

QueReSeek: 検索履歴の逆引きによるコミュニティベースの Web ナビゲーション

QueReSeek: Community-Based Web Navigation by Search History Reverse Lookup

丹 英之^{*1}
Hideyuki TAN

大向 一輝^{*2*3}
Ikki OHMUKAI

武田 英明^{*2*3*4}
Hideaki TAKEDA

^{*1} 株式会社アルファシステムズ
Alpha Systems Inc.

^{*2} 国立情報学研究所
National Institute of Informatics

^{*3} 総合研究大学院大学
The Graduate University for Advanced Studies

^{*4} 東京大学
The University of Tokyo

In this paper, we propose the sharing system of the search query that is issued to the search engine by the community. This system constructs the Query-URL index from community's search history, and, presents related search queries from the URL of the Web page that is in browsing. The query of the community's issued to the search engine is the keyword of the knowledge which the community has interest. Consequently, this system can offer the link to the other Web contents with community's interest keyword connection. We think that it is possible to promote the Web browsing which is sharing knowledge in the community.

1. はじめに

インターネット利用形態の一つである Web 閲覧は、検索エンジンサイトの発達と共に、インターネットから情報・知識を引き出すための入口として、様々な背景を持つユーザに利用されるようになった。しかし、情報爆発時代、情報大航海時代と表現される様に、Web コンテンツの数は急激に増大し、元々高いとはいえない S/N 比は一段と低下する傾向にあり、コンテンツの信頼性を伴う質の問題にぶつかる。そして、情報リテラシーの高いユーザでない限り、情報の波に飲み込まれてしまい、ユーザが求める情報・知識が掲載された Web コンテンツへ辿り着くには時間を要し、更には目的とする Web コンテンツへ到達できず検索行動を諦めてしまうユーザが現れることになる。これは、大手検索エンジンサイトの利用方法に関する攻略本が数多く出版されていることから伺え、検索エンジンサイトの提供するサービスへの興味もさることながら、検索エンジンを上手に使いこなしたいというユーザが数多く存在していると考えられる。

大手検索エンジンサイトでは、ユーザの入力したキーワード(検索クエリ)と連動した広告を提示している。これはターゲット・マーケティングの一つであり、検索連動型広告と呼ばれ、入力された検索クエリからユーザの意識の範囲にありそうな広告情報を選択して提示することで実現される。この広告手法は、検索クエリの内容と関連性を持たない広告提示と比較して、広告表示に付与されたリンクのクリック率が非常に高く、ユーザの購買行動にまで誘導できる場合が多く、商業的にも成功している。この点を考慮すると、先に述べた検索エンジンを使いこなせていないユーザが多いのにも拘らず、その使いこなせていないユーザが発行する検索クエリの結果である Web コンテンツ集合の一つとして提示される広告情報は、ユーザへの影響度が大きいと言えることになる。もちろん、広告情報が検索結果である Web コンテンツ集合の上位に表示されることで、ユーザに対して視覚的に影響を与えていることも考えられる。また、検索行動は、知りたいことに対してわざわざ検索クエリを想定し入力する、という心理的な障壁を超える何らかの動機による能動的な行動であり、心理的なコストが発生している。すると、検索結果が望んでいる

もので無かった場合、この心理的なコストに対する見返りを得るのに失敗していることとなる。ここで、得られた結果の判断に、検索行動を自己肯定してしまう作用が働き、有効な Web コンテンツであると感じてしまうことが多い、ということも考えられる。いずれにせよ、検索結果である Web コンテンツ集合は、ユーザが求めるものに関して編み出した検索クエリによって、ユーザに対して絞り込まれたものとなるので、ユーザに影響を与えた Web コンテンツと検索クエリの間には強い結び付きが存在すると言える。

今後増加するであろう情報の波に飲み込まれてしまうユーザを救済するには、ユーザの意図する Web コンテンツを何らかの方法で探し出し、それを提示すれば良い。しかし、ユーザの意図を把握するのは、非常に困難である。ここで、先に述べた Web コンテンツと検索クエリの関係を利用する。すると、閲覧中の Web コンテンツが検索エンジンでの検索結果となる検索クエリを探すことで、ユーザの意図する Web コンテンツの範囲を絞ることができる。そして、この検索クエリをユーザへ提示することで、検索クエリを介した内容的に近傍にある Web コンテンツへのリンク、つまりユーザの求める Web コンテンツへの近道を提供できると考える。

