ市のマウンテンビューに位置する(図-1)。

スタンフォード大学でデータマイニングに関する研究を行っていたPageとBrinは、当時急速にその利用が伸びてきたWWWに興味を持ち、Webデータを研究の対象とするようになった。データマイニングの研究には膨大なディスク容量とコンピューティングパワーが必要となるが、Webデータの急速な増加に対して、彼らが使えるディスク容量やコンピューティングパワーは十分なものではなく、同じスタンフォード大学から誕生したYahoo!^{☆4}の成功事例に倣ってビジネスプランを2年もの間書き続けた。

ビジネスプランを送った相手の1人である、Sun Microsystems社の共同設立者でありスタンフォード大学出身でもあるAndy Bechtolsheimに、ある日の早朝、PageとBrinは会うことになった。会合では、PageとBrinが彼らのビジネスプランを説明し終わらないうちに、Bechtolsheimは携帯電話でどこかに電話した後、「十分にビジネスプランは聞いた。あと残りの時間で話し続けるより、この小切手でビジネスをすぐに始めてみないか」と2人に告げ、10万ドルの小切手を渡した。これがGoogle社の起業の第一歩となる。

その後、シリコンバレーで有名なKleiner Perkinsと

☆1 <http://www.google.com/>

☆2 本稿は、筆者が2001年3月末にGoogle社の技術担当取締役のCraig Silverstein氏および主席渉外部長のDavid Krane氏とディスカッションした際の内容、および参考文献に示される文献の調査結果に基づく。

☆3 <http://www.netscape.com/>

☆4 <http://www.yahoo.com/>

☆5 これら2大ベンチャーキャピタルは、Apple, Amazon.com, Cisco Systems, Netscape, Yahoo!へも巨額の出資を行っている。

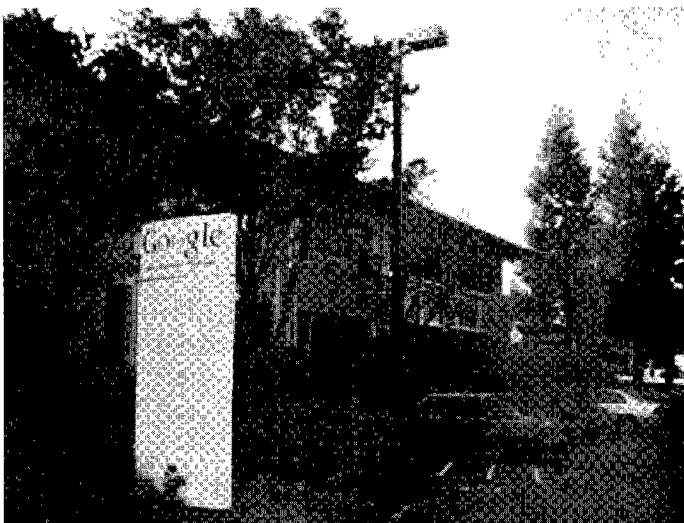


図-1 Google社のエントランス

Sequoia Capitalの二大ベンチャーキャピタル^{☆5}から総額2,500万ドル(約30億円)の投資を受け、1998年9月にGoogle社が誕生した。

図-2に、Google社で稼働するPC台数と収集Webページ数の推移を示す。当初30台のPCでスタートした後、2001年3月には8,000台に達している^{☆6}。一方で、Googleが収集するWebページ数もPC台数とほぼ同じペースで増加しており、当初5,000万ページであったものが、2001年2月には13.5億ページに達している。

Googleに対するアクセス数(検索要求)も飛躍的に伸びており、1999年6月時点で50万アクセス/日であったものが、2000年6月には4,000万アクセス/日、2001年3月末時点では7,000万アクセス/日となっている。これは、1999年6月のNetscape、2000年6月のYahoo!^{☆7}、2000年12月のBiglobe^{☆8}との提携が大きな要因となっている。また、2001年4月には@Nifty^{☆9}が提携を開始した^{☆10}。

Google社内には、毎日のアクセス数の推移を起業当時から記録した巨大な手書きのグラフがある。これによれば、2001年3月末時点で、全アクセス数の約1/3がYahoo!からであり、さらに1/3がGoogleのホームページから、残りの1/3が提携している各種サイトからのアクセスとなっている。

