

ユーザの意思を反映した Web 検索の効率化

Efficient Web Retrieval Based on User's Intention

大石 哲也 峯 恒憲 長谷川 隆三 藤田 博 越村 三幸
 Tetsuya Oishi Tsunenori Mine Ryuzo Hasegawa Hiroshi Fujita Miyuki Koshimura

1. はじめに

近年、多くの人が、必要な情報を得るために Web 検索を利用している。しかしながら、現在世の中に存在している Web ページはすべてを把握できないほど多くなっているため、ユーザが必要とする Web ページをなかなか見つけられないこともしばしばある。

そのためユーザの必要とする情報を把握するために様々な研究が行われてきた。例えば、膨大な Web ページの中から必要な Web ページを得るために、ユーザのブックマークを利用した研究(e.g.[7])や、複数のユーザ間で Web ページを推薦しあう研究(e.g.[4])、Web 検索エンジンへの問合せを拡張する研究(e.g.[5])、Web 探索経路選択の研究(e.g.[6])などが行われている。さらには、ユーザに関するあらゆる情報をクライアントマシンから抽出し、個人に特化した Web 検索を行う研究(e.g.[1])も行われており、ユーザの趣向を考慮した Web 検索を実現している。また、Web 検索の効率化のために関連する検索ワードを提案するシステム(e.g.[2])が実用化されている。しかし、これらの研究は時間を考慮に入れた Web 検索を実現していない。

ユビキタス時代といわれる現代において、ユーザの視点や状況を基にタイムリーな情報をユーザに推薦できる技術が求められる。ユーザの趣向を判断する材料として、電子メールや Web 閲覧履歴など様々な存在するが、本研究では時間要素を含むスケジュールに着目した。スケジュールにはユーザの趣向と、ユーザが情報を必要としているタイミングが含まれている。そのスケジュールの特性を用い、ユーザにとってタイムリーな情報を含んだ Web ページを推薦することが可能になると考えた。

そこで、本稿では、スケジュールを用いた Web 検索システムを提案する。以下、2章でシステムの概要について述べ、3章でシステムが妥当であるかを調査するための予備実験について述べる。最後に、4章で結論を述べる。

2. システムの概要

2.1 コンセプト

本研究では、ユーザの状況を把握しタイムリーな情報をユーザに推薦する手法の開発を目的とする。状況を把握する手段として様々な方法が存在するが、本研究ではスケジュールを用いてユーザの状況を把握し、その状況に応じて Web 検索を行いユーザに情報を推薦する。

ユーザによって情報を必要とする状況は異なる。人によって予定の何日も前から情報を必要としたり、直前に情報を必要としたりするかもしれない。また、以前必要と思っていた情報が再度必要になる状況もある。このような状況を把握するためには、ユーザの個性や時間の概念を考慮することが必要であるが、本研究ではまず、ユーザの状況を時間に応じて判断することを考え、そのための手段としてスケジュールを利用することとした。

スケジュールの利点は、現在だけでなく、過去や未来の情報を利用することができることである。ユーザが現在必要としている情報を過去の情報から推薦することが可能になる。また、未来の情報を利用することで、ユーザが今後どのような状況になるのかを推測することができ、ユーザにとって必要な情報を事前に推薦することも可能になる。

例えば、同じ議題に対する会議が定期的にスケジュールに組み込まれているとする。初期段階では、詳細にスケジュールや議事録(スケジュール実施後の事後情報)を入力するが、回を重ねる毎に簡略化して入力すると考えられる。このように、もしスケジュールが簡略化して入力されていたとしても、過去のスケジュール、議事録、閲覧履歴などの情報から現在のユーザの状況を把握することができる。その結果、ユーザが必要としている Web ページを推薦できるような機能を実現する。また、未来に過去と同じようなスケジュールがあるうとなかろうと事前に Web ページを推薦することも可能になる。

2.2 概要

本研究におけるスケジュールを用いた Web 検索において、Web ページを推薦する流れの例を図.1に示した。

各スケジュールに対する Web ページの推薦の流れは、関連するスケジュールの有無によって異なる。関連するスケジュールがない場合は、ユーザが入力したスケジュールの内容を利用して Web ページを推薦する。関連するスケジュールがある場合は、関連するスケジュールの議事録と閲覧履歴も利用して Web ページを推薦する。

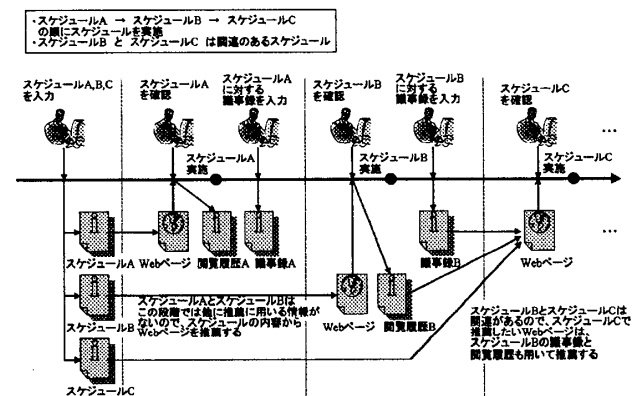


図.1 Web ページ推薦の流れ

以降、本研究においてまず最初に実現しようとするシステムについて、ユーザ側の機能とシステム側の機能に分けて述べる。

2.3 ユーザ側の機能

ユーザは以下を行うことができる。

- スケジュールを入力
- システムが推薦した Web ページを閲覧

● 過去のスケジュールに対して、議事録を入力ユーザがこれらの情報をシステムに与えることで、システムはユーザの趣向にあった Web ページをタイムリーに推薦することができる。