本稿では、Web 閲覧に伴ったコミュニティ内における情報・知識共有の補助を目的として、検索エンジンへ投入された検索クエリをコミュニティ内にて共有し、閲覧中の Web コンテンツに「コミュニティの関心キーワード」を逆引きして提示することで、影響度が大きくユーザに効果的な Web コンテンツへと閲覧行動を誘導する手法、QueReSeek を提案し、この提案手法を実装したシステムについて述べる。

2. 検索時における情報の提示

本研究では、ある共通の目的を持つ検索エンジンユーザの集団をコミュニティと呼ぶ。ここで、ユーザが検索エンジンへ発行する検索クエリは、ユーザの持つ背景知識を元として、望む Web コンテンツに関連するであるという思慮を含んだ文字列であると捉えてみる。すると、このコミュニティを形成するユーザが発行する検索クエリは、コミュニティを形づくる共通の目的と潜在的に関連を持つであろうと想定できる。つまり、コミュニティ内から発生した検索クエリは、潜在的にコミュニティの目的のためであり、コミュニティ内にて関心のある Web コンテンツと結び付く

連絡先: 丹英之, 株式会社アルファシステムズ, 川崎市中原区
上小田中 6-6-1, tel:044-738-4125, mailto:tanh@alpha.co.jp

ことになる。検索エンジンの返す結果は、複数の Web コンテンツへのリンクであるので、関心キーワードは関心のある Web コンテンツ集合、複数の Web コンテンツへと繋がるのである。よって逆を辿ると、コミュニティが関心を持つ Web コンテンツは、検索クエリであるコミュニティの関心キーワードへ繋がる。これは、検索クエリがその Web コンテンツを一言で表していること、つまり、Web コンテンツが一言に抽象化されたということになる。この繋がりを用いることで、コミュニティに属しているユーザが閲覧中の Web コンテンツに対し、コミュニティ内から発生した関心キーワードを提示できるはずである。すると、関心キーワードによって繋がる他の Web コンテンツへと導く新しいリンクをユーザへ提供できる。よって、コミュニティ内にて情報・知識を共有するための Web 閲覧を促進できると考えられる。提案する手法では Web コンテンツの URL から検索クエリを求めるので、「逆引き検索エンジン」とも表現できる。

コミュニティの関心キーワードである検索クエリの共有によって、検索を得意としないユーザは、検索クエリ発行のためのヒントとなる情報を入手できる。一方、検索を得意とするユーザには、不得意なユーザの発行する検索クエリによって、求める Web コンテンツの内容とは外れた検索クエリが提示されてしまうことも予想される。しかし、ユーザの閲覧している Web コンテンツ中には提示された検索クエリが存在しており、コミュニティを形成する共通の目的に対する知識分野とは全く関連性を持たないと想像するのは難しく、着眼点を変えた Web コンテンツの見方を示唆できる。つまり、本研究の手法によって、ユーザの求める情報・知識が掲載されている Web コンテンツへの閲覧行動の誘導、そして、閲覧中の Web コンテンツに関連する、気づきに似た新しい着想を与えることが可能となる。

以下に、検索のクエリ共有として、検索クエリ入力時においてクエリ文字列を提示する試み、及び、閲覧中の Web コンテンツに関連情報を提示する試みについて、関連する先行研究を紹介する。そして、本研究の位置付けについて述べる。

2.1 検索クエリ入力時におけるクエリ文字列の提示

ユーザへ過去に利用された検索クエリを提示するサービスがある[Google 2004][NTT レゾナント 2005]。これらは、ユーザの検索行動の真っ只中である検索クエリ入力中のコンテキストにおいて過去に入力されたクエリ群や整備された辞書を元に、あるアルゴリズムによって順序付けされたインデックスから、インクリメンタル検索を行い、結果を順に提示する。Web ブラウザ上で実現された、予測変換入力メソッドの拡張による検索クエリ入力の補助である。一般的には人気のあるテーマに関する Web コンテンツを検索する際において、ユーザビリティが向上する。