Googleのアーキテクチャ

Google社の技術面を取り仕切るのは、技術担当取締役のCraig Silversteinである。また、運用は、運用技術主幹のJim Reeseが取り仕切る。

■PCとLinux OS採用の理由

Googleが持つ8,000台の各PCは、Intel社のPentium III プロセッサ(533MHz~800MHz)、Maxtor社の80GBのIDEディスク2台、256~512MBのメモリから構成される。また、OSは、Red Hat Linuxの標準ディストリビューションからグラフィック部分を除いたものを採用している。

このように、一般的なPCとLinux OSというコモディティ化しているハードウェア、ソフトウェアにより、検索システムを構築する理由は、①低コスト化、②スケールアップの容易性、の2点である。

PCはその構成要素であるCPU、RAM、HD共に年々価格が低下しており、価格性能比から考えてWWWサ

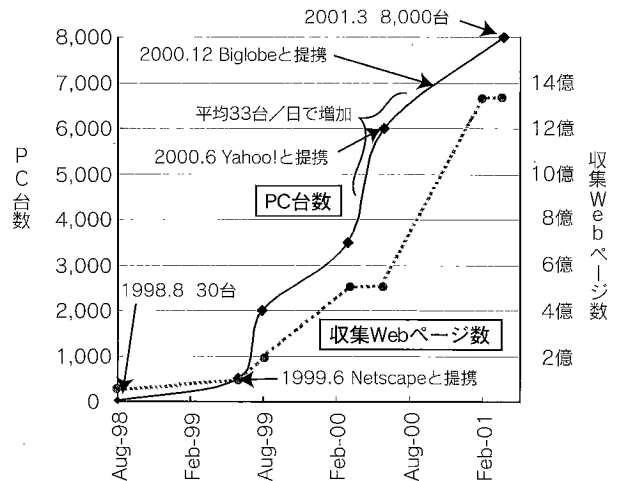


図-2 稼働PC台数と収集Webページ数の推移

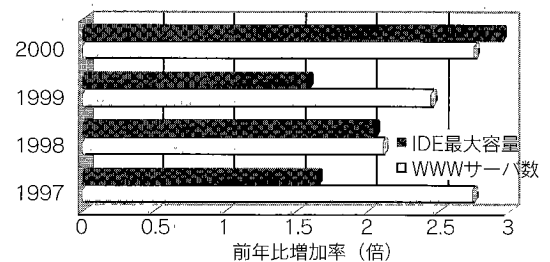


図-3 IDEディスクの最大容量とWWWサーバ数の前年比増加率^{☆11}

ーバ数の増加に対して十分にスケラビリティが保てると判断している。実際、ハードディスクの密度は年率2倍以上の伸びを示しており、WWWサーバ数が年率約2倍程度で増加していることを考えると、十分に容量的にスケラビリティが保たれる(図-3)。

■クラスタ構成

Google社で利用しているハードウェアは、RAIS (Redundant Arrays of Inexpensive Servers)^{☆12}構成をとる。19inchラック(210cmH×60cmW×75cmD)に1UサイズのPCを80台設置する。通常であれば44Uのラックには、1UサイズのPCが44台しか入らないが、Rackable Systems社^{☆13}の技術を用いることによって、ラックの表面と裏面の両方から奥行きが通常の半分のPCを挿入し、80台のPCを1ラックに格納する(図-4)。ただし、発熱量が大きくなるため空調用のファンをラック上部に設置している。ラック内の80台のPCは、

