様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ５ 月 １９ 日現在

機関番号：２２６０４
研究種目：挑戦的萌芽研究
研究期間：2012～2014
課題番号：２４６５００４０

研究課題名（和文）次世代検索エンジンのためのコンテクスト検索手法の確立

研究課題名（英文）Establishment of Context Search Method for Next-generation of Search Engines

研究代表者
高間　康史（Takama, Yasufumi）

首都大学東京・システムデザイン研究科・教授

研究者番号：２０３１３３６４

交付決定額（研究期間全体）：（直接経費）　　3,000,000円

研究成果の概要（和文）：本研究課題では，動向に関する問いにタスクを限定することで，幅広いドメインで利用可能という現在の検索エンジンの特徴を備えつつ，より高度な検索機能を提供するコンテクスト検索手法について研究を行った．既存検索エンジンを利用した予備実験を行い，ユーザの検索意図について調査した結果などに基づき3種類の基本検索機能を提案し，通常のWebブラウザおよびAPIによりアクセス可能なプロトタイプ検索エンジンを開発した．プロトタイプエンジンを用いて実験協力者による評価実験を行い，情報要求を多様な検索意図に基づく複数のクエリに分解し，検索可能であることや，データリソース間の関係性発見に有効活用できる可能性を示した．

研究成果の概要（英文）：Although a Web search engine is a necessary tool for us to access huge amount of information available on the Web, we think that there is a significant difference between function provided by existing search engines and users' information needs. Aiming at narrowing the gap, we focus on the task of answering trend-related queries. By focusing on the specific task, more advanced search functions can be provided compared with existing Web search engines. As the task of answering trend-related queries is supposed to be common in various domains, we expect it could be used for various purposes. Prototype context search engine provides three basic functions, which can be accessed via both of ordinary Web browsers and API. Experimental results using the prototype search engine show that users can break their information needs into several queries using those functions. It is also shown it can be used for finding relationship among different data resources.

研究分野：Webインテリジェンス

キーワード：情報検索　動向情報　検索エンジン　コンテクスト検索　対話的情報アクセス

１．研究開始当初の背景

【背景１】Web の魅力の一つは世界中の最新の情報が入手可能な点であり，Blog や Twitter の普及によりその傾向はさらに強まっている．Blog から映画の興行収入を予測する研究など，学術的関心も最新の情報から近未来を予測する方向が主流となっている．一方，Web が利用されるようになって 20 年弱経ち，膨大な情報が蓄積されているが，過去を知るための情報リソースとしての活用は Google タイムライン検索などわずかしかない．新しい情報のみに着目するのではなく，過去の情報も有効活用する段階に来ていると考える．

【背景２】現在の検索エンジンが提供する検索機能はキーワードベース，ページ単位の低レベルなものであり，利用者は多様な情報要求を低レベルの検索に分割・実行する必要がある．この様な状況を解消するために，情報利用者の実情にあった検索機能を提供する次世代検索エンジンの実現が望まれている．次世代検索エンジンに関して自然言語検索や質問応答システムなどが研究されているが，既存検索エンジンに取って代わりうるものは未だ提案されていない．一方，研究代表者・研究分担者は，新聞記事や統計データを対象とした動向情報の要約・可視化に関する研究に取組む中で，動向に関する問いは一般的なものであり，これに答えるための検索タスクをカバーすることが次世代検索エンジンとして有望であるとの着想に至った．

２．研究の目的

動向に関する問いに答えるための検索タスクをコンテクスト検索と定義し，その検索手法を確立する．動向に関する問いに共通する基本検索タスクを定め，その検索モデルについて研究する．基本検索タスクを実行する次世代検索エンジンのプロトタイプを実装し，有効性について評価する他，基本検索タスクを利用してコンテクスト検索を行うプロセスを対話的情報アクセスの枠組みとして捉え，インタラクション設計及び評価指標について研究を行う．

３．研究の方法

アイテム（商品，人物，出来事など）の栄枯盛衰に関する時間的動向情報を，次世代検索エンジンが対象とするコンテクストと定義する．評判・口コミなどに表れる主観的動向，統計データとして定量的に測定可能な客観的動向に分類し，入手可能な情報リソースについて抽出方法も含めて調査・研究する．特に主観的動向では Web 利用形態の変化も考慮に入れる．これらの動向情報を対象とした基本検索タスクを定め，多様な情報源を統合した検索モデルについて研究する．基本的検索タスクを利用した対話的情報アクセスとしてコンテクスト検索を定義し，インタラクション設計や評価指標について研究する．

