

Web の探索行動と情報評価過程の分析

An Analysis of Information Seeking and Evaluating Processes on the Web

¹⁾ 名古屋柳城短期大学図書館
Nagoya Ryujo College Library

²⁾ 名古屋大学附属図書館研究開発室
Nagoya University Library Studies

種市 淳子¹⁾ 逸村 裕²⁾
TANEICHI, Junko ITSUMURA, Hiroshi

Abstract

In this study, we discussed information seeking behavior on the Web. First, the current Web-searching studies are reviewed from the perspective of: (1) Web-searching characteristics; (2) the process model for how users evaluate Web resources. Secondly, we investigated information seeking processes using the Web search engine and online public access catalogue (OPAC) system by undergraduate students, through an experiment and its protocol analysis. The results indicate that: (1) Web-searching patterns are repeated with regularity, between the process levels, such as key words, results, pages, and connected pages with links; (2) More experienced users of search engines tend to use some typical Web-searching patterns when using the OPAC system; (3) Users tend to evaluate search results of Web resources based on its look and information design than information quality and authority.

I. はじめに

World Wide Web (WWW) が、情報の生産と利用にもたらした影響は計り知れないものがある。その成長と共に、質の低い情報源という初期の評価を覆し、学術論文にも数多く引用されるように¹⁾、今日では情報源としての Web の有用性は高まっている。Web 上の一般的な情報探索手段としてサーチエンジンが普及し、Web に特徴的なブラウジング行動が研究活動においても重要な役割を果たしていることが明らかにされる²⁾につれて、ブラックボックスの多い Web の探索プロセスを

記述できるモデルの構築や、そのための調査が重要視されるようになってきた。

近年では、Web の探索行動研究は確実に増えつつあるが³⁾、Web の探索行動の特質や構造はまだ解明されていない点が多い。Jansen らは、Web の従来の情報検索システムと異なる特殊な検索環境が、研究を制限する理由になったと述べている⁴⁾。

本論の目的は、大学生の Web 探索行動と情報評価のプロセスを分析することにある。まず、先行研究のレビューを行い、Web と従来の情報検索システムを比較し、Web の探索行動の特徴につ

いて述べる。また、情報源の評価行動に注目し、Web の信頼性評価に影響を与える要素と評価モデルの検討を行う。

次に、先行研究の知見を検証するために、大学生のサーチエンジンと OPAC の検索過程についてプロトコル分析法により調査した結果について述べる。

Ⅱ. 先行研究の検討

1. Web と従来の情報検索システムの比較

Jansen & Pooch (2001) は、Web の探索に関する研究の動向を広範に調査し、そこに見られる共通の傾向を見出して、Web 探索研究における方法論の枠組みを提案している。Jansen らは、Web の探索に関する研究と、従来の情報検索システム (IR) 及びオンライン・パブリック・アクセス・カタログ (OPAC) に関する研究を比較し、それぞれの検索システムに見られる典型的な様相と Web の特殊性を論じている。

Jansen & Pooch は、Web 探索、従来の IR、OPAC という 3 種の異なる情報検索システムにおける主要な先行研究として、それぞれ 3 種、合計 9 種の研究を取り上げて比較している (Web の

研究: Hoelscher (1998)⁵⁾; Jansen ら (2000)⁶⁾; Silverstein ら (1999)⁷⁾、IR の研究: Hsieh-Yee (1993)⁸⁾; Koenemann & Belkin (1996)⁹⁾; Siegfried ら (1993)¹⁰⁾、OPAC の研究: Millsap & Ferl (1993)¹¹⁾; Peters (1989)¹²⁾; Wallace (1993)¹³⁾。

(1) Web の探索行動における比較

Web の 3 種の研究は、すべてサーチエンジンのログ分析によるものである。Hoelscher (1998) の Fireball 研究⁵⁾ は、ドイツのサーチエンジン Fireball¹⁴⁾ から収集した約 1,600 万件のデータによるものであり、Jansen ら (2000) の Excite 研究⁶⁾ は、サーチエンジン Excite¹⁵⁾ の利用者による約 5 万件のデータを基にしている。Silverstein ら (1999) の Alta Vista 研究⁷⁾ は、サーチエンジン Alta Vista¹⁶⁾ から収集した約 10 億件のデータに基づいた調査であり、データ件数や収録期間の長さにおいて、現在も最大規模とされる研究である。

Jansen らは、これらの研究で示されたデータを、共通の指標を用いて比較した。(しかし、表中の値は定義し直されたものであり、いくつかの項目では厳密な比較はできない)

Web 探索における典型的な行動として、1) 利用者は 2 語の検索語を使用し、複雑な論理演算

表 1 Web 探索研究における比較 (Jansen ら, 2001)