^{☆6} グラフでは滑らかに増加しているように示されているが、実際には一度に数千台規模を繰り返している。

^{☆7} Yahoo!のサイトに検索結果がなければ、自動的にGoogleの検索エンジンが利用されYahoo!の検索結果として表示される。

^{☆8} <http://www.biglobe.ne.jp/>

^{☆9} <http://www.nifty.com/>

^{☆10} Netscape、Biglobe、@NiftyのサーチエンジンのバックエンドとしてGoogleが利用されている。

^{☆11} WWWサーバ数は英国Netcraft社発表のデータ、IDEディスクの最大容量は米国Maxtor社が販売するIDE-HDの最大容量から算出した。

^{☆12} RAIDと同じ考え方であり、「安いPCを並べて冗長性を持たせたサーバ」という意味。

^{☆13} <http://rackable.com/>

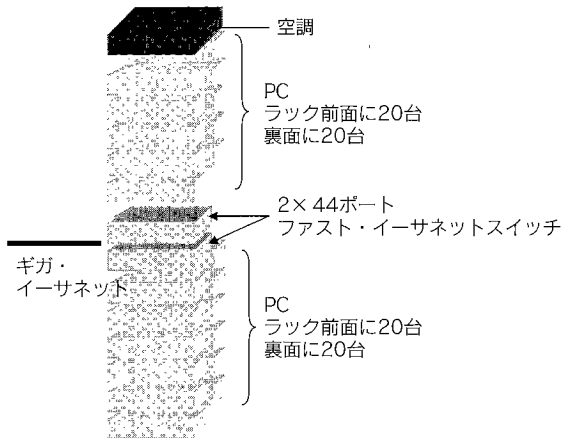
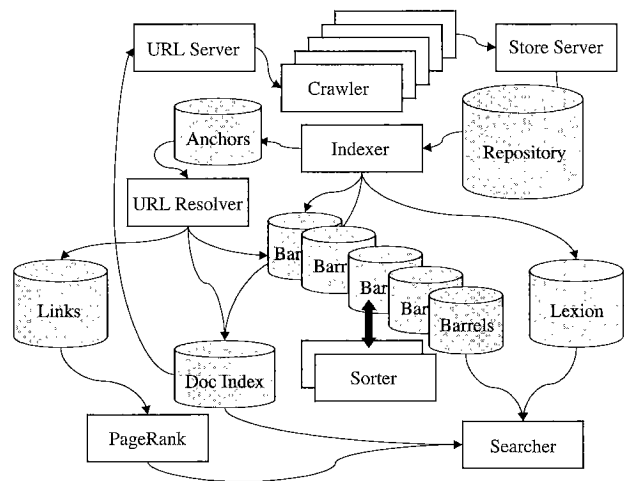


図-4 ラック構成



(文献1)より引用
図-5 Googleのソフトウェア構成

2台の44ポートのファスト・イーサネットスイッチにより接続され、ラック間はギガ・イーサネットにより接続される。4つのラック(PC320台)で1クラスタを構成し、各クラスタはまったく同一の構成で、同一のデータを持つ。

2001年3月末現在、シリコンバレーに2カ所、シカゴ、アトランタ、ニューヨークにそれぞれ1カ所の合計5カ所にデータセンターを設置し、各データセンターが1,000～3,000台規模のPCを備える。世界中からのアクセスは、これらのデータセンター単位で分散され^{☆14}、次に各データセンター内のクラスタへと負荷分散がなされる^{☆15}。今後、アジアとヨーロッパにデータセンターが追加される予定となっている。

Googleの検索技術

検索ソフトウェアの全体構成を図-5に示す。図-5は、スタンフォード大学でGoogleの運用を行っていた1998年当時の構成である¹⁾。2001年3月末時点での最新のソフトウェア構成は不明であるが、現時点もほぼ同じ構成をとるため^{☆16}、以下ではこの1998年当時のソフトウェア構成について解説する。

■検索ソフトウェアの全体構成¹⁾

図-5に示すように、Webページは複数のCrawlerにより収集され、Store Serverにより圧縮された後、Repositoryに格納される。Store Serverは、新規に見つかったWebページに通し番号(docID)をふった後、Repositoryに格納する。

索引付けは、IndexerとSorterが担当する。Indexerは、Repositoryに格納された圧縮されたWebページを

伸張し、Anchorの抽出、wordの抽出、word位置情報の抽出、フォントサイズ・大文字小文字の区別を解析し^{☆17}、これらの情報と共にdocIDからwordIDへの索引を作成しBarrelsへ格納する。Sorterは、上記の索引を元にwordIDからdocIDを得るための逆索引を作成し、Barrelsへ格納する。

wordからwordIDへのハッシュテーブルは、Lexiconとして持つ。Lexiconは各PCのメモリ上に載る大きさに調整されており、文献1)では256MBに調整するためにword数を1,400万に限定している^{☆18}。索引作成においては、あらかじめ作成済みのLexiconを用い、複数のPC上で並列処理する。この際、Lexicon上にないwordについては、重複なく新規にwordIDを付加する必要があるため、一時ファイルに書き出し、後で順に処理する。