４．研究成果

Web から入手可能な，動向情報に関する情報リソースについて調査した結果，当初想定していた「主観的動向」，「客観的動向」の分類を見直し，以下の二種類に分類して収集を行った．

・コンテンツとしての動向情報：各アイテムの価格や販売量に関する統計データの様な，各企業や組織・団体によりコンテンツとして公開される動向情報．
・Web 利用に基づく動向情報：各アイテムをキーワードとして既存検索エンジンで検索した際のヒット数や，ブログ記事数などといった，Web 上でのユーザ活動により発生する動向情報．

上記分類に基づき，前者に関しては総務省統計局から人口，雇用者に関する動向情報など，NHK 放送文化研究所から内閣政党支持率に関する動向情報，内閣府から景気ウォッチャー調査，農畜産業振興機構から野菜の価格や消費者物価指数などのデータを収集し，プロトタイプシステムで検索可能とした．後者に関しては，Yahoo! Japan が提供する年間検索ワードランキングで上位に含まれる人名や企業名，サービス・製品名などを中心とした約 400 語についてのヒット数，ブログ記事数，検索数などを収集して検索可能とした．

収集した動向情報について，特徴的な変動を示した期間に基づく検索を可能とするため，以下に定める変動タイプを抽出した．

・最大値（MAX）／最小値（MIN）：各動向情報が最大値／最小値を取る月
・急上昇（SI）／急下降（SD）：3 ヶ月以内に，その動向情報の｜最大値−最小値｜の 1/5 以上の単調増加／減少が見られる期間
・山形（PEAK）／谷形（BOTTOM）：その動向情報の｜最大値−最小値｜の 1/10 以上の単調増加／減少が見られた後，減少／増加に転じた期間

基本検索タスクの検討にあたっては，既存検索エンジンを用いて行われる対話的情報アクセスについて調査する事を目的とした予備実験を行った．一般に，検索エンジンを用いて行われる検索の意図には，Informational, Navigational, Transactional があることが知られているが，本研究では「ある画像が撮影された場所を既存検索エンジンのキーワード検索のみで探してもらう課題」を行ってもらうことで，図 1 に示すように，より詳細な検索意図の分類を定めた．新たな情報を得るために行う Discover と，情報の確認を目的とする Verify に大別し，それらを特定アイテムを指定して関連動向情報を検索する Navigational，アイテムを限定しない Informational な検索に分類する．さらに，特徴的変動を指定して行う検索を

Pinpoint，指定しない場合を Broad とし，異なる複数の状態が検索条件に該当しうる場合を Multiple，一意に定まる場合を Single とする．Multiple か Single かは特徴的変動に対応して決まり，急上昇／急下降や山形／谷形は Multiple，最大値／最小値は Single となる．

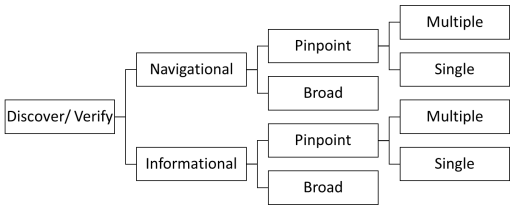

図 1. 検索意図の分類

これらの検索意図に対応するため，以下の 3 つの基本検索機能を提案する．基本検索機能 (1) は Navigational に対応し，(2) は Informational に対応する．Pinpoint, Broad は変動タイプを指定するか否かに対応する．また，動向に関する問いの中で，同時期に流行したアイテムの検索は需要が多いと考え，基本検索機能 (3) を定める．これは検索結果として任意アイテムを期待するため，Informatoinal に対応するとみなせる．

(1) 指定したアイテムに関する動向が特徴的変動を示した期間の検索
(2) 指定した期間に特徴的変動を示したアイテム・動向の検索
(3) 指定したアイテムに関する動向が特徴的変動を示した期間に同様の変動を示したアイテム・動向の検索