本研究の手法は、Web コンテンツ閲覧時においても常に検索クエリを提示するので、検索クエリをユーザへ提示するタイミングの点において異なる。

2.2 閲覧中の Web コンテンツに関連する情報の提示

閲覧中の Web コンテンツの URL を元に、関連する情報を提示する例も幾つかある。例えば[Google 2000]では、Web ブラウザの機能拡張プラグインとして実装され、閲覧中の Web コンテンツに関する検索エンジン内インデックスでの評価度が表示される。閲覧中の Web コンテンツの周辺空間を、その Web ページからのリンクの近傍 Web ページ群で構成される構造的周辺空間、及び、検索エンジンでの類似ページ検索による内容的周辺空間から構成される空間とし、これらの情報から Web 空間上に交通標識をメタファーとする Web 標識なる誘導標識を生成し、これを周辺情報として提示することで、ユーザの求める Web コ

ンテンツへ Web 閲覧を誘導する試みもある[池田 2003]。また、ソーシャルブックマーク(SBM)サービスを Web コンテンツに関する提示情報の源として利用するアプリケーション[Glucose 2007]も存在する。SBM サービスは、不特定多数のユーザとブックマークを公開・共有するサービスであり、Web コンテンツ登録の際、ユーザはそのコンテンツに対して自由記述のタグと呼ばれる分類用の語句を付与することができる。このアプリケーションでは閲覧中の Web コンテンツに関して、被ブックマーク数、及び、付与されたタグを提示する。この提示されるタグは、SBM サービスにて提供されるタグ付与 Web コンテンツ一覧ページへとリンクされているので、SBM を媒介とした Web 閲覧の誘導と言える。SBM では、付与されたタグであるメタデータとユーザの民主的な評価により Web コンテンツが非階層構造で整理される。この分類法はフォークソノミーと呼ばれ、自発的なユーザ達によって選別された情報が集積されているので、ユーザにとって効果的な Web 閲覧の誘導が期待できる。

本研究の手法は、コミュニティから投入された検索クエリに関連情報としてユーザへ提示するので、扱う情報の源が異なる。

2.3 QueReSeek

ここでは QueReSeek の概要について述べる。まず、検索クエリを共有して再利用する点であるが、本研究の手法では、Web コンテンツ閲覧中のコンテキストにおいて、可能であれば常にユーザへ検索クエリを提示する。これは、検索エンジンの結果のページから Web コンテンツの表示状態へページが遷移した後、Web コンテンツを確認している最中でも関連する検索クエリが視野に入ることであり、更にユーザの意図している内容の Web コンテンツへと繋がる可能性がある検索クエリを、ユーザ自身で発見することができるようになる。また、コミュニティ内にて初めて利用される検索クエリを発行したユーザは、検索エンジンを介してコミュニティに新しい概念を導入したことになる。このユーザを先人と表現すると、広大な Web ネットワークを探索した先人の道筋を、そのコミュニティに属している後人が踏襲できることを意味しており、コミュニティ内にて情報・知識共有を伴った Web 閲覧を促進できることになる。そして、明確な目的を持たない Web 閲覧時においても閲覧中の Web コンテンツに関連する検索クエリが提示されるので、そのユーザが属しているコミュニティでは、どのような視点から閲覧中の Web コンテンツを捉えているのかについて発見する手掛りを提供できる。

この Web 閲覧中のコンテキストにおける関連検索クエリの提示は、SBM からのタグの提示と類似している。SBM のタグは自由記述であるため、付与された Web コンテンツとは異なる内容として表現されている可能性もある。しかし、この非常に大きな自由度のためコンテンツに対する評価や感想を表現する機能を持つタグが付与される場合もある。また、ユーザは Web コンテンツ確認後に SBM への登録を判断して行動するため、SBM にはユーザに対する影響度の大きい Web コンテンツが集積されていると言える。一方、検索クエリは、ユーザが求めている Web コンテンツについての思慮を含んだ文字列であり、入力と共に検索エンジンへ投入される。そして、検索エンジンサイトの採用しているアルゴリズムによって、Web コンテンツと結び付けられている。言い換えると、ロボットによって回収された Web コンテンツを対象とした全文検索の結果集合に含まれる文字列であり、Web コンテンツの破片である。つまり、SBM のタグと検索クエリでは、Web コンテンツに対し発生する順番が逆である。本研究で着目した検索クエリは、今後増加していく新しい Web コンテンツにも自動的に対応できる。

