

# 検索エンジンを用いた情報検索におけるユーザ行動の分析

## Analysis of User's Behavior in Information Retrieval Using Search Engine

桑折 章吾<sup>1\*</sup> 加藤 優<sup>1</sup> 高間 康史<sup>1</sup>  
Shogo Kori<sup>1</sup>, Yu Kato<sup>1</sup>, Yasufumi Takama<sup>1</sup>

<sup>1</sup> 首都大学東京大学院システムデザイン研究科

<sup>1</sup> Graduate School of System Design, Tokyo Metropolitan University

**Abstract:** 本稿では、検索エンジンを用いた情報検索におけるユーザ行動を分析した結果について報告する。我々は、「動向に関する問い」を対象とした検索エンジン構築を目指し、その基本的検索機能について検討を進めている。既存検索エンジンを用いた検索でも、ユーザは異なる意図に基づく基本的検索を組み合わせて目的を達成しているとの考えに基づき、本稿では検索意図の観点からユーザの情報検索行動を分析し、得られた結果に基づき動向に関する基本的検索機能について考察する。

## 1 はじめに

本稿では、検索エンジンを用いた情報検索におけるユーザ行動を分析した結果について報告し、得られた結果に基づき我々が目指す「動向に関する問い」を対象タスクとした検索エンジンの基本検索機能について考察する。Webの魅力の一つとして、世界中のリアルタイムな情報が収集可能である事が挙げられる。近年では、ソーシャルメディアの普及によりリアルタイムな情報がますます注目されている。その一方で、Webが利用されるようになってから20年弱が既に経過し、Web上には膨大な量の情報が蓄積されている。このように蓄積された情報に着目し、過去の情報を知るためのリソースとしてWebを有効的に活用していくことも検討すべきであると考え、既存検索エンジンが提供する機能と、ユーザの情報収集目的との乖離が大きいという問題がある。すなわち、既存検索エンジンが提供するものは、キーワードベースの検索要求指定、ページ単位での結果出力といった低機能にとどまったままであり、情報要求をキーワードに分解するのに要するユーザの負担が大きいと考える。

検索エンジンの知的化・高機能化に関するアプローチとしては、対象ドメインを限定することが考えられるが[8][10]、本稿で検討している検索エンジンでは、ドメインに依存せず、広く一般的に利用可能であることを目指している。提案する検索エンジンでは、対象タスクに特化したいくつかの基本検索機能を検討してい

るが、それらは組み合わせて用いることで多様かつ高度な検索目的を達成可能である必要がある。既存検索エンジンでも、ユーザは異なる意図に基づく基本的検索を組み合わせて目的を達成しているとの考えに基づき、本稿では検索意図の観点からユーザの情報検索行動を分析する。

本稿では、ユーザの情報検索行動を調査するために行った実験について述べ、得られた結果に基づき動向に関する基本検索機能を考察する。実験では既存検索エンジンを使いWebから答えを見つける問題を実験協力者に出題した。入力されたクエリを検索意図毎に分類して分析した結果、ユーザは自らの情報要求を満たすために異なる意図に基づく基本検索機能を多様に組み合わせることを示す。分析結果に基づき、構築中の検索エンジンに必要な基本検索機能について考察する。

## 2 関連研究

### 2.1 次世代検索エンジン

Webが普及してから20年弱が経ち、Web上は情報過多となってきた。現在、Web上に蓄積された情報を探す方法としては、検索エンジンが最も用いられている。しかし、既存の検索エンジンは指定したキーワードを含むWebページを返すという汎用的ではあるが低機能なものにとどまっているため必要とする情報

\*連絡先：首都大学東京大学院  
システムデザイン研究科情報通信システム学域  
〒191-0065 東京都日野市旭が丘6-6  
E-mail: kori-shogo@sd.tmu.ac.jp

に辿り着くまでに何度も検索を繰り返す必要がある場合が発生する。このような手間を省くために、対象を絞ることでより効率的な検索を実現することを目指す次世代検索エンジンの研究・開発がなされている。