上記基本検索機能を備えた検索エンジンのプロトタイプを開発した．プロトタイプシステムは Web ブラウザだけでなく，API を利用して任意のアプリケーションからも利用可能とすることで，インタラクション設計に関する研究にも活用している．図 2 は，Web ブラウザでアクセスした場合のプロトタイプインタフェースのスクリーンショットである．図の上部は入力フォーム部分に「 2011/03-2011/12 の間に最大値をとったアイテムの検索」(基本検索機能 (2)) をクエリとして入力した例であり，その検索結果を図の下部に示している．検索結果として，アイテム名，統計データ名，クエリを満たす期間，該当地域についての情報が返される．検索結果画面において，表示されるアイテム名（左端）をクリックすることで，その動向の折れ線グラフが表示される．

提案する検索エンジンは，「基本検索機能の組み合わせで，多様な情報要求に対応することができる」という既存検索エンジンの利点を継承しつつ，タスクを動向に関する問いに限定することでより高度な基本検索機能を提供する点に特徴がある．この点について

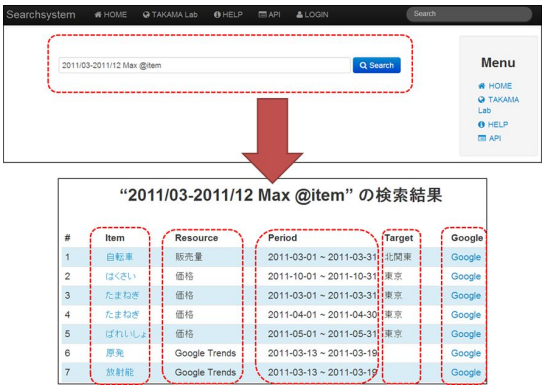

図 2. プロトタイプインタフェースのスクリーンショット：(上) クエリ入力フォーム，(下) 検索結果画面

有効性を示すために，開発したプロトタイプシステムを用いて評価実験を行った．情報要求に対して適切なクエリを表現できることを検証するために，以下に示す検索課題 8 種類を用意し，12 名の工学系大学生／大学院生に解答してもらった．

(1) 自転車に関する動向がピーク（山形変動）を迎えた期間は？
検索意図：Navigational-Pinpoint-Multiple
クエリの例：[自転車 PEAK @period]
(2) 2008 年 5 月～2008 年 12 月に谷を迎えたアイテムは？
検索意図：Informational-Pinpoint-Multiple
クエリの例：[2008/05-12 BOTTOM @item]
(3) 自転車に関する動向が 2008 年～2010 年の中で特徴的な変動を迎えた期間は？
検索意図：Navigational-Broad
クエリの例：[自転車 2008/01-2010/12 BROAD @period]
(4) iPad の（Web 上でのユーザ活動に関連する）動向が最小値を迎えた期間は？
検索意図：Navigational-Pinpoint-Single
クエリの例：[iPad !S MIN @period]
(5) iPad のヒット数が最大である期間に，同じ変動をしたアイテムは？
検索意図：Informational-Pinpoint-Single
クエリの例：[iPad S+ヒット数 MAX @item]
(6) 2005 年中に特徴的変動をしたアイテムは？
検索意図：Informational-Broad
クエリの例：[2005 BROAD @item]
(7) 安倍内閣の内閣支持率が最大である期間に，最小値を迎えたアイテムは？
検索意図：Navigational-Pinpoint-Single; Informational-Pinpoint-Single
クエリの例：[安倍内閣 MAX @period]; [2007/08-09 MIN @item]
(8) 2004 年に急上昇を迎えたアイテムの中で，2008 年にも急上昇したアイテムは？
検索意図：
Informational-Pinpoint-Multiple; Informational-Pinpoint-Multiple
クエリの例：[2004 SI @item]; [2008 SI

@item]