項 目	Fireball 研究	Excite 研究	Alta Vista 研究
データ収集期間	31日 1998/07/01~1998/07/31	1日の一定時間 1997/03/10	43日 1998/08/02~1998/09/13
検索システム名	Fireball Search Engine	Excite Search Engine	Alta Vista Search Engine
収集されたクエリーデータ数	16,252,902	54,573	993,208,159
セッション1回あたりのクエリー実行数	報告なし	平均: 1.6, 標準偏差: 0.69 1: 67% (36,564) 2: 19% (10,391) 3: 7% (3,820) 4: 3% (1,637) <4: 4% (2,183)	平均: 2.02, 標準偏差: 123.4* 1: 77.6% (221,527,914) 2: 13.5% (38,539,006) 3: 4.4% (12,560,861) <3: 4.5% (12,846,335) *標準偏差の大きな値は, softbots による影響と見られる
1クエリーあたりの検索語数	平均: 1.66, 標準偏差: 0.70 0: 報告なし 1: 54.59% (8,873,001) 2: 30.80% (5,005,653) 3: 10.36% (1,683,129) <3: 4% (691,119)	平均: 2.21, 標準偏差: 1.05 0: 5.02% (2,584) 1: 30.81% (15,854) 2: 32.46% (16,191) 3: 17.96% (9,242) <3: 15% (8,186)	平均: 2.35, 標準偏差: 1.74 0: 20.6% (204,600,881) 1: 25.8% (256,247,705) 2: 26.0% (258,243,121) 3: 15.0% (148,981,224) <3: 12.6% (125,144,228)
ブール演算の使用	2.55% (414,461) *提供データ中の最大値	8.54% (4,661)	報告なし
クエリー設定時のエラー	報告なし	10% (5,457)	報告なし
近接演算の使用	25.3% (4,111,843)	9% (4,776)	20.4% (202,614,464)* *ブール演算子を含む
セッション1回あたりに見る文書数	10件以下: 59.51% (9,621,347) <10件: 40.47% (6,545,887)	10件以下: 58% (31,652) <10件: 42% (14,735)	10件以下: 85.2% <10件: 14.8% *数値は報告なし

は使用しない、2) 利用者が結果から見る文書は10件以下であると結論している。すなわち Web の探索行動は単純化されたパターンをもつことを示しているといえる。

(2) Web と IR 及び OPAC における探索行動の比較

従来の IR 及び OPAC に関する研究は、Web 探索研究の比較と同様に、共通の指標をもとに集計され（初心者とエキスパートの両方のデータがある場合は、初心者のデータを使用している）、1つの表に統合された（表2）。

3種の異なる検索システムに共通する傾向は、検索オプションの使用率、すなわち検索オプションを含むクエリー件数の割合が8%程度と低いことである。また、Web と従来の IR では結果を見る件数において（10件程度）、Web と OPAC では検索式に用いる検索語数において（1～2語）、それぞれ類似する傾向が示されている。

しかし、その他の指標が示す傾向は異なる。従来の IR は、検索式が長く、ブール演算の使用率は37%と高い割合を示す。従来の IR とは対照的に、すべての項目で最も単純なパターンを示すのは Web であり、検索式のエラーも10%と最も低い。

従来の情報検索システムの検索における研究では、効率よく検索を進めるうえでの計画や手法として、検索戦略（search strategy）や検索戦術（search tactics）という語が使われた。Bates（1979）は、検索戦略を「検索全体へのアプローチの仕方や計画」、検索戦術を「検索効率や検索効果をあげるための手段」と定義し、合計46種類もの検索戦術をあげている^{17,18)}。

従来の情報検索システムは、探索状態を常に保

持しているため、結果を積み上げて構造化することが可能であった。検索戦略や検索戦術は、そのような検索環境の中で有効性を発揮するものである。一方、Web の探索では、探索状態が保持されず、結果を積み上げてフィードバックに基づく論理的な検索を行うことはできない。表2が示す Web 探索パターンの単純性には、Web のシステム上の特質が関わっていると考えられる。Web は、従来の情報検索システムに用いられた複雑な検索戦略や検索戦術を要しない（あるいは意識しない）システムであることが示されている。

2. サーチエンジンの探索パターン

サーチエンジンの調査は、Web 探索研究における主要な課題として、近年増加している。Pew Internet & American Life Project（2004）の全米1,399人に対する電話調査によれば、サーチエンジン利用者の35%は1日1回サーチエンジンを使い、その内の3分の2は1日数回使う¹⁹⁾。また、サーチエンジン利用者の87%は探したい情報をほぼ見つけられると答え、68%はサーチエンジンを公平で偏りのない情報源と答えている¹⁹⁾。

Spink ら（2002）は、1997年、1999年、2001年に、サーチエンジン Excite¹⁵⁾ のログ分析による調査を行っている²⁰⁾。それによると、1クエリーあたりで、検索語数の平均は2.4～2.6語、結果から見た特定のページ閲覧数の平均は1.6～1.7件である。また、結果から見る特定のページ閲覧数が1件の割合は、28.6%（1997年）、42.7%（1999年）、50.5%（2001年）と4年で倍増している。つまり、サーチエンジン利用者の2人に1人は検索結果を1件しか見ないことになる。

表2 異なる情報検索システム研究の比較（Jansen ら，2001）

項 目	Web	従来のIR	OPAC
検索1回の長さ (検索1回あたりのクエリー設定数)	1-2	7-16	2-5
クエリーの長さ (1クエリーあたりの検索語数)	2	6-9	1-2
検索1回あたりに見る適合文書数	10件以下	10件程度	50件以下
高度な検索オプションの使用 (検索オプションを含むクエリーの数)	9%	9%	8%
ブール演算の使用	8%	37%	1%
エラーになる確率 (クエリー設定時のエラー)	10%	17%	7-19%

このような傾向（検索結果から1件以上を見る意欲が減少しているような探索パターン）は、彼らがその後に行なった別の調査で一層明らかになっている。Jansen & Spink（2003）は、サーチエンジン FAST²¹⁾ のログ分析による調査を行い、利用者が結果から見た Web ページとその閲覧時間を調べている²²⁾。それによると、1) 利用者の54%は単に1つのクエリーを用い、結果は1件しか見ない、2) 特定のページを見る時間は、75%以上が15分以下、40%が3分以下、14%は30秒以下であることを示している。また、利用者が結果から見たページの適合性（relevance）を分析したところ、52%の確率で適合性があったと報告している。