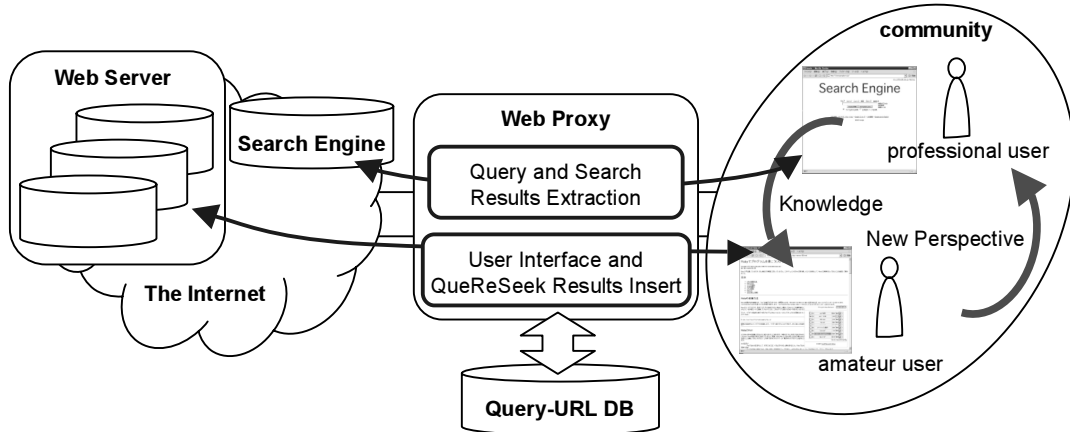


図1 本研究で提案する QueReSeek を実装したシステムの構成と利用イメージ

本稿の後半では、この検索クエリをコミュニティ内にて共有する QueReSeek の実装、及びその実用性の一評価について述べる。

3. QueReSeek の実装

本研究で提案する QueReSeek を実装したシステムの構成と利用イメージを、図 1 に示す。QueReSeek を実現するには、大きく分けて二つの機能を持つモジュールが必要になる。一つは、コミュニティから検索エンジンへ投入される検索クエリを回収するモジュール(図中、Query and Search Results Extraction)、もう一つはコミュニティに属しているユーザへ閲覧中の Web コンテンツに関連する検索クエリを提示するモジュール(図中、User Interface and QueReSeek Results Insert)である。インターネットでの Web 閲覧は、検索エンジンも含め Web コンテンツ提供側である Web サーバと、その Web コンテンツをユーザに提示する Web ブラウザの二つのアプリケーション間通信によって実現される。今回は、これら二つのアプリケーション間に位置できる Web プロキシのネットワークボロジに着目し、Web プロキシへ二つの機能モジュールを実装した。これにより、既に普及している Web サーバ、Web ブラウザに手を入れることなく、また、ユーザに新たなアプリケーションを配布することなく、提案手法を実現した。検索クエリ回収モジュールは、コミュニティの検索エンジンサイト(Google, Yahoo, Excite, msn, goo, Infoseek)の利用を検出し、投入された検索クエリを抽出する。この検索クエリについて、Google, Yahoo にて再検索を行い Web コンテンツの URL, ランキング, 推定総ヒット数を取得し、“Query-URL DB”へ格納する。検索クエリ提示モジュールは、QueReSeek の結果を提示するためのインタフェースを生成する JavaScript を Web コンテンツへ挿入する。この際、“Query-URL DB”を参照し、閲覧中の Web コンテンツの URL からコミュニティが投入した検索クエリを検索し、ユーザへ提示する。これらの機能により、コミュニ

ティ内にて検索クエリを共有し、前章で述べた効果を発揮する。

Web コンテンツ閲覧中、関連する検索クエリが存在する場合の Web ブラウザは、図 2 に示す画面をとり、Web コンテンツ表示領域の右下にサブウィンドウとしてインタフェースが現れる。サブウィンドウ内は、図 3 に示すようにテーブル形式で検索クエリが表示される。左カラムより、検索エンジンを示すアイコン、検索結果での順位、検索クエリ文字列、推定総ヒット数、他検索エンジンへのリンクアイコン、表示される検索クエリ文字列が相応しくない場合のプーイングボタン、コミュニティ内で検索クエリとして利用された回数を示す。検索クエリは検索エンジンでのランキング、ヒット数、コミュニティで投入した回数の順でソートし、上位 10 セットを表示した。