Webページ内のAnchor情報は、Anchorsに格納され、URL Resolverにより相対URLは絶対URLに変換される。この時、Webページ間の相互のリンク情報はLinksに格納し、後述するPageRankを算出するのに用いる。また、コロラド大学のWWW(World Wide Web Worm)^{☆19}で実現されているように、Anchor部分のテキストは、そのAnchorが指し示すWebページを説明するwordとして、Barrels内の索引に追加される^{☆20}。

Doc Indexには、docIDからRepositoryへのポインタが格納される。

2000年9月からGoogleでは毎月第3週に前月にCrawlerが収集したデータに対して、図-5中の各種データ更新を行っている。このため、あるWebページが更新された場合、実際の検索に反映されるのに60～90日を要している^{☆21}。

☆17 現在のGoogleでは大文字小文字を区別していない。

☆18 現在のインプリメンテーションは不明。したがって、各国語ごとにLexiconを作成しているかどうかは分からない。

☆19 <http://www.cs.colorado.edu/www>

☆20 あるWebページへのリンク元のanchor部分のテキストを、そのWebページを説明する重要なキーワードと位置づける。

☆14 ラウンドロビンDNSを用いる。現在、Border Gateway Protocolの利用へと移行中。

☆15 データセンター内での負荷分散は、Custom load-balancing softwareにより行われているが、技術的詳細は公開されておらず不明。

☆16 Craig Silverstein氏によれば、ほぼ同じ構成とのことである。

(1) waseda

THE WASEDA BIOETHICS WEBSITE (Jan. 19, 2001)
"the peaceful 21st century" version. ... Here you're at the Gateway of "WASEDA
BIOETHICS WEBSITE"!! You're just the th visitor!! ...
kenko.human.waseda.ac.jp/rihito/biogate-j.html - 11k - Cached - Similar pages

(2) waseda homepage

THE WASEDA BIOETHICS WEBSITE (Jan. 19, 2001)
... Presented By Naoto Kawahara (Graduate School) Please send me your candid
opinion on this Homepage!! naox@human.waseda.ac.jp (Naoto Kawahara). ...
kenko.human.waseda.ac.jp/rihito/biogate-j.html - 11k - Cached - Similar pages

図-6 Googleの検索結果例

検索時には、図-5のSearcherがWebサーバから起動され検索が行われる。まず、入ってきた検索語がLexiconにより対応するwordIDに変換され、Barrels内の逆索引により、wordIDから対応するdocIDが検索される。検索結果のランク付けにおいては、検索語がWebページ中にどのような状態で現れているかという情報と後述のPageRankを組み合わせて算出される^{☆22}。また、複数の語が検索語として指定された場合は、指定された検索語の語間が近いものに対してランキングを上げている。

■主要検索機能

Googleが持つ特徴的な検索機能について説明する。

①アブストラクト表示

Googleで特徴的なのは、図-6に示すように同じサイトでも検索語を変えた場合、アブストラクトが変わるという点である。Googleでは、全WebページのデータをRepositoryにキャッシュしており、アブストラクトを検索語に応じて動的に生成した後、検索語に対応する部分を強調表示する。

②Webページのキャッシング

各Webページの情報をすべてキャッシュしており、オリジナルなページにアクセスできない場合でも、キャッシュされているWebページを参照できる。他の検索エンジンにはない機能であるが、キャッシュが著作権法上、許されるかどうかは微妙である^{☆23}。

③類似ページ検索 (GoogleScout)

類似ページ検索は、検索の結果得られたWebページと似たページを表示する機能でありGoogleScoutとも呼

^{☆21} 2001.2.19にhttp://www.webmaster.com/にGoogleが投稿したFAQに基づく。

^{☆22} 具体的にPageRankにどれだけの重みを置いているかは公開されていないが、PageRankの重みを大きくすることでScam Webページ(汚いやり方で検索エンジンの検索結果上位にくるように仕組んだWebページ)を検索結果から排除しているものと考えられる。

^{☆23} 米国の法律上も明確な判断はなされておらず、Googleでは、キャッシュに対してクレームがあった場合、手作業で該当ページに対するキャッシュを行わないように設定するという対策をとる。