亀井ら [4] は、WWW に存在するソフトウェア開発に関する知見や情報を検索するための検索エンジンを提案している。過去に多くのソフトウェアが開発され、それらに関するノウハウや関連情報などが Web 上で多数公開されている。しかし、それらは体系化されず Web 上に点在しており、網羅的・効率的に情報収集することは困難である。そのため現状では、似たようなソフトウェアが開発されていたり、同じようなミスでソフトウェア開発が滞ることがある。亀井らの提案する検索エンジンは、ソースコードそのものやそのコメント、開発日記、Tips などソフトウェアの知見に関する情報にドメインを特化することで、既存検索エンジンよりも効率的な検索を目指している。

対象ドメインを限定しない検索エンジンとして、動向情報を対象としたコンテキスト検索エンジンが提案されている [6][7]。動向情報とはある商品の価格や売上の状況、ある会社の業績状況、内閣や政党の支持状況などの事であり、幾つかの統計量の時系列データを基として、その変化を通時的にとらえつつ、それらを総合的にまとめ上げることで得られるものである [5]。動向情報は検索エンジンの検索数やヒット数などの主観的動向情報と、アイテムの価格や販売量、生産量などの客観的動向情報に分けられる [6]。文献 [6] では、主観的動向情報として Google Trends<sup>1</sup> で公開されている検索数や Yahoo! 検索ランキング<sup>2</sup> で公開されている急上昇ワードなどを収集対象としている。客観的動向情報としては、ベジ探<sup>3</sup> で公開されている野菜の価格や自転車産業振興協会<sup>4</sup> で公開されている自転車の生産台数などの統計データを収集対象としている。収集した動向情報はデータベース (MySQL) に格納し、Web アプリケーションフレームワークに Ruby on Rails を用いてシステムを実装している。プロトタイプシステムのインタフェースを図 1 に示す。このシステムでは「指定アイテムに関する動向情報のピーク (最大値) 時期の検索」、「指定期間に動向情報の最大値を持つアイテムの検索」の 2 つの基本検索機能を実装している。

The image shows two screenshots of a web search interface. The top screenshot is titled 'アイテムから探す' (Search by item) and contains a text input field for '検索ワード:' and a '検索' (Search) button. The bottom screenshot is titled '期間から探す' (Search by period) and contains a form for '検索対象期間' (Search target period) with '年' (Year) and '月' (Month) dropdowns, and a '検索' (Search) button.

図 1: コンテキスト検索エンジンのインタフェース

## 2.2 情報検索におけるユーザ行動

既存検索エンジンを用いた情報検索では、ユーザは異なる意図に基づき基本検索機能を組み合わせて目的を達成している。ユーザの検索意図はクエリとして表現されるが、うまく表現できない場合もある。そのような場合、検索を繰り返しても、膨大な検索結果の中に必要な情報を含むページが埋没してしまい、必要とする情報にたどり着く事は難しい。

藤田ら [3] らは、ユーザの連続した検索からクエリ変更意図を推測することで、ユーザの検索の先を読み、自動で検索する先読み検索を提案している。提案手法では、クエリログからユーザのクエリ変更意図について分析し、その結果に基づき SVM によるクエリ変更意図の自動分類を行う。クエリ変更意図毎に先読み検索を行っている。

南ら [9] は、ユーザが問題解決を目的に複数の検索結果を確認しながら情報を集めて行く際の作業効率向上を目的とし、検索結果のフィルタリングを行っている。ユーザの Web ページ閲覧時の行動をモニタリングして検索タスクにおけるユーザ意図を動的に抽出する手法を提案し、検索結果のフィルタリングシステムを実装している。

旭ら [1] は、「iPod を買う」→「iPod を使う」→「iPod が壊れて修理する」のようにある話題の中で行われる一連の行動の流れを行動連鎖と呼び、ブログのエントリ内、エントリ間という 2 つの観点からシーケンシャルパターンマイニングを用いて行動連鎖の抽出を行っている。抽出した結果に基づきユーザが目的とする行動に応じて必要な Web ページをランキングしてユーザに提示するシステムを提案している。順序だてて行動連鎖をユーザに提示することで、ユーザは自分にとって必要な情報を効率よく調べることが可能となる。また、行動プロセス提示によりユーザは問題解決のためにどのような事を調べれば良いのかを把握することができる。

<sup>1</sup><http://www.google.co.jp/trends/>

<sup>2</sup><http://searchranking.yahoo.co.jp>

<sup>3</sup><http://vegetan.alic.go.jp>

<sup>4</sup><http://www.jbpi.or.jp>

### 3 ユーザ行動の分析

検索意図の観点からユーザの情報検索行動を調査する実験を行った。3節で行った実験の概要および、その結果を分析し検索意図を分類した結果について示す。4節では3節で定めた分析意図によりログデータにラベル付けを行い分析を行った結果を示すと共に、構築中の検索エンジンに必要な基本検索機能について考察する。