4. 評価

ここでは、QueReSeek の有効性を検証するにあたり、実装したシステムについて行った簡単な評価について述べる。

実装したシステムの起動初期には、図 1 の“Query-URL DB”は空である。もちろん、ユーザの継続的な利用によってインデックスは増加するので、検索クエリを提示できる可能性は高くなる。しかし、既にコミュニティが Web プロキシを利用している場合、そのログを事前に解析することによって、システムの起動初期段階においても“Query-URL DB”を用意できる。本稿ではこの方法を用い、実用性を計る指標の一つとして、閲覧中の Web コンテンツに対し、Web プロキシのログから回収した検索クエリをどの程度提示できるのかについて評価した。

4.1 実験の準備

Web プロキシのログ解析によって得られた検索エンジンの利用状況として、検索エンジン利用回数と割合、及び、抽出したユニーク検索クエリ数を表 1 にまとめた。解析対象とした Web プロキシのログは、筆者の属する組織の内、約 1,500 ユーザが 1 ヶ



図2 検索クエリ提示時の Web ブラウザ画面

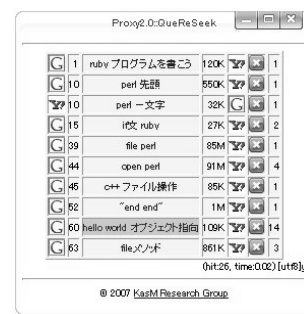


図3 検索クエリ提示サブウィンドウ

表 1 Web プロキシログ解析による検索エンジン利用状況

検索エンジン利用回数と割合		
Google	286,203	87.2%
Yahoo!	30,892	9.4%
Excite	4,076	1.2%
msn	3,588	1.1%
goo	2,620	0.8%
Infoseek	1,057	0.3%
合計	328,436	
ユニーク検索クエリ数 84,870		

表 2 構築した“Query-URL DB”の詳細

登録された検索クエリ数	84,135
延べURL数	11,366,183
ユニークURL数	8,106,381

月利用したログである。この解析により、約 85,000 の検索クエリを抽出できた。この抽出した検索クエリを用いて構築した“Query-URL DB”の詳細として、登録できた検索クエリ数、及び、延べ URL 数、ユニーク URL 数を表 2 にまとめた。抽出した検索クエリの再検索で、推定総ヒット数が 0 件以上のものを“Query-URL DB”へ登録したので、735 セットの検索クエリが本手法において意味をなさないことになる。これらの検索クエリを観察してみると、ログからの抽出時において文字コードの判定に失敗している文字列や、文章のような長い文字列である場合が多かった。ログから抽出した検索クエリでの再検索により、約 84,000 の検索クエリと約 800 万の URL が登録されている“Query-URL DB”を構築できた。

4.2 実験条件と得られた結果

先で述べた“Query-URL DB”と二つの機能モジュールを実装した Web プロキシを用意し、筆者が所属する部署内にて実験協力者を募り、システムの提供する機能について簡単な説明を行った後、休憩時間も含めた業務時間中の Web 閲覧を通常通り行ってもらった。計測期間は 5 日間で、ユーザ数は延べ 20 人であった。得られた結果として、閲覧された Web コンテンツ数、検索クエリを提示できた Web コンテンツ数、そして、計測開始から追加された検索クエリ数、及び、提示した検索クエリに付与したリンクを辿った回数を表 2 にまとめた。提示できた Web コンテンツは、閲覧されたコンテンツに対し、約 33%の割合で関連する検索クエリを提示できたことになる。ニュースや Weblog など更新頻度の高い Web コンテンツでは、検索エンジンのクロールとインデキシングにある程度の期間が必要である。実験協力者はこれら更新頻度の高い Web コンテンツを閲覧する機会が多いと想定していたが、検索クエリを提示できた Web コンテンツが 3 割以上であったのは、“Query-URL DB”にある約 800 万のユニーク URL が、閲覧された約 6,000 の Web コンテンツに対して十分な量であったとも考えられる。