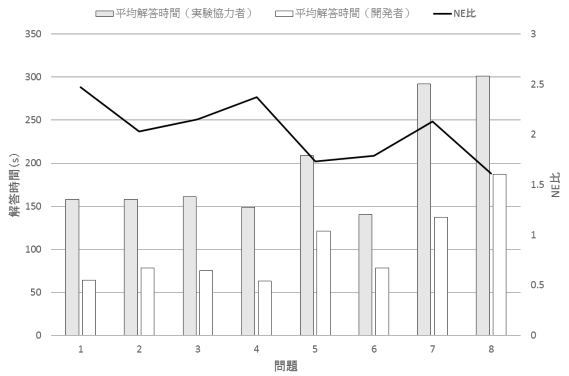

図 3. 解答時間と NEM 比

問題毎の実験協力者の平均解答時間，開発者の解答時間，および NEM 比（開発者とユーザの操作時間の比率）を図 3 に示す．解答時間を棒グラフ（左側の軸），NEM 比を折れ線グラフ（右側の軸）で示している．NEM 比は最小で 1.61（問題 8），最大で 2.47（問題 1）となった．また，いくつかの問題においては，設計者と同等の時間で解答クエリ送信を行う実験協力者も複数名存在した．これらの結果は，既存検索エンジンを利用した先行研究と同等のものである．また，入力されたクエリの内，正しい構文に従ったものの割合は全実験協力者平均で 79.2%であり，代表的な商用検索サービスと比較しても十分高い数字となっている．従って，プロトタイプシステムを用いて適切なクエリを十分効率的に生成可能といえる．

多様な検索意図に基づく検索が実行されるかについても評価を行った．実験協力者各自が任意に選択した期間の中でインパクトが大きかったモノや出来事ベスト 3 を選び，順位付けはせずに回答してもらった．プロトタイプシステムの他，既存検索エンジンを補足的に利用しても構わないとした．

回答例を以下に示す．括弧内は各実験協力者が設定した期間である．各自，自身の高校時代や学部時代などを振り返り，当時関心のあったことを検索を通じて確認し，回答していた．

・「民主党政権誕生，ルンバ発売，ユニクロ海外出店」（2008-2010）
・「東日本大震災，なでしこジャパン W 杯優勝，新型インフルエンザ流行」（2009-2011）

実験協力者の情報アクセスパターンについて分析するために，入力されたクエリについて，検索意図を図 1 に従い分類した結果，全実験協力者が Informational な検索は 1 回以上行っていたが，Navigational な検索を一切使用しない協力者が 4 名，Informational よりも多用した協力者が 3 名いた．これらより，既存検索エンジンを用いた場合と同様に，利用者毎に異なる検索行動が見られ，各自複数の検索意図を組み合わせてプロトタイプシステムを利用することを確認した．以上より，多様な情報検索要求を基本検索機能に分解して満たすことができるという，既存検索エンジンの重要な特徴を継承していると考える．

コンテクスト検索エンジンの活用が期待できる応用の一つとして，データセット間の関係性を発見するタスクを想定し，有用性についての考察を行った．データ市場においてやりとりされるデータセットの中には内容を公開できないものも存在するため，内容を公開することなく，その価値を見積もることを可能とするためにデータジャケットの概念が提案されている．データジャケットはデータセットの変数名といったメタデータや概要を記述したものであり，これを利用することで価値を生み出すデータセットの組合せなどを検討する．IMDJ（Innovators Marketplace on Data Jackets)ではデータジャケットを利用し，市場の多様な利害関係者がワークショップ形式で議論を通じながら自身の問題解決に繋がるデータセットの組合せを発見する．複数のデータセットを組み合わせることを考えた場合，組合せ可能なデータセット間には何らかの関連するデータ・情報が必要である．データセットの一般的な形式として表形式を想定した場合，以下の二種類の関連性が考えられる．

・行の関連性：関連するインスタンスの存在
・列の関連性：関連する属性の存在

一般に，表形式のデータの場合は商品や顧客といったインスタンスが行に相当し，列に属性を配置する．行の関連性は，２つのデータセットが同一のインスタンスに関するデータを含んでいる場合などに相当する．LOD において，共通するリソースを含む RDF トリプルを接続する場合などがこの一例である．列における共通性とは，２つのデータセット間で共通，あるいは関連する属性が存在する場合であり，データジャケットはこの種の関連性の発見に有効と考える．また，近年 LOD の活用例として多く見られるような，地図上に複数データセットからの情報をマッピングするサービスなども，このタイプに該当する．

これらのアプローチとは異なり，コンテクスト検索エンジンでは動向情報の関連性の観点から，データセット間の関係を見いだすことが可能である．同時期に流行したなどの関係性は，時系列性のあるデータセットで，収集期間にオーバラップがあれば計算可能であるため，より多様なデータセット間の関係性発見に貢献することが期待できる．