Jansen らは、この実際に見たページの適合率（52%）と結果を1件しか見ない割合（54%）には関連があると述べている。言い換えれば、サーチエンジンのランキングシステムにより2件の内1件という高い確率で適合情報を得られることが、結果を見て判断するまでにかける時間を短くする要因の一つとなっているのではないかと考えられる。

僅かな時間の中で、情報はどのように評価されるのか、その判断基準となるものは何か。次節では、Web の評価に関する問題を検討する。

3. Web の信頼性評価の指標

Web Credibility Project²³⁾ では、サイト設計技術の発展と Web の信頼性に関する研究の促進を目的として、いくつかの定量的調査が行われ、その研究成果をもとに、10項目からなる Web の信頼性のためのガイドライン（2002）を提示している²⁴⁾。

- 1) 情報の正確性を確認しやすくする（引用、参照、典拠の明示等）
- 2) 組織等の実体がよく見えるようにする（住所、会社の写真等）
- 3) 組織やサービスの専門性を強調する（提供者の権威と証明、著名団体加入の明示等）
- 4) 実際に運営する人の存在を示す（構成員の明示と説明文書、イメージ等）
- 5) 連絡をとりやすくする（住所、電話番号、電子メールアドレスの明示）
- 6) プロフェッショナルな印象のデザイン設計（レイアウト、統一感、目的に合うデザイ

ン等）

- 7) 利用者にとって便利で使いやすいようにする
- 8) 更新と情報の確認は頻繁に行い、そのことを明示する
- 9) 広告など、宣伝的なコンテンツの掲載をできるだけ避ける
- 10) エラーの回避（誤字、リンク切れの確認等サイトの維持管理）

人々がどのような要素をもとに Web サイトの信頼性を評価しているかを分析した2つの調査では、一般の消費者と特定領域の専門家が Web サイトの評価に用いる基準は異なることを明らかにしている。

Fogg ら（2002）の一般消費者に対する調査は、Web 上の質問紙による2,864人の回答をもとに行われた。回答者は、10の主題カテゴリーの内1つから任意に割り当てられた2つの Web サイトを評価し、コメント（信頼性を評価した結果良い、悪い等）を記述した。その結果、コメント全体の46.1%でデザイン設計が評価されるなど、Web サイトの表面的要素をもとに判断する傾向が見られた²⁵⁾。

Stanford ら（2002）の別調査では、専門家が評価に用いる指標は、一般の消費者と異なることを示した²⁶⁾。医者やフィナンシャルプランナーなど健康及び金融の専門家15人が、10の健康又は金融分野のサイト評価を行ったところ、重視したのは、表面的要素ではなく、情報源の質に関わるものであった。また、健康専門家は、定評ある情報源や著者から引用された記事が掲載されているサイトを高評価し、金融専門家は、投資家に対して教育的で偏りのない情報が提供されているサイトを高評価するなど、領域の違いによっても基準が異なった（表3、表4）。

ここでは、主題の専門知識をもたない一般消費者は、視覚的な印象に基づいて Web サイトを評価している点に留意する必要がある。それは、Web に典型的な探索パターン（結果を見て判断を下すまでに時間をかけない）と明らかに一致する行動である。

4. Web の信頼性評価の過程モデル

Wathen & Burkell（2002）は、情報源の信頼性評

価に関する先行研究の広範なレビューを行い、そこから得た知見を統合するものとして、Web 情報源の信頼性評価の過程モデルを示した(図1)²⁷⁾。

このモデルによると、Web サイトに入った段階では、まずデザイン設計といった表面的要素を用いて評価を行う。この段階で基準を満たさない場合はサイトを去り、満たした場合は次の段階に進む。そこでは、メッセージの妥当性などから内容的な評価を行う。ここで基準を満たさない場合はサイトを去り、満たした場合は最終段階である個々の状況の照合(個々の抱える問題状況により様々な基準があるとされる)へ進むことになる。

表3 専門家と一般消費者による健康サイトの評価基準

評価項目	専門家	一般消費者
所属機関	43.9%	16.5%
情報の典拠	25.8%	6.5%
会社の業務目的	22.7%	17.8%
情報の焦点	19.7%	33.0%
広告	13.6%	21.3%
デザイン設計	7.6%	41.8%
情報バイアス	4.6%	14.8%
情報のデザイン	3.0%	28.3%
記述の質	3.0%	8.3%
情報の正確さ	1.5%	12.2%

Wathen らのモデルは、実際の Web 探索場面における利用者の情報評価の過程をダイナミックな行動と捉え、段階的なモデルを示している点で興味深い。しかし、それは文献レビューをもとに考案されたものであり、実証されたものではない。従って、実際に調査を行って検証する必要がある。

Wathen らのモデルとこれまで述べてきた先行研究の知見については、Ⅲ章において大学生を被験者とした実験調査の結果をもとに検証する。

Ⅲ. 大学生の Web 探索行動調査

大学生のサーチエンジンと OPAC の検索過程

表4 専門家と一般消費者による金融サイトの評価基準

評価項目	専門家	一般消費者
情報の焦点	40.3%	18.9%
会社の業務目的	35.8%	21.0%
情報バイアス	29.9%	8.5%
デザイン設計	16.4%	54.6%
情報のデザイン	13.4%	33.0%
所属機関	10.5%	24.5%
記述の質	10.5%	9.0%
広告	1.5%	8.3%
情報の正確さ	0.0%	8.0%
情報の典拠	0.0%	0.0%