#### 3.1 ユーザ行動調査のための実験

実験で用いた問題を図2に示す。実験では、二枚の画像から検索エンジン（Google）を用いて画像の撮影場所を特定する問題を出題した。図2の問題のように答えをどのような視点から探せばよいか、画像をクエリとしてどの様に表現すればよいかは明確ではない場合、実験協力者は自ら解答への道筋を考えなければならない。答えを見つけるアプローチの仕方が様々であり、多様な検索行動が生じることが期待できるためこの問題を選択した。実験は実験協力者3名を対象に行い、各協力者は平均して約20分で正解を出すことができた。実験協力者がどのような検索を行い、どのようなページを開いているかを正確に調査するため実験中にoCam<sup>5</sup>を用いて画面のキャプチャを行った。



図2: 実験で用いた問題

#### 3.2 検索意図の分析

入力されたクエリを分析し、検索意図を図3の様に分類した。実験協力者の検索意図はVerify（検証）とDiscover（発見）の二つのタイプに大別される。また、何かに関する情報を探す際には、対象ページを限定しないInformationalと、特定のWebページの発見

を目的とするNavigationalに分類できる[2]。さらにDiscover-Informationalには、正確に目標を定めた検索（Pinpoint）と幅広い検索結果を期待した検索（Broad）が存在する。Discover-Informational-Pinpointには条件を満たす情報を一つだけ探す検索（Single）と複数の情報を探す検索（Multi）があり、Multiにはそれらが一覧のようにまとめられているWebページを期待した検索（List）と一つずつ別ページに存在することを期待した検索（Item）が存在する。同じクエリであっても検索される段階によって検索意図が異なると考えられる場合があった。

今回の実験で入力されたクエリの例を以下に示す。今回の実験ではItemに該当する検索は行われなかった。

- Verify-Informational  
「市場 スペイン バルセロナ」  
…写真がバルセロナ（スペイン）の市場で撮影したものであることを確認
- Verify-Navigational  
「サン・ジョセップ市場 Google マップ」  
…サン・ジョセップ市場の場所をGoogleマップで確認
- Discover-Navigational  
「バルセロナ wiki」  
…バルセロナについて書かれているWikipediaのページを期待
- Broad  
「ヨーロッパ 市場」  
…ヨーロッパにある市場について幅広い情報を期待
- Single  
「サン・ジョセップ市場 住所」  
…サン・ジョセップ市場の住所が書かれているWebページを期待
- List  
「スペイン 市場 一覧」  
…スペインの市場が一覧のようにまとめられているWebページを期待