また、提示した検索クエリへ付与したリンクを辿った回数は、49 回であった。これは、閲覧された Web コンテンツへ検索クエリを提示できたうちの約 2%ということになる。リンクを辿った検索クエリの約 15%が“Slashdot”, “kizasi”, “CodeZine”など Web サイト名であり、何らかの情報を求めるための検索ではなく、行き先が決まっている検索のための“ナビゲーションクエリ”であった。リンクを辿った検索クエリの約 3 割が、実験開始後に追加されたものであった。計測が短期間であったこと、及び、実験協力者から関連情報提示サブウィンドウが Web コンテンツを隠すので閲覧を妨げる、という意見があり、ユーザインタフェースの再検討を要するため、本稿では検索クエリを提示できた Web コンテンツの割合だけの議論に留める。

表 3 QueReSeek で検索クエリを提示できた割合

閲覧されたWebコンテンツ数	6,047
検索クエリを提示できたWebコンテンツ数	2,019 33.4%
追加された検索クエリの増加数	593
提示した検索クエリのリンクを辿った回数	49

5. まとめと今後の課題

本研究では、検索エンジンへ投入された検索クエリをコミュニティ内にて共有し、ユーザの閲覧する Web コンテンツに関連情報として検索クエリを提示する「逆引き検索エンジン」の QueReSeek を提案した。そして、これを実現するシステムを実装し、その実用性について一評価を行った。

Web ナビゲーションという観点からは十分な評価を行うことができなかった。これには、Web コンテンツに関連する検索クエリを複数発見した場合においての、提示すべき検索クエリの選出方法、そして、検索クエリを提示するユーザインタフェース等の課題も包含しているので、これから先は検索クエリの提示の仕方について再検討を行う。

提案する手法は、ユーザに対し閲覧中の Web コンテンツに関連するまだ存在を知らない未知の Web コンテンツへのリンクを提供できる。特に Web ネットワークは日々成長しており、一つの検索エンジンだけでは全てを覆いきれていない現状では、特定の検索エンジンしか利用しないユーザにとって、複数の検索エンジンの結果を集約できるシステムでもあり、新しい Web コンテンツへの出会いの機会を増やすことが可能である。

今回は、Web プロキシのログから抽出した検索クエリ集合を用いたが、例えばコミュニティの共通目的に関係する知識分野の用語辞典などから“Query-URL DB”を作成し、提示する検索クエリに制約を設けることで、Web 閲覧行動へ e-Learning 的な要素を追加することも考えられるので、この方向でも検討したい。

本稿の実装では、Web プロキシにて提案手法を実現し、Web プロキシのユーザ集合を検索クエリの共有関係を持つコミュニティとした。提案手法を実現するにあたり、引数に URL 文字列をとり、戻り値に検索クエリセットを返す機能を Web サービス API に、そして、ユーザへの提示インタフェースの機能を先行研究で挙げた単体のクライアントアプリケーション、もしくは、Web ブラウザの機能拡張モジュールへと分割すれば、多数のユーザへ本研究のサービスを提供できることになる。この場合、コミュニティの定義が異なることになるが、この状況下において提示されるであろう検索クエリは、フォークソノミーのタグにも匹敵する Web コンテンツの分類機能を持つことになるかもしれない。

今後は、提案手法について有効な評価・検証を進めていきたい。

参考文献

- [Google 2004] Google Suggest Web Page, <http://www.google.com/webhp?complete=1>, 2004.
- [NTT レゾナント 2005] goo サジェスト β with ATOK Web Page, <http://suggest.search.goo.ne.jp/suggest/>, 2005.
- [Google 2000] Google Toolbar Web Page, <http://toolbar.google.com>, 2000.
- [池田 2003] 池田新平, 是津耕司, 小山聡, 田中克己: Web コンテンツの周辺情報提示によるナビゲーション支援, 第 14 回データ工学ワークショップ(DEWS2003)論文集, 電子情報通信学会データ工学研究専門委員会, 2003.
- [Glucose 2007] Glucose nano Web Page, http://dev.glucose.jp/wiki/index.php/Glucose_nano, 2007.