注) 表3、表4中のパーセンテージは各項目に対するコメントがコメント全体に対して出現した割合を示す。

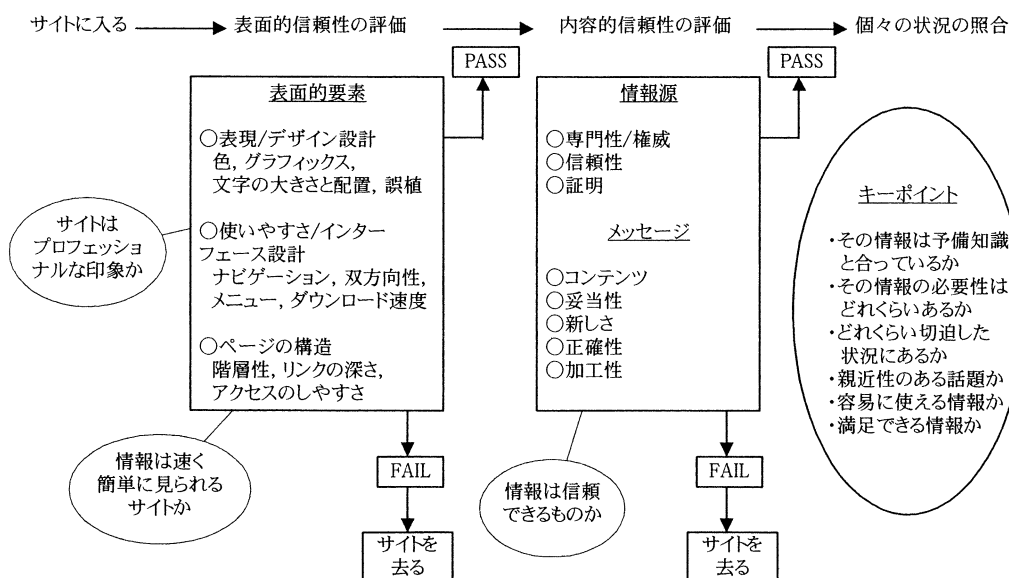


図1 Web 情報源の信頼性評価の課程モデル (Wathen ら, 2002)

をプロトコル分析法により調査し、大学生の Web 探索行動を分析した結果について、以下に述べる。

1. 実験の目的

- 1) Web の探索行動の特徴を検証する
- 2) Wathen らのモデルをもとに、Web 情報源の評価とフィルタリングがどのように行われているかを検証する。

2. 実験の方法

大学生のサーチエンジンと OPAC の検索過程における発話と行動のデータを、2 名 1 組による対話法により採録した (2004 年 7 月 3 日)。対話法の利点として、被験者の緊張感が軽減されること、話し合うことにより今起こっていることをどう捉えているかを相手に伝える必要があり、発話の必然性が明らかになることがある^{28, 29)}。

分析は、採録したデータから書き起こしたトランスクリプションを用いてプロトコル分析法により行った。その際、齋藤ら (2001)³⁰⁾ による行動レベルを援用し、検索行動と情報評価の過程をモデル化して検証した。

1) 被験者 大学 1 年生 4 名 (男 3、女 1)。サーチエンジン利用歴は 3 年以上とした。被験者は、実験以前にサーチエンジンや OPAC 等の検索方法に関する基本的な情報探索法の指導を受けている。

2) 手順 被験者 4 名は、2 人 1 組にグループ分けされた。課題の教示と基本的な操作説明の後、被験者は端末を 2 人 1 組で使用し、2 つの課題に取り組んだ。被験者は、課題 1 はサーチエンジン Google³¹⁾、課題 2 は総合目録データベース Webcat³²⁾ で検索し、選んだページは画面の印刷ボタンを押すことを指示された。制限時間は、課題 1 が 15 分、課題 2 が 7 分とした。被験者の実験中の発話は録音され、画面は録画された。実験終了後に、結果の満足度や情報の選択理由についてインタビューを行った。

3) 課題 課題は、被験者が所属するゼミで取り組んでいるテーマの一つである「若年犯罪」に関わるものとした。これは、できる限り日常に近い場面を設定し、自然な行動を引き出すためである。

課題 1. 「青少年の凶悪化」が騒がれ、更なる少年法改正論議も活発です (新聞記事コピー参照)。これらの報道を検証するために、凶悪事件等の少年犯罪件数の推移を調べてゼミで発表することになりました。発表準備のための情報を Google で探して下さい。

課題 2. 上記テーマで参考になりそうな文献を Webcat で探して下さい。

3. 実験の結果

(1) 探索行動の特徴

サーチエンジンの実験結果によると、1 クエリーあたりの検索時間は平均 3.1 分、検索語数は平均 2.5 語、結果から見た特定ページの閲覧数は平均 3.0 件であった (表 5)。これを、先行研究の調査結果^{20, 22)} と比較すると、検索時間及び検索語数において、類似する傾向が示されている。

結果を確認する行動では、実行されたクエリー全体の 87% において、被験者は一覧表示画面の 1 ~ 2 頁めをブラウズしただけで結果の確認を中止し、それより後に続く結果を見なかった。また、ヒット件数はほとんど気にしていない。その後のインタビューで確認したところでは、被験者は結果のページ順が後になるほど情報要求と離れていくという認識を共通してもっていた。

Web の探索パターンでは、行動レベルの前段階と後段階 (キーワードと検索結果、検索結果とページ、ページとリンク先ページ) を規則的に反復する行動特徴が見られる。また、ルートを浅く規則的に反復する行動と、1 つのルートをやや深く進みながら探索する行動の、2 つの異なるパターンが見られた (図 2)。