ばれる。本機能は、他の多くの検索エンジンにも備わっている機能である。しかし、Googleでは、http://www.waseda.ac.jp/のようにWWWサーバ名以外を持たないWebページに対して本機能を選択するとドメイン名を考慮した類似ページが表示される。上記の例だと、大学(jpを第1、acを第2レベルドメインとして持つ)のWWWサーバのトップページが表示される。つまり、Googleではドメイン名も類似ページを求める上で利用していると考えられる。一方、http://www.waseda.ac.jp/index-j.htmlのようにhtmlファイル名をURLに持つページに対して本機能を選択すると、同類のキーワードが含まれるページが表示される^{☆24}。

④ランキング

本機能は、Googleの最大の特徴であり、PageRank法が用いられる。PageRankは、Google社の登録商標、かつ、特許出願中の技術である。以下で説明する。

■PageRank法²⁾ ~4)

PageRank法は、PageとBrinがスタンフォード大学在学中に提案した手法である。PageRank法では、あるページAからページBへのリンクをページAからページBへの支持投票とみなし、その票数からそのページの重要度を算出する。この考え方自体は、論文の引用率に基づいた評価や計量社会学において1950年代から存在する。

PageRank法を用いれば、検索結果の上位に有名なWebページ、すなわち他のサイトから沢山のリンクが張られているWebページを表示できる。たとえば、「文学部」で検索すると、第1位に慶應義塾大学、第2位に京都大学、第3位に早稲田大学が検索結果として表示されるが、これは、慶應義塾大学文学部のホームページが他の沢山のサイトからリンクが張られ有名であることを示す。つまり、検索エンジンを使うユーザから見ると、他の多くの人がよいと思ってリンクを張るようなWebページを効率よく探すことができ、検索に要する時間の短縮が可能となる。

これまでも、張られているリンク(Back Link)数を重要度として用いる検索エンジンは存在したが、PageRank法の特徴的な点は、リンク元であるページAの重要度によって、1票の重みを変える点にある。この方法を用いれば、それほど重要でないページCが無意味なページから多数のリンクを受けているような場合には、ページCの重要度が増してしまうことを避けることができる。たとえば、Scam Webページ(汚い手口の

^{☆24} 後述のDDM法が利用されている可能性もあるが、利用されているかどうかは不明。

Webページ)と一般的に呼ばれるWebページは、重要度を上げるために複数のダミーサイトからのリンクを張っている^{☆25}。これらのWebページは、リンク数のみをカウントして重要度を判定する一般のサーチエンジンでは、検索結果の上位に出してしまうが、PageRank法を用いるとこれらのダミーサイトの重みを低く設定できるため、ページCの重要度を低くすることができる。これが、GoogleがScam Webページに強いといわれる理由である。

PageRankは、次式を繰り返し適用することにより求められる³⁾。

$$R(p) = \frac{\epsilon}{n} + (1 - \epsilon) \cdot \sum_{(q,p) \in E} \frac{R(q)}{\text{outdegree}(q)} \quad (1)$$

$R(p)$ はページ p のPageRank、 $R(q)$ はページ q へのリンクを持つページ q のPageRankを示す。 n は対象とするグラフ G (Webページをノードとし、Webページ間のリンクをエッジとしたグラフ)のノード総数(Webページ数)、 $\text{outdegree}(q)$ はページ q からの外向きリンク数である。また、 $\sum_{p \in G} R(p) = 1$ であり、 ϵ は通常0.1~0.2の間に設定され、**dampening factor**と呼ばれる³⁾。

モデル的には、グラフ G のノード p_i から $(1 - \epsilon)$ の確率でノード p_i からのリンクをたどりノード p_j へ進むが、一方で ϵ の確率でリンクをたどることなくノード p_k へジャンプするという無限のランダムウォークと考える。つまり、「ユーザは $(1 - \epsilon)$ の確率で現在のWebページからのリンクをたどり、 ϵ の確率でまったく無関係なWebページへジャンプする」というモデル化を行っている。このモデル化により、どこからもリンクを持たないWebページのPageRankは ϵ/n となる。