適用例として，2011 年 3 月から 12 月の間に動向情報が最大値を迎えたアイテムをプ

ロトタイプシステムで検索した場合，原発の検索数などが検索された．当該期間は東日本大震災直後であるため妥当な結果と言えるが，自転車の販売量も検索された．当時のニュース記事などを確認したところ，交通機関が止まった場合の交通手段や，省エネのために自転車を購入する人が増加しており，それが反映した結果と言える．原発と自転車の間には一見関係はないように考えられるが，動向を切り口とすることで異なるアイテム間の関係性が発見できたといえる．

別の例として，インフルエンザと同時期に動向情報が急上昇するアイテムを検索したところ，空気清浄機が検索された．これは，空気清浄機の高機能なものには，インフルエンザへの効果をうたったものがあることに対応している．この例も，関係が必ずしも自明とは言えないアイテム間の関係を発見可能なことを示しており，コンテクスト検索エンジンの検索結果を手がかりとして，両アイテムに関係するデータセットの組合せなどを検討することに繋がることが期待できる．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 4 件）

① 高間康史，加藤優，桑折章吾，石川博，動向に関する問いを対象とした検索エンジンの提案，人工知能学会論文誌，Vol. 30, pp. 138-147, 2015.（査読有）
DOI: 10.1527/tjsai30.138

② 松下光範，探索的検索－情報へのアプローチの変化，薬学図書館，Vol. 60, pp. 64-70, 2015.（査読無）

③ Yasufumi Takama, Kohei Ishiguro, Support of Exploratory Analysis of Exchange Rate Data Based on Context Search and Granularity-dependent Similarity Calculation of Temporal Data, International Journal of Affective Engineering, Vol. 13, pp. 235-244, 2014.（査読有）
DOI: 10.5057/ijae.13.235

④ Yasufumi Takama, Masaki Okumura, Interactive Visualization System for Monitoring Support Targeting Multiple BBS Threads, International Journal on Intelligent Decision Technologies, Pre-Press, 2014.（査読有）
DOI: 10.3233/IDT-140232

〔学会発表〕（計 13 件）

① Yanjun Zhu, Yasufumi Takama, Yu Kato, Shogo Kori, Hiroshi Ishikawa, Introduction of Search Engine Focusing on Trend-related Queries to Market of Data, MoDAT2014, 2014/12/14, 深セン（中国）

② Yasufumi Takama, Koichi Tashiro, Proposal of Support Tools for Analyzing RDF Database Using TETDM, SCIS&ISIS2014, 2014/12/3-6, 北九州国際会議場（福岡・北九州市）

③ Sayuri Anbe, Ichiro Kobayashi, An Approach to Category Classification of Cosmetics Reviews based on Brand Names, SCIS&ISIS2014, 2014/12/3-6, 北九州国際会議場（福岡・北九州市）

④ Naoya Otsuka, Mitsunori Matsushita, Constructing Knowledge Using Exploratory Text Mining, SCIS&ISIS2014, 2014/12/3-6, 北九州国際会議場（福岡・北九州市）

⑤ Ryo Yamashita, Mitsunori Matsushita, Content Discrimination of Comics Based on Users' Reviews, 3rd Asian Conference on Information Systems, 2014/12/1-3, Nha Trang（ベトナム）

⑥ 高間康史，Chengzhu Yin，桑折章吾，山口晃一，動向に関する問いに答えるコンテクスト検索エンジンのデータ市場への応用に関する検討，人工知能と知識処理研究会，2014/11/29，九州工業大学サテライト福岡天神（福岡・福岡市）

⑦ Chengzhu Yin, Hiroshi Ishikawa, Yasufumi Takama, Proposal of Time Series Data Retrieval with User Feedback, GrC2014, 2014/10/22-24，登別グランドホテル（北海道・登別市）

６．研究組織
(1)研究代表者
高間　康史（TAKAMA, Yasufumi）
首都大学東京・システムデザイン研究科・教授
研究者番号：20313364

(2)研究分担者
小林　一郎（KOBAYASHI, Ichiro）
お茶の水女子大学・人間文化創成科学研究科・教授
研究者番号： 60281440

松下　光範（MATSUSHITA, Mitsunori）
関西大学・総合情報学部・教授
研究者番号：50396123