OPAC の検索では、試行回数に対してノーヒットになる確率が Group1 では 64%、Group2 では

表 5 サーチエンジンの実験結果

項目	Group1	Group2	平均
検索時間 (1クエリーあたり)	1.3分	5.0分	3.1分
検索語数 (1クエリーあたり)	2.9	2.0	2.5
ページ閲覧数 (1クエリーあたり)	2.0	4.0	3.0

検索結果の満足度について実験後のインタビューで確認したところでは、Web では満足1名、ある程度満足3名としたのに対し、OPAC では4名がやや不満としており、不満の理由を“0件（ノーヒット）が多い”“結果を見ても内容がわからないので判断できない”と答えている。

情報源の評価やフィルタリングがどのように行われているかを検証するために、被験者の発話プロトコルからトランスクリプション(付録資料1)を作成して分析した。分析結果をもとに、検索事例における情報評価の過程を図示して、Wathen

図3は、Webの探索における情報評価の事例である。被験者は、キーワードを設定して検索実行し、表示された結果一覧において明確な主題を示すタイトルが目に入り詳細ページの閲覧に入る。そして、グラフに示されたデータを評価して「要る」と判断している。被験者は、明確なタイトルや見出しが付されたページや、画像等を用いて情報を見やすく構成されたページを選ぶ傾向があった。Web情報を選ぶ際のポイントについて、その後のインタビューでは“フィーリング、感覚的”“直感で良さそうなもの”と答えており、情報の内容的な信頼性を評価する発話や、発信者やドメインを確認する行動(例：トップページへたどるなどして発信者の所属を確かめる)は見られなかった。

Group1

キー	結果	ページ	リンク		
① search	p1	link	p1-1	link	p1-1-1
			p1-1	return	
			p1-1	link	p1-1-1
② search	p1	link	p1-8	link	p1-8-1
			p1-8	return	
			p1-8	link	p1-8-1
	p1	link	p2-5		
	p2	link	p2-8		
			p2-8	return	
			p2-8	link	p2-8-1
			p2-8	return	
			p2-8	link	p2-8-1
			p2-8	return	
③ search	p1	link	p1-7	link	p1-7-1
			p1-7	return	
			p1-7	link	p1-7-1
			p1-7	return	
④ search	p1	link	p1-3	link	p1-3-1
			p1-3	return	
			p1-3	link	p1-3-1
			p1-3	return	
	p1	link	p1-2		
	p2	link	p2-2		
			p2-2	return	
			p2-2	link	p2-2-1
			p2-2	return	
			p2-2	link	p2-2-1
⑤ search			p2-6		
			p2-7		
			p2-7	return	
			p2-7	link	p2-7-1

Group2

キー	結果	ページ	リンク		
① search	p1	link	p1-1	link	p1-1-1
			p1-1	return	
			p1-1	link	p1-1-1
			p1-1	return	
			p1-1	link	p1-1-1
			p1-1	return	
			p1-1	link	p1-1-1
			p1-1	return	
			p1-1	link	p1-1-1
			p1-1	return	
② search	p1	link	p1-7	link	p1-7-1
			p1-7	return	
			p1-7	link	p1-7-1
			p1-7	return	
	p1	link	p2-9		
	p2	link	p3-6		
			p3-6	return	
			p3-6	link	p3-6-1
			p3-6	return	
			p3-6	link	p3-6-1
③ search	p1	link	p5-9		
			p5-9	return	
			p5-9	link	p5-9-1
			p5-9	return	
			p5-9	link	p5-9-1
			p5-9	return	
			p5-9	link	p5-9-1
			p5-9	return	
			p5-9	link	p5-9-1
			p5-9	return	
④ search	p1	link	p1-1	link	p1-1-1
			p1-1	return	
			p1-1	link	p1-1-1
			p1-1	return	
			p1-1	link	p1-1-1
			p1-1	return	
			p1-1	link	p1-1-1
			p1-1	return	
			p1-1	link	p1-1-1
			p1-1	return	
⑤ search			p1-4		
			p1-4	return	
			p1-4	link	p1-4-1
			p1-4	return	

– 7 –

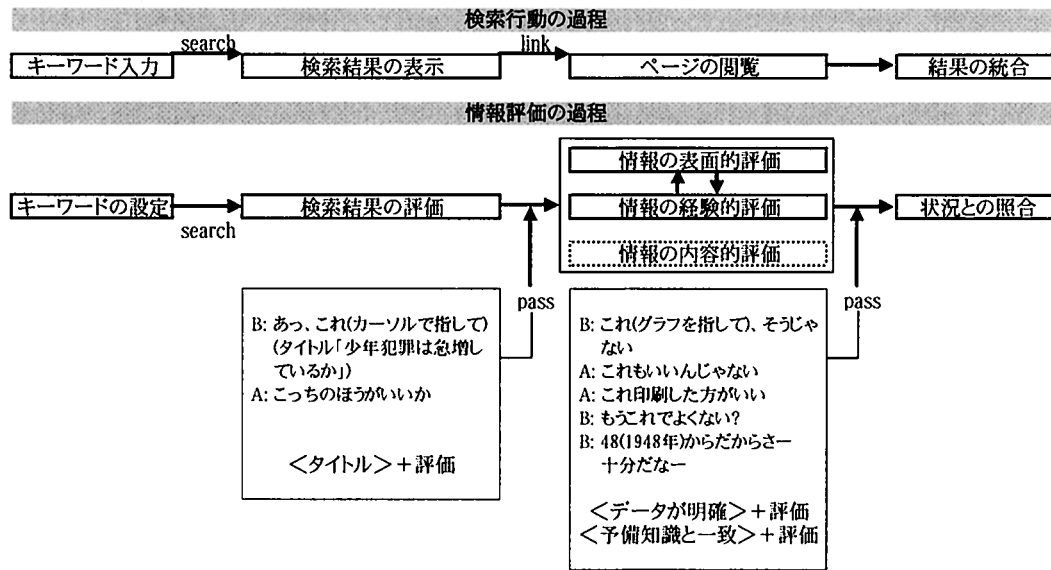


図2 Web探索における情報評価の事例

4. 考察

(1) Webの探索パターン

本調査の結果は、先行研究が示す Web 探索行動の単純性を裏付けるものとなった。被験者は、2～3語の検索語を用いて、結果は最初の1、2ページだけを見て判断する傾向がある。また、被験者はヒット件数をほとんど見ない。これは、過去のサーチエンジンの検索経験から結果の表示順が後になるほど情報要求とは離れていくという認識をもっており、ヒット件数に関係なく、結果の前の方に出てくるページを見ればよいと考えているからであった。

