

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 1 日現在

機関番号：12102

研究種目：若手研究(B)

研究期間：2012～2014

課題番号：24700239

研究課題名(和文) 未来情報の探索研究における評価基盤の構築

研究課題名(英文) Test Collection for Information Retrieval Research on Future Information

研究代表者

上保 秀夫 (Joho, Hideo)

筑波大学・図書館情報メディア系・准教授

研究者番号：00571184

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：過去や現在に関する情報に焦点を当ててきた従来の情報検索研究に対し、本研究では未来に関する情報の検索技術および検索行動に焦点を当て、その技術評価基盤の構築を行った。その結果、未来情報に関する検索行動が、過去や現在の検索行動と比較して、どのように異なるのか明らかになった。また、未来情報の検索技術を評価することが可能な研究用データセットを構築し公開した結果、米国、英国、フランス、ドイツ、インド、中国、日本の研究グループに活用され、技術開発に貢献した。

研究成果の概要(英文)：While existing research on Information Retrieval tend to focus on past or recent information, this research aimed to build an evaluation infrastructure for search technologies of future-related information retrieval. Based on a series of user studies, this research allowed us to identify the differences and similarities between future-related information searching behaviour and that of past or recency-related behaviour. In addition, we built and released a test collection that allowed researchers to evaluate the performance of future-related information retrieval systems. The test collection was used by the research groups in U.S., UK, France, Germany, India, China, and Japan, and helped their research and development on new search technologies.

研究分野：情報検索

キーワード：情報検索 未来情報 テストコレクション

1. 研究開始当初の背景

サーチエンジンを含む情報検索システムは人間が発信した情報をその検索対象としているが、人間が発信する情報は概念の定義のように時制の影響が少ないものを除くと、多くの場合「過去」に関するものである。これは人類が過去の経験や記録を編纂して様々な知識を構築してきたことを考えれば不思議なことではない。したがって情報検索システムが対象とするコーパスは主に過去の情報が蓄積されたものであると言える。その結果として情報検索分野が取り扱う利用者の情報要求も多くの場合、過去の情報をその主な対象としてきており、情報検索システムの評価基盤であるテストコレクション（詳細は後述）も同様の性質を持っていると考えられる。例外として **Topic Detection and Tracking** があり、これは随時発信される新聞記事などからあるトピックに関する新規情報を含む文書のみを検出する技術であり、「現在」に焦点をあてた情報検索技術である。

本研究では情報探索における「未来」の時制に着目した。情報検索分野で使用されているコーパスにも、イベントの告知や書籍の近刊案内、組織や個人の将来計画など発信された時点では未来となる情報は含まれている。しかしこれまで、非常に少数の例外を除いて、この「未来」に関する情報に焦点を当てた包括的な研究は非常に限られている。一方で、近年のツイッターや Facebook などのソーシャルメディアの浸透により、未来情報の発信量が増加しており重要な研究テーマになっている。

未来を予測する研究はこれまでも行われてきた。株価や為替の変動を過去のデータを分析するものや、ツイッターなどに出現するキーワードの頻度の突起度を計測し、流行の兆しを検出する技術などがその例である。しかしこれらはあくまで過去と現在の分析を基にして未来を予測するのが目的であって、未来に関する情報そのものを検索していない。また予測する対象があらかじめ定義されていることが多く、情報検索システムのように利用者の情報要求があらかじめ規定できない場合には必ずしも有効的であるとは言えない。一方で、未来情報への効果的なアクセスは、あらゆる分野における組織や個人が短・中・長期的展望を見据えて行動することを可能にするため、重要な研究課題である。

2. 研究の目的

以上のような学術的背景を元に、本研究は、未来情報を含む時間的要素を考慮した検索に着目しており、特に(1) ユーザの情報検索行動の解明と(2)情報検索システムの評価を可能にする評価基盤の構築を目的とした。

3. 研究の方法

(1)の目的である、ユーザの未来情報検索行動の解明は①意識調査と②行動分析を元に行った。①意識調査はオンラインアンケートを用いて、20代から60代までの男女を対象に行った。アンケートは8つの選択式質問と3つの自由回答式質問で構成されており、回答者の直近のサーチエンジン利用体験について、その時間的要素を中心に調査することを目的とした。質問内容の例として、探していた情報の対象時間（過去の情報、今日の情報、未来の情報、等）、探していた情報の時間的性質（長期的な興味関心、定期的に繰り返される興味関心、最近の興味関心、等）、探していた情報の利用予定（検索後すぐ、その日中、数日中、予定なし、等）などが含まれる。

②行動分析は実験室型実験を用いて、大学生および大学院生を対象に行った。実験参加者には、図1のような検索課題の背景情報と検索内容（未来情報のみを掲載）を記述した文書を提示し、15分間自由にインターネット上で情報検索を行ってもらった。ただし、Facebookなどのソーシャルメディアを使って知人に直接尋ねることは禁止した。また、適合したウェブページを見つけた場合は、ブックマークをするように指示を与えた。