<sup>5</sup><http://ohsoft.net/product-oCam.php>

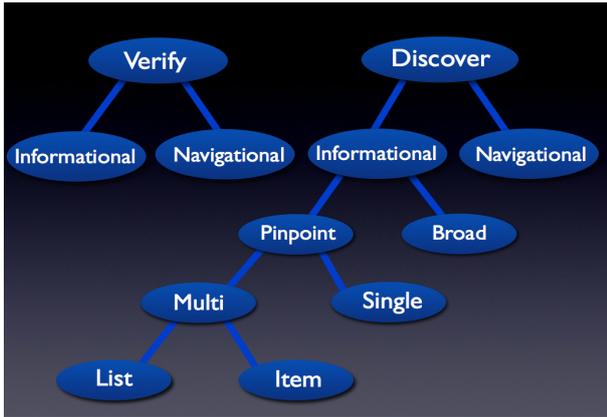


図 3: 検索意図の分類

## 4 実験結果と基本検索機能の考察

### 4.1 ログデータへのラベル付け

ユーザ行動をさらに分析するために図2と同じ様な問題を用いて再び実験を行い、収集したログデータに図3に示したラベルを付けた。実験は計3問行い、問題1では3人、問題2では3人、問題3では5人の実験協力者を対象とした。問題1~3の全実験協力者が答えを導くことができたが、解答に要した時間にはばらつきが見られた。表1に、全実験協力者についてログデータに付与された各ラベルの数を示す。なお実験協力者A~Hは解答が早かった順に上から並べている。表より、DiscoverのNavigational, Broad, Listは利用者が少ない一方、VerifyのNavigationalとInformational, DiscoverのPinpoint (Single, Multi)はほぼ全員が利用していることがわかる。また、3問全てに共通して前半ではPinpoint (Single/Multi/List)、中盤ではVerify-Informational、後半ではVerifyに該当する検索のパターンが多く見られた。問題3で最も解答が早かった実験協力者Fのラベル付け結果を表2に、最も解答が遅かった実験協力者Hのラベル付け結果を表3に示す。ここで、前半、中盤、後半はラベルの総数を3分割したものである。表2、表3より両者とも前半ではPinpoint (Single/Multi/List)、中盤ではVerify-Informational、後半ではVerifyに該当するクエリが比較的多く入力されていることがわかる。また、最も解答が遅かった実験協力者は中盤でMultiに該当する検索の回数が多かったり、後半でSingleに該当する検索の回数が多いなど他のラベルに該当する検索が多く見られ、欲する情報を見つけるのにてまどっていることがわかる。

表 1: 全実験協力者のラベルの数

		Verify		Discover					総数
		Navigational	Informational	Navigational	Broad	Single	Multi	List	
実験1	A	1	3	1	0	5	0	0	10
	B	3	2	0	0	2	1	0	8
	C	1	3	1	3	5	1	0	14
実験2	D	2	4	0	0	2	1	0	9
	E	1	5	0	0	1	8	4	19
	F	5	26	0	2	2	6	0	41
実験3	F	2	5	0	0	0	0	3	10
	C	0	4	1	0	6	3	0	14
	B	1	6	0	1	0	1	1	10
	G	0	16	0	1	1	3	1	22
	H	3	5	0	1	10	14	0	38

表 2: 実験協力者 F のラベル付結果 (実験 3)

		前半	中盤	後半
Verify	Navigational	0	1	1
	Informational	0	3	2
Discover	Navigational	0	0	0
	Broad	0	0	0
	Single	0	0	0
	Multi	0	0	0
	List	2	1	0

表 3: 実験協力者 H のラベル付結果 (実験 3)

		前半	中盤	後半
Verify	Navigational	0	0	3
	Informational	0	5	0
Discover	Navigational	0	0	0
	Broad	1	0	0
	Single	0	0	10
	Multi	10	4	0
	List	0	0	0

### 4.2 基本検索機能の考察

我々が構築中のコンテキスト検索エンジンで想定する検索タスクは3節で示したものと異なるが、既存検索エンジンと同様にユーザの検索意図を満たす機能が必要であるとの考えに基づき、3.2, 4.1節に示した結果に基づきコンテキスト検索エンジンが備えるべき基本検索機能について考察する。

構築中の動向情報を対象としたコンテキスト検索エンジンでは、入力されるクエリとしてアイテムや期間、

変動タイプが考えられる。ここでいう変動タイプとはアイテムのピーク、急激に値が変わった時、最大ピーク、最小ピーク、最初に訪れたピークなど、特徴的な動向の変化を指す。図3に示した既存検索エンジンにおける検索意図のラベル分類を元に、コンテキスト検索エンジンの検索意図を考慮して体系化しなおしたものを図4に示す。コンテキスト検索エンジンでは検索対象がWebページではないため、既存検索におけるNavigationalに直接対応するものは存在しない。そこで本稿ではInformationalはアイテムを指定しない場合、Navigationalは指定した場合の検索としている。変動タイプを指定した場合はPinpoint、指定しなかった場合はBroadとし、最大ピークの様各動向情報に一つしか存在しない変動タイプを指定した場合はPinpoint-Single、複数存在するものを指定した場合はPinpoint-Multiに分類している。以下に上記のラベルに該当するコンテキスト検索エンジンでの検索例を示す。

- Informational-Pinpoint-Multi  
「2013年に売れ始めたアイテムは？」  
入力：期間  
出力：アイテム  
変動タイプ：上昇傾向
- Informational-Pinpoint-Single  
「自転車最も売れた時期に同様に売れたアイテムは？」  
入力：アイテム  
出力：アイテム  
変動タイプ：最大ピーク
- Informational-Broad  
「2013年に特徴的な変動を示したアイテムは？」  
入力：期間  
出力：アイテム  
変動タイプ：指定なし
- Navigational-Pinpoint-Multi  
「自転車が急激に売れた年は？」  
入力：アイテム  
出力：期間  
変動タイプ：急激な上昇
- Navigational-Pinpoint-Single  
「自転車の生産台数が一番少なかった年は？」  
入力：アイテム  
出力：期間  
変動タイプ：最小値
- Navigational-Broad  
「自転車の生産台数の動向について知りたい」  
入力：アイテム