Webの探索パターンにおいて、検索行動レベル間を規則的に往復する行動が見られたことは、被験者に常に元の位置に戻れるようにする、という意識が働いているためであると考えられる。また、ルートを規則的に浅く反復する行動と、ルートをやや深く進みながら探索する行動という異なる特徴が見られたことは、Webの探索パターンのバリエーションを示している。

OPACの検索では、ノーヒットの事例が多かった。検索課題の性質による影響を考慮する必要もあるが、ノーヒット原因の多くは、サーチエンジンに特徴的な検索パターンをそのままOPACに用いたことによるものであった。サーチエンジンは、Web上の膨大な文書の本文そのものを検索対象とする全文検索(full-text search)システムで

ある。Googleのようなロボット型のサーチエンジンは、主題階層、構造化された統制語や件名の理解が必要とされるOPACの検索システムとは明らかに異なる構造をもっており、その検索経験にもとづく認識は、OPACにおける主題検索を困難にする一因となっている。

(2) 情報評価モデルの検討

Web探索における情報の評価では、結果の見切りが速いことが特徴的であった。これは、Webページの閲覧に入った段階で、見出しや情報のデザイン設計といった視覚的要素を基に不要な情報のフィルタリングを行うためであった。

実験結果をもとに Wathen らのモデルを検証したところ、情報評価の最初の段階で表面的要素を用いる点では一致している。しかし、Wathen らのモデルで、その次の段階とされた内容的評価、すなわち情報の妥当性や正確さ、著者の権威や専門性といった情報の信頼性に影響を与える指標に照らして評価する発話や行動はほとんど見られなかった。大学生のWebの探索過程において、情報源の内容的信頼性の評価は欠落する傾向が示された。

IV. 結論

Webの探索プロセスについて先行研究のレビューと大学生に対する実験調査から分析を行った。

先行研究の検討では、Web と従来の情報検索システムを比較し、Web の探索パターンの単純性を示した。また、Web の信頼性評価における指標と情報評価モデルについて検討した。

次に、これらの知見を検証するために、大学生のサーチエンジンと OPAC の検索過程についてプロトコル分析法による調査を行い、1) Web の探索は単純なパターンが規則的に繰り返される反復行動である、2) サーチエンジンの検索経験を積んだ利用者は、OPAC の検索においても Web に特徴的な検索パターンを用いる傾向がある、3) Web 情報源の評価は、視覚的要素をもとに判断される傾向があり、内容の質的評価は欠落している、という結果を得た。

V. 今後の課題

1. 調査上の課題

本論の調査では、プロトコル分析法という質的な調査方法を用いることにより、ログ分析等の量的な手法で検証することが難しい側面、すなわち検索過程で被験者が問題状況をどのように捉えているか、なぜその結果を選んだのか（あるいは選ばなかったのか）といった行動の分析を行うことができた。今後は、調査結果をもとに検証実験を行い、少数事例からの知見の一般性を高めていくことが課題となる。

2. システム上の課題

インターネットに直接情報を求めようとする利用者に対し、図書館がその存在感を示すためには、信頼性の高い情報資源の提供と、発見及びアクセスの支援、ナビゲーション機能等をもつ情報検索システムの構築が必要である。

学術分野における有用な情報資源への主題によるアクセス支援を行うものに、サブジェクトゲートウェイがある。サーチエンジンとの対比でいえば、サブジェクトゲートウェイの特徴は、「質を重視」している点にあり、質の高さを維持するために「人間が処理をする」ことが条件として求められている³³⁾。有用な情報資源の探索、資源に関する記述（メタデータ）の作成において、図書館情報の専門家と主題の専門家との協力が重要である。

利用者アクセスの改善に向けた対応も必要であ

る。Bates (2003) は、LC アクションプラン³⁴⁾ に基づく報告書の草案において、一般に好まれている検索手法はブラウジングであり、これまでの検索式中心の考え方をブラウジング中心へ転換する必要があると述べている³⁵⁾。Bates は、利用者アクセスを改善する方策として、利用者アクセス語彙の構築、書誌ファミリーによるリンク構造、目録以上の内容情報（要約、抄録、目次）へ誘導する段階的なインターフェース、という3つの具体的な提言を行っている³⁵⁾。

「深層ウェブ (deep web)」と呼ばれる通常のサーチエンジンでは検索できない情報源の問題もある。ネットワーク上の情報資源は「表層ウェブ (surface web)」と「深層ウェブ (deep web)」に分類できる。静的な HTML 等で構成されロボット収集が容易な部分が「表層」、データベース等からアクセスの都度動的に生成され容易に収集できない部分が「深層」であり、深層ウェブの情報量の方が400~500倍と、圧倒的に大きいと言われている³⁶⁾。深層ウェブに対する収集技術の確立が課題とされる。