【検索作業の背景となるストーリー】

あなたが使ってきた携帯の調子が最近悪くなってきたので、スマートフォンに変えることを検討しています。そこで少し調べてみると、アップル社の iPhone の他に、複数の会社がアンドロイド (Android) というシステムを使ったスマートフォンを出していることがわかりました。そこで、アンドロイドを使ったスマートフォン (アンドロイドフォン) に心が傾いているのですが、まだ本当にそれでよいのか確信が持てないので、より詳しく調べてみることにしました。

【検索作業の情報ニーズ (未来)】

本当に今が買い時なのか判断したいので、アンドロイドフォンの関連技術が今後どのように発展する予定なのか、ロードマップが知りたい。どのような技術をいつ頃をめどにリリースしようとしているのかを具体的に説明しているページを探してください。

図1: 検索課題例

データ収集には、検索作業で使用するパソコンのウェブブラウザに搭載したログ記録機能による検索ログ（入力した検索質問（クエリ）、クリックした URL など）、作業後のアンケート、保存された文書の適合性情報、などが含まれた。

(2)の目的である評価基盤の構築は、テストコレクションと呼ばれる、特定の技術課題に焦点を当てた情報検索システム評価用データセットを構築することで実現した。テストコレクションは、ユーザの情報ニーズ、文書集合、そして情報ニーズごとの文書適合性情報で成り立っている。本研究では、①ユーザのクエリに含まれる時間的検索意図を検出する課題と、②未来を含む時間的要素を考慮した検索意図に対する順位付け課題の二つを設定し、テストコレクションの構築を行った。

①クエリの意図解析では、未来情報、過去情報、最新情報、非時間的情報という4つの区分に対し、それぞれ75クエリ、計300クエリをサンプリングした。クエリの利用日時は、2013年5月1日と設定した。例えば、「Disney prices 2014」や「long term weather forecast」は未来情報の検索意図があるクエリの例である。

②時間的文書検索では、大きな検索トピックを背景に4つの時間区分に焦点を当てた情報ニーズを定義する手法を考案した。本研究では、50の大きな検索トピックを基に、それぞれの情報ニーズを定義したため、計200の時間的検索トピックを構築することになった。構築されたトピックの例を以下に示す。

Title	Girl with the Dragon Tattoo
Description	I've recently watched a film called Girl with the Dragon Tattoo, and really liked it. Therefore, I would like to gather information about the movie.
Past question	How did the casting of the film develop?
Recency question	What did the recent reviews say about the film?
Future question	Is there any plan about its sequel?
Atemporal question	What are the names of main actors and actresses of the film?
Search date	28 Feb 2013 GMT+0:00

図2: 検索トピック例

4. 研究成果

研究の主な成果

本研究の成果は大きく分けて二つある。1つは、未来情報を探すサーチエンジンユーザの検索行動を、過去や現在など他の時間区分と比較して明らかにしたことである。これは、アンケートを用いた意識調査と実験室実験を用いた行動分析からなる。以下にそれぞれに結果を概観する。

図2は、回答者から得た情報ニーズの時間軸分布を示したものである。図からわかる通り、多くの検索は現在の情報を得るために行われている。一方で、未来(右方向)や過去(左方向)の情報を得る検索も一定の割合で行われていることが明らかになった。

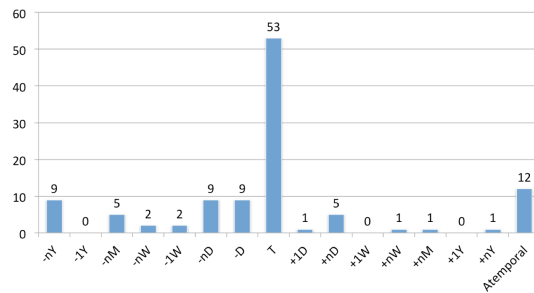


図3: 情報ニーズの時間分布

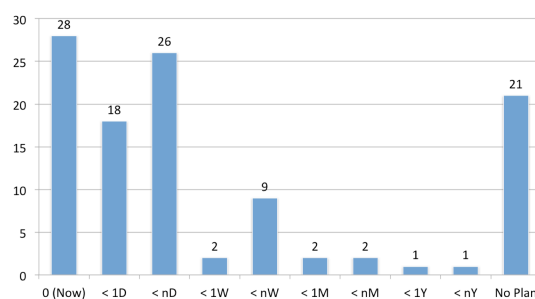


図4: 検索情報の利用機会

図3は検索によって得た情報を使う機会の時間軸分布を示したものである。左端がすぐに使う割合で多くなっているが、数日中またはそれ以降に使う予定の割合も多い。このことから情報を得たタイミングと使う時期には隔たりがあることがわかった。これは、得た情報の内容が未来において変化する場合、問題となる可能性を示唆している。本調査の詳細は、学会発表③で報告している。