2.4 システム側の機能

システムは以下を行うことができる。

- 現在のスケジュール情報からユーザが必要としていると思われる Web ページを推薦
- 現在のスケジュールと同じようなスケジュールが過去にあり、その議事録が入力されている場合に、議事録情報からも Web ページを推薦
- 推薦する Web ページの中で、過去にシステムが推薦しユーザが閲覧した Web ページがある場合に、その Web ページを優先的に推薦

3. 予備実験

3.1 実験方法

スケジュールを用いた Web 検索の有効性を確かめるため、いくつかの予備実験を行っていく予定である。その実験の概要について述べる。

予備実験では、ユーザが入力したスケジュールを整形された状態(XML 形式)で取得できることを前提とする。スケジュールには【時刻】、【タイトル】、【詳細】(【タイトル】に付け加える内容)、【議事録】が含まれているものとする。これらの情報を利用して、以下のような観点で予備実験を行う。

- (1) 【タイトル】と【詳細】の形態素(形態素解析には MeCab[3]を用いる)を用いて Web 検索を行い、検索結果の内容を以下の場合に分けて調査
 - 形態素をそのまま利用する場合
 - 名詞だけを利用する場合
 - 名詞の中から必要な単語が抽出されていると仮定し、それを利用する場合
- (2) 【タイトル】と【詳細】に、(1)と同様の場合分けをした【議事録】の形態素を加えて Web 検索を行い、検索結果の内容を調査
- (3) 【タイトル】と【詳細】の形態素を用いて Web 検索を行い、検索結果を【閲覧履歴】を用いて優先順位を変更した結果を調査
- (4) 現在時刻を変更しながら、現在時刻に近いスケジュールの【タイトル】と【詳細】の形態素を用いて Web 検索を行い、検索結果の内容を調査

3.2 考察

Web 検索は、検索に用いるキーワードによって検索結果が大きく異なる。そのため、【タイトル】と【詳細】から得られた形態素をそのまま利用するよりは、必要な単語が抽出されていることが好ましい。そのため、形態素の中から必要な単語を抽出する技術が必要である。

【議事録】は過去のあるスケジュールにおいて、そのスケジュールが入力された段階よりも知識がある状態で入力された情報である。【議事録】に URL や固有名詞などの情報が含まれていると、そのスケジュールに類似した別のスケジュールに対して、よりよい Web 検索の結果を推薦することが可能になる。また、【閲覧履歴】は実際にユーザが閲覧したという事実を反映するので、ユーザの趣向を反映した Web 検索が可能になる。

スケジュールには【時刻】が含まれるので、現在時刻にスケジュールの【時刻】が差し迫った時に、ちょうどよく Web ページを推薦することが可能になる。

以上の考察により、スケジュールを用いて Web 検索を行うことで、ユーザの視点や状況に応じたタイムリーな情報を与えることが可能になる。

もし、時間を考慮しない場合、つまり、【議事録】と【閲覧履歴】を用いない場合、同様のスケジュールが繰り返し入力されていると、推薦する Web ページが毎回ほぼ同じになってしまい、ユーザにとってタイムリーな情報を推薦できないことがある。この点から、時間を考慮することが重要であることがわかる。

4. 結論

本研究では、ユーザの視点や状況を基にタイムリーな情報をユーザに推薦するためにスケジュールを用いることが有効であるかどうかを簡単な予備実験を行って調査する。考察程度であるが、スケジュールを用いることでタイムリーな情報を与えることが可能になるという感触は得られた。

本研究は、スケジュールを用いることで、他の研究と違って、時間軸の概念を導入する。ある瞬間にユーザが必要としている情報をユーザに推薦できる技術は、これからの時代に求められる技術であり、非常に重要である。しかし、現段階ではまだ実用的なレベルには達していない。今後は、スケジュールからどのようにユーザの趣向を抽出するか、議事録をどのように有効に利用するか、ユーザにタイムリーな情報を推薦するための適切なタイミング、など細かい点に留意して研究を進めたい。また、過去のスケジュールを用いて、未来のスケジュールを再入力することなく Web ページを推薦できるような技術の研究も進めたい。

参考文献

- [1] Jaime Teevan, Susan T. Dumais, Eric Horvitz: "Personalizing Search via Automated Analysis of Interests and Activities", ACM SIGIR, pp.449-456, 2005
- [2] Yahoo!JAPAN, <http://www.yahoo.co.jp/>
- [3] "高速形態素解析システム「MeCab」", <http://mecab.sourceforge.jp/>
- [4] 亀井剛次, 湯川高志, 吉田仙, 桑原和宏: "パーソナルレポジトリ間の協調情報検索", 人工知能学会論文誌 19 巻 4 号 SP-C, pp.292-299, 2004
- [5] 岡部正幸, 山田誠二: "トランスダクティブ学習による最小文書判定からのクエリ拡張", 人工知能学会論文誌 21 巻 4 号 I, pp.398-405, 2006
- [6] 長谷川忍, 柏原昭博: "ハイパー空間における適応的ナビゲーションプランニング支援", 人工知能学会論文誌 21 巻 4 号 I, pp.406-416, 2006
- [7] 佐野浩司, 佐山弘樹: "BisNet: Web ブラウザのブックマーク機能を利用した情報共有システム", 人工知能学会論文誌 20 巻 4 号 B, pp.281-288, 2005