3. 利用教育上の課題

利用者自身が情報源を注意深く評価し賢明な判断を行うためのスキルを身につけるためには、利用教育の問題が重要である。

高校のカリキュラムでは、2003年度より普通教科「情報」が必修となり、大学図書館における利用者の多くはコンピュータの操作やインターネットの使い方等を一通り学んでくることになる。しかし、検定教科書の記述を見る限り、コンピュータやソフトを使いこなす技能に関しては詳しく説明されているが、情報源の評価をどのように行うかについてはほとんど触れられていない。

図書館においては、利用者の情報探索行動に応じた情報利用教育が必要である。情報源の信頼性評価のためのスキルは重要な課題の一つであり、情報の評価法を軸とした指導プログラムの作成と実践など、具体的な対応の検討を進めていく必要がある。

注・引用文献

- 1) Koehler, W. An analysis of web page site constancy and permanence. *Journal of the American Society for Information*

- Science and Technology*. vol. 50, no. 2, 1999. p.162-180.
- 2) 越塚美加. 文献のブラウジングが研究過程に与える影響. 学術情報センター紀要. no. 8, 1996. p.131-142.
- 3) Kari, Jarkko. Web information seeking by pages: An observational study of moving and stopping [online]. *Information Research*, vol. 9, no. 4, 2004. <<http://InformationR.net/ir/9-4/paper183.html>> (last access 1/29/2005)
- 4) Jansen, B. J.; Pooch, U. A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology*. vol. 52, no. 3, 2001. p.235-246.
- 5) Hoelscher, C. How internet experts search for information on the web. Paper presented at the World Conference of the *World Wide Web, Internet, and Intranet*, Orlando, FL, 1998.
- 6) Jansen, B. J.; Spink, A.; Saracevic, T. Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing and Management*. vol. 36, no. 2, 2000. p.207-227.
- 7) Silverstein, C.; Henzinger, M.; Marais, H.; Moricz, M. Analysis of a very large web search engine query log. *SIGIR Forum*, vol. 33, no. 1, 1999. p.6-12.
- 8) Hsieh-Yee, I. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science*, vol. 44, no. 3, 1993. p.161-174.
- 9) Koenemann, J.; Belkin, N. A case for interaction: A study of interactive information retrieval behavior and effectiveness. Paper presented at *Conference on Human Factors in Computing Systems*, Vancouver, Canada, 1996.
- 10) Siegfried, S.; Bates, M.; Wilde, D. A profile of end-user searching behavior by humanities scholars: The Getty online searching project report no. 2. *Journal of the American Society for Information Science*, vol. 44, no. 5, 1993. p.273-291.
- 11) Millsap, L.; Ferl, T. Search patterns of remote users: An analysis of OPAC transaction logs. *Information Technology and Libraries*, vol. 11, no. 3, 1993. p.321-343.
- 12) Peters, T. When smart people fail: An analysis of the transaction log of an online public access catalog. *Journal of Academic Librarianship*, vol. 15, no. 6, 1989. p. 267-273.
- 13) Wallace, P. How do patrons search the online catalog when no one's looking?: Transaction log analysis and implications for bibliographic instruction and system design. *RQ*. vol. 33, no. 3, 1993. p.239-252.
- 14) <http://www.fireball.de/>
- 15) <http://www.excite.com>
- 16) <http://www.altavista.com>
- 17) Bates, M. J. Information search tactics. *Journal of the American Society for Information Science*. vol. 30, no. 4, 1979. p.205-213.
- 18) Bates, M. J. Idea tactics. *Journal of the American Society for Information Science*. vol. 31, no. 5, 1979. p.280-289.
- 19) Pew Internet & American Life Project. The popularity and importance of search engines [online]. 2004. <http://www.pewinternet.org/pdfs/PIP_Data_Memo_Searchengines.pdf> (last access 1/29/2005)
- 20) Spink, A.; Jansen, B. J.; Wolfram, D.; Saracevic, T. From e-sex to e-commerce: Web search changes. *IEEE Computer*. vol. 35, no. 3, 2002. p.107-111.
- 21) <http://www.alltheweb.com/>
- 22) Jansen, B. J.; Spink, A. An analysis of web documents retrieved and viewed. In *Proc. of the 2003 Internet Computing Conference*, Las Vegas, Nevada, 2003. p.65-69.
- 23) <http://credibility.stanford.edu/>
- 24) Fogg, B. J. Stanford guidelines for web credibility: A research summary from the Stanford Persuasive Technology Lab. Stanford University [online]. 2002. <www.webcredibility.org/guidelines> (last access 1/29/2005)
- 25) Fogg, B. J.; Soohoo, C.; Danielsen, D.; Marable, L.; Stanford, J.; Tauber, E. How do people evaluate a web site's credibility?: Results from a large study [online]. 2002. Stanford Persuasive Technology Lab, Stanford University. <<http://www.consumerwebwatch.org/dynamic/web-credibility-reports-evaluate-abstract.cfm>> (last access 1/29/2005)
- 26) Stanford, J.; Tauber, E.; Fogg, B. J.; Marable, L. Expert vs. online consumers: A comparative credibility study of health and finance web sites [online]. 2002. <<http://www.consumerwebwatch.org/dynamic/web-credibility-reports-experts-vs-online-abstract.cfm>> (last access 1/29/2005)
- 27) Wathen, C. N.; Burkell, J. Believe it or not: Factors influencing credibility on the web. *Journal of the American Society for Information Science and Technology*. vol. 53, no. 2, 2002. p.134-144.
- 28) Miyake, N. Constructive interaction and the iterative process of understanding. *Cognitive Science*. 10, 1986. p.151-177.
- 29) Suchman, L. A. Plans and situated actions: The problem of human-machine communication. Cambridge University Press. 1987.
- 30) 齋藤ひとみ; 三輪和久. WWW の情報探索における人間の行動プロセスの実験的検討. 電子情報通信学会技術研究報告. ET2001-14, 2001. p.23-28.
検索エンジンによる Web の探索行動には次の4つの行動レベルがあるとしている.
 - ・検索レベル (キーワードレベル) : 検索エンジンの検索
 - ・検索結果レベル: 検索レベルで検索した結果の閲覧
 - ・ページレベル: 検索結果から選択したページの閲覧
 - ・リンク先ページレベル: ページからリンクでつながれているページの閲覧

