

PC内の操作履歴を用いた Web 閲覧行動の分析

Analysis for web browsing using local event logs

金澤功尚† 森田哲之‡ 田中明通‡ 加藤泰久‡ 高橋時市郎†
 Katsuhisa Kanazawa Tetsuhisa Morita Akimichi Tanaka Yasuhisa Kato Tokiichiro Takahashi

1. はじめに

今日、検索エンジンを調査ツールのひとつとして用いることは一般的になっている。しかし、再調査のために前回の調査過程をトレースしようとする、その過程を思い出せずトレース出来ないことはしばしば起こる。

Google Desktop[1]はキーワードを含む