PageRankの求め方は、文献4)に詳しいが、この問題は、上記のランダムウォークの推移確率行列の優固有ベクトル(**dominant eigenvector**)を求めることに等しい^{☆26}。

実際の計算においては、 $\text{outdegree}(v) = 0$ となるWebページ v をあらかじめ除いたグラフからPageRankを求める。これは、 $\text{outdegree}(v) = 0$ であるようなWebページ v は、他のどのページに対してもPageRank計算上影響を与えないからである。また、文献2)では7,500万のWebページ(リンク数3.22億)に対して式(1)を約52回繰り返し適用することにより、一般的なワークステーションを用いて約5時間でPageRankが収束することを確認している。また、文献3)でも約100回の繰り返しで十分であると報告されている。

☆25 たとえば、2001.11.1のWired News (<http://www.wired.com/>)「Did Smut Spammers Scam Google?」でも取り上げられたwww.jennifer-smith.comは、ダミーサイトへのリンクのみを持つ。

☆26 日本語での解説は、京都大学の馬場肇氏のWebページが詳しい(<http://www.kusastro.kyoto-u.ac.jp/~baba/wais/pagerank.html>)。

Googleが現在どのようなマシンでPageRankを計算しているかは不明であるが、Compaq Systems Research Centerが行った実験^{☆27}では、3.28億のWebページ(61億のリンク)に対して667MHz Compaq AlphaServer ES40(主記憶16GB)を用いてPageRankを計算したところ、グラフ全体をメモリ上に載せることにより約1日で計算を終えている。これは、推移確率行列が疎な行列であるため、メモリ上への格納方式を工夫することによって可能となっている。

■その他の技術

PageやBrinがスタンフォード大学時代に研究開発したその他の主要な2手法を紹介する。

DIPRE⁵⁾

Dual Iterative Pattern Relation Expansionの略であり、パターン認識を用いてWeb上から特定の情報を取り出そうとする試みである。文献5)では、本を例題として取り上げ実験を行っている。まず、「著者名、書名」のペア5組を用いてWebデータを検索する。この際、著者名と書名が近接して現れるWebページのみを抽出する。その後、得られたWebページから、「著者名」と「書名」がどのような形式で表現されているかを抽出する。たとえば、「<i>書名</i> by 著者名」のような形式を抽出する。次に抽出された複数の形式に合致するWebページを検索し、最終的に、さまざまな「著者名、書名」の組を得る手法である。

本手法については、現在Google社が進めているディレクトリ型検索サービスであるOpen Directory Project^{☆28}の各ディレクトリに合致するWebページを自動抽出するために利用ができないかどうか検討されている。

DDM⁶⁾

Dynamic Data Miningの略であり、データマイニングにおいてルールの探索を並列に行うことにより、Web情報等の膨大なデータを対象としたデータマイニングを高速化する手法である。

DDMの特徴は、解析対象であるアイテム同士の関係をWeightという実関数で定義すると共に、アイテムをノードとし、Weight値をエッジとしたハイパーグラフの形状から、Weight値が求められていなくても、グラフ形状からアイテム間に関連性があると推測することが可能になっている点である。このようにグラフ形状から関連性を判断する手法をHEP(Heavy Edge Property)と呼ぶ。

さらに、従来のデータマイニングシステムでは、アイテムの追加、削除を動的に行うことが困難であった

☆27 <http://www.research.compaq.com/src/mercator/research.html>

☆28 現在英語版のみがあり、人手により作成されている。

が、DDMでは、アイテムの追加、削除、さらにはWeight関数の修正までが動的に可能であり、Webデータの追加に併せて動的に再構築ができる利点を持つ。

2,400万のWebページを対象に、Webページ内の語句の関連性を調べたところ、関連性の強いものとして(フォトショップ、パワーマック)や(モルジブ、ザンビア、フィジー、グアドループ、エストニア、ルワンダ)等^{☆29}が得られている。

■フォルトトレラント機能・セキュリティ対策

Googleでは、(1)各クラスタがすべて同一のデータを保持し、(2)各クラスタ内でもデータを多重化、さらに(3)全マシンをモニタし故障時にはリスタートさせることでフォルトトレランス性を維持している。

セキュリティ対策では、OSにLinuxを採用しているが故に、ソースが公開されていない商用Unixに比較し、セキュリティレベルが高いという立場をとる。また、httpのアクセスのみを許し80番ポートのみを外部からアクセス可能としている。