- 31) <http://www.google.co.jp/>
- 32) <http://webcat.nii.ac.jp/>
- 33) 緑川信之；伊藤真理；松林麻実子．サブジェクトゲートウェイ：ネットワーク上の知識集積．知的コミュニティ基盤研究センター・モノグラフシリーズ 1．筑波大学知的コミュニティ基盤研究センター，2003．
- 34) Library of Congress. Bibliographic control of web resources: a Library of Congress action plan [online]. 2003. <<http://lcweb.loc.gov/catdir/bibcontrol/actionplan.pdf>> (last access 1/29/2005)
- 35) Bates, M. J. Task force recommendation 2.3 research and design review: improving user access to library catalog and portal information: final report. Library of Congress [online]. 2003. <<http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf>> (last access 1/29/2005)
- 36) The deep web: surfacing hidden value [online]. <<http://www.brightplanet.com/technology/deepweb.asp>> (last access 1/29/2005)

(付録資料)

プロトコル分析に用いた発話資料の一部

データ採録年月日: 2004年 7月 3日
採録場所: N大学附属図書館
被験者: N大学・学部 1年生2名(男2)
検索システム: サーチエンジンGoogle

時間	-	検索式	情報の利用件数 ()内はヒット数	キーワード	行動レベル 結果	ページ	リンク	発話記録 (話者A:キーボード操作)
0:00:00	開始							
	1	"少年"and"犯罪"	一覧 10 / 詳細 1 / 取得 - (364,000件)	入力	一覧p1	詳細p1-1		A よし、少年犯罪の件数でやってみるか B 少年と犯罪、分けたほうがいいかも A ああ、じゃあ、少年と犯罪で A え、Google検索しま-す A あ、件数って A 件数も入れた方がいいかな、あー B (詳細表示に目を通しながら)件数入れた方がいいい?
	2	"少年"and"犯罪"and"件数"	一覧 20 / 詳細 3 / 取得 - (18,400件)	入力	一覧p1	詳細p1-8		A ざっきといっしょだね B いっしょ A これ出そう? A あれ、左右だし B あるかなこれ、なさそうだ A なさそうかな、ていうかさ、今のグラフ絶対にあったのにさ AB あー(声をそろえて) おしいな(A) B なんだ... A あーもうちょっと調べてみようか (画面を指差す) A 遅い、やばいこれ時間かかりそう A あ、きた B あー B これいいんじゃないの A うん B うん、これってさあ、少年犯罪件数... A うん、これがさあ、89年からなんだよね、一応 B とりあえずでも使えそうだけど B (実験者に対して)使えそうだったら印刷すればいいんですか? B 印刷は...これだ A ???、これ B これいるの A "推移"も入れてみたら
0:01:26					一覧p1 一覧p2	詳細p2-5		
					一覧p2	印刷		
0:02:28					一覧p2	詳細p2-8		

3	“少年”and“犯罪”and“件数” “and”推移”	一覧 10 / 詳細 1 / 取得 1 (3,830件)	入力	一覧p1	詳細p1-7	<p>B “推移”入れるとえらく長い</p> <p>B そうすっか? これじゃないか?</p> <p>A 遅い</p> <p>B あっ、これ</p> <p>A こっこのほうがいいか</p> <p>B これさっきのやつ、グラフが出てこないんだよ、これ</p> <p>A やばい、遅い</p> <p>B あれ? あ、出た</p> <p>B これそうじゃない</p> <p>A これもいいんじゃない</p> <p>A これ印刷した方がいい</p> <p>B 印刷は…これか…OK</p> <p>B もうこれでよくない?</p> <p>B 48からだからさー十分だなー</p> <p>A だよ、ね、ふつうに</p> <p>B うん、ね、これでいいじゃん</p> <p>A これはいいでしょ</p> <p>B ねえこれさ、どっからとんだんだろね</p> <p>A 戻ったんだ</p> <p>B 戻ったんだよねー、戻ったわけだ</p> <p>B こきてさー(笑いながら)、どれにしたんだっけ</p> <p>A ま、いいよ</p> <p>A どうする? あと</p> <p>B ていうか、何か…、もう終わりじゃん(笑いながら)</p> <p>A いや、まだ、でも終わりでやっただほうがいい</p> <p>B マジで? な、何があるかな</p> <p>A あー</p> <p>B 何か…</p> <p>A こんくらい?</p> <p>A また出ましたよ</p> <p>A 「犯罪白書」ってよさそうじゃない?</p> <p>B あ、いいかもしれん</p> <p>A (入力)</p>
0:03:26				一覧p1		<p>A これは違う</p> <p>B これは違うね</p> <p>A これさっきのやつか、とぶなっつの</p>
0:03:55				一覧p2	詳細p2-2	
0:04:36			入力	一覧p1	詳細p1-3	
0:05:12	4	“少年”and“犯罪”and“件数” “and”推移”and“犯罪白書”		一覧p1	詳細p1-2	