出力：期間、変動タイプ  
変動タイプ：指定なし

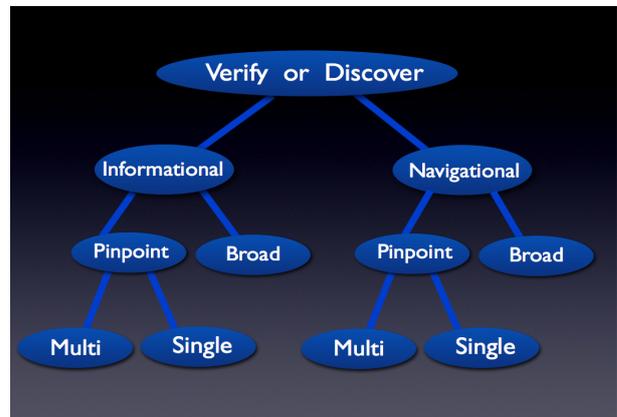


図4: コンテキスト検索エンジンにおける検索意図の分類

## 5 終わりに

本稿では、既存検索エンジンを用いた情報検索におけるユーザ行動を分析した結果について報告し、「動向に関する問い」を対象とした検索エンジンの基本検索機能について考察した。実験結果に基づきユーザの検索意図を分類し、該当するラベルをログデータに付与することで実験協力者の検索行動を分析した。また、ラベルを「動向に関する問い」を対象とした検索エンジンの場合に置き換えることによって、必要な基本検索機能について考察した。今後は、考察結果に基づき基本検索機能の実装を進める予定である。統計局が平成25年6月にAPIを公開するなど情報公開の流れもあり、今後は官公庁も含めて公開されるデータは増える事が期待されるため、動向情報を対象としたコンテキスト検索エンジンがより幅広い分野に対して有効になることが期待できる。

## 参考文献

- [1] 旭直人, 山本岳洋, 中村聡史, 田中克己: 行動連鎖を用いた情報検索支援とWebからの行動連鎖の抽出, DEIM Forum, A7-2, 2009
- [2] C. D. Manning, P. Raghavan, H. Schütze: Ch. 19: Web search basics, Introduction to Information Retrieval, Cambridge University Press, 2008.
- [3] 藤田遼治, 太田学, 徳永徹郎: ユーザのクエリ変更意図に基づく先読み検索, DEIM Forum 2012, A4-4, 2012

- [4] 亀井 俊之, 門田 暁人, 松本 健一: WWW を対象としたソフトウェア検索エンジンの構築, 電子情報通信学会技術研究報告 ソフトウェアサイエンス, Vol.102, No617, pp.59-64, 2007
- [5] 加藤 恒昭, 松下 光載, 平尾 努: 動向情報の要約と可視化に関するワークショップの提案, 情報処理学会研究報告/自然言語処理研究会報告, 2004(108), pp.88-94, 2004
- [6] 加藤 優, 桑折 章吾, 高間 康史: 「動向に関する問い」を対象タスクとしたコンテキスト検索の提案, 人工知能学会, インタラクティブ情報アクセスと可視化マイニング研究会 (第3回), pp.7-12, 2013
- [7] 加藤 優, 高間 康史: Web コンテキスト情報に基づく同時期流行アイテム検索手法の提案, FSS2012, pp.115-118, 2012
- [8] 小久保 卓, 小山 聡, 山田 晃弘, 北村 泰彦, 石田 亨: 検索隠し味を用いた専門検索エンジンの構築, 情報処理学会論文誌, Vol.43, No.6, pp.1804-1813, 2002
- [9] 南 翔太郎, 岡 誠: 閲覧行動モニタリングに基づく検索意図の抽出と検索結果の分類, 情報処理学会報告, HCI-142(8) , pp.1-6, 2011
- [10] 山田 泰寛, 廣川 左千男: 専門検索サイトの動的統合による次世代検索システム DAISEN における検索サイトエディタの開発, 第1回情報科学技術フォーラム, 一般講演論文集第2分冊, pp.11-12, 2002