■今後の技術的課題

今後の技術的課題として下記4項目を挙げている。

(1) システムイメージの整合性

5カ所のデータセンター、さらにはデータセンター内の複数クラスタが持つデータの整合性をどのようにして保つか。

(2) システム稼働状況のモニタ

膨大な数のPCの稼働状況をどのようにして効率よく取得するか。

(3) パフォーマンス・デバグging

負荷分散をどのように効率化するか。

(4) ビットエラー

10TBオーダーのデータを扱うため、ディスク、メモリのあらゆる場所でのビットエラーを避けることは確率上困難となっており、どのような対策をとるべきか。

Googleのビジネスモデル

2001年3月末時点でGoogleは約200名の社員を持ち、その内約100名が技術者である^{☆30}。Google社では収入源として、他の多くの検索エンジンが収入源として利用している広告の他に、Googleの検索エンジンの機能自体を別の検索エンジンのバックエンドとして売るサービスを展開している。2001年3月末時点で120以上のバックエンド(世界30カ国)を提供している^{☆31}。

^{☆29} 文献6)では英語表記である。

^{☆30} 日本人技術者は2001年3月末時点でスタンフォード大学の修士課程を修了した女性研究者1名のみである。

主な会社としては、Cisco, Washington Post, Virgin Net, Red Hat等がある。

日本向け戦略ではユーザ数が3,000万ともなっているi-mode向けの検索サービスを2001年2月に開始すると共に、日本語版のGoogle Tool Bar^{☆32}の提供、日本語広告の順で力を入れようとしている。

検索結果に現れる広告では、従来の検索エンジンがバナー広告を使うのに対して、Googleでは2行のテキスト広告を採用している。一般的にバナー広告は、ポップアップメニュー、アニメーション等を多用する。しかし、Googleでは、2行のテキスト広告を採用することで、ユーザが一般的に感じる「バナー広告を表示するが故に、検索結果の表示が遅い」といった問題を解決している。2行広告は、検索語と関連づけられており、特定の検索を行った場合のみ表示される。これにより、2行広告は一般のバナー広告に比較して4~5倍のクリック率を達成している。

おわりに

Googleの強みは、13.5億ページという世界最大のデータ量とPageRank法によるランキング技術であろう。今後もGoogle社が成長を続けられるかどうかを予測することは困難であるが、少なくとも技術的に他の検索エンジンとの差異を保つことができる限り、Googleの存在価値はある。さらに、巨大クラスタシステムは、今や世界中のどこを探しても存在しない規模になっており、世界最大のクラスタシステムを持つことと、世界最大のWebデータの蓄積から得られるさまざまなノウハウの蓄積は、次のビジネスへと展開していくためのエネルギーとなることは間違いない。

参考文献

- 1) Brin, S. and Page, L.: The Anatomy of Large-scale Hypertextual Web Search Engine, Proc. of the 7th Int. World Wide Web Conf., pp.107-117 (1998).
- 2) Page, L., Brin, S., Motwani, R. and Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web, Stanford Digital Library Technologies, Working Paper SIDL-WP-1999-0120 (1998).
- 3) Henzinger, M.: Link Analysis in Web Information Retrieval, IEEE Bull. of the Tech. Committee on Data Engineering, Vol.23, No.3, pp.3-8 (2000).
- 4) Haveliwala, T. H.: Efficient Computation of PageRank, Stanford Digital Library Technologies, Technical Rep. 1999-31 (1999).
- 5) Brin, S.: Extracting Patterns and Relations from the World Wide Web, Proc. of the Workshop on the Web and Databases (WebDB '98) (in conjunction with EDBT '98) (1998).
- 6) Brin, S. and Page, L.: Dynamic Data Mining: Exploring Large Rule Spaces by Sampling, <http://www-db.stanford.edu/~sergey/ddm.ps> (1998).

(平成13年6月14日受付)

^{☆31} Googleの多言語への対応は、こうしたバックエンドユーザとの連携、および各国のユーザによるボランティアにより行われている。

^{☆32} Internet ExplorerのツールバーにGoogleの検索機能を追加する仕組み。表示ページのPageRankの大きな値が棒グラフで表示される。