実験室実験による行動分析からは以下の結果が得られた。(1)ユーザは「動向」や「将来」などの時間表現を主題を表す語に追加することで時間的検索を制御しようとする。(2)しかし、時間表現で拡張されたクエリの効果にはばらつきがあり、検索トピックに依存している。(3)検索エンジンは時間的検索意図を検出し、検索結果の時間的制御を可能にする検索オプションを表示する必要がある。(4)現状の検索エンジンのパフォーマンスは、最新情報の検索に対しては良いが、未来と過去の情報に関しては向上の余地がある。(5)ユーザは一般的に未来に関する検索を最も難しいと感じている。本調査の詳細は、雑誌論文①で報告している。

もう1つの成果は、未来情報を含む時間軸を考慮した検索アルゴリズムの開発評価を可能にする評価基盤(テストコレクション)を構築・公開したことである。テストコレクションは、第11回NTCIRワークショップにおけるTemporal Information Access (Temporalia)タスクとして公開された。その結果、国内外から9つの研究グループが

Temporalia タスクに参加し、それぞれの検索アルゴリズムを開発した。テストコレクションを用いた開発システムの評価分析の結果、以下のことが明らかになった。(1) クエリの時間的検索意図検出に関しては、最新情報の意図検出が難しく、過去情報の意図検出は比較的容易である。(2) 同一のアルゴリズムで全ての時間区分の検出を効果的に行うのは現状の技術では困難である。(3) 未来情報意図のクエリと最新情報意図のクエリの区別は難しい。(4) 時間的検索に関しては、クエリ意図解析同様に、同一のアルゴリズムで全ての時間区分の検索を効果的に行うのは現状の技術では困難である。(5) 検索が困難な時間区分は元の大きな検索トピックに依存している場合がある。Temporalia テストコレクションの詳細は、学会発表①②で報告している。

得られた成果の国内外における位置づけとインパクト

サーチエンジンのアルゴリズム開発やユーザの検索行動理解において、時間的要素に着目する研究が近年増えている。しかしそれらは多くの場合、最新情報の検索に焦点を当てている。そのような状況で、未来という時間軸を考慮したユーザ行動研究と検索技術評価基盤の構築は、新しい貢献と位置づけることができる。本研究の成果は、国際論文誌 1 件と国際会議およびワークショップ 3 件によって公表されている。また、構築したテストコレクションは、NTCIR-11 Temporal Information Access (Temporalia) タスクとして第 11 回 NTCIR (引用文献[1]) に採択され、研究目的で公開された。Temporalia テストコレクションは、これまで米国、英国、フランス、ドイツ、インド、中国、日本など国内外の研究グループに利用されており、新しい検索アルゴリズム開発に貢献している。

今後の展望

未来情報を含む時間的要素を考慮したテストコレクションの構築は、今後とも高度化させていくことが考えられる。その試みとして、Temporalia タスクは、第 12 回 NTCIR においても採択され、より高度なテストコレクションの構築作業に着手している。具体的には、クエリの時間的意図分布解析や検索結果の時間的多様化、などの技術的課題に取り組むものである。

引用文献

[1] Joho, H. and Kishida, K. (2014) "Overview of NTCIR-11". In: Proceedings of the NTCIR-11 Conference, pp. 1-8, Tokyo, Japan.

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 1 件)

① Joho, H., Jatowt, A., and Blanco, R. (2015) "Temporal Information Searching Behaviour and Strategies", Information Processing and Management, 印刷中. 査読有 doi:10.1016/j.ipm.2015.03.006

[学会発表] (計 3 件)

① Joho, H., Jatowt, A., Blanco, R., Naka, H., and Yamamoto, S. (2014) "Overview of NTCIR-11 Temporal Information Access (Temporalia) Task". In: Proceedings of the NTCIR-11 Conference, pp.429-437, NII, Chiyoda-Ku, Tokyo, Japan. 2014.12.10. <http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings11/pdf/NTCIR/OVERVIEW/01-NTCIR11-OV-TEMPORALIA-JohoH.pdf>

② Joho, H., Jatowt, A., and Blanco, R. (2014) "NTCIR Temporalia: A Test Collection for Temporal Information Access Research". In: Proceedings of the TempWeb 2014 Workshop, WWW 2014, 845-849, Seoul, Korea. 2014.04.08. doi: 10.1145/2567948.2579044

③ Joho, H., Jatowt, A., and Blanco, R. (2013) "A Survey of Temporal Web Search Experience". In: Proceedings of the TempWeb 2013 Workshop, WWW 2013, pp.1101-1108, Rio de Janeiro, Brazil. 2013.05.13. <http://dl.acm.org/citation.cfm?id=2488126>

[その他]

ホームページ等

<http://ntcirtemporalia.github.io>

6. 研究組織

(1) 研究代表者

上保 秀夫 (JOHO HIDEO)

筑波大学・図書館情報メディア系・准教授
研究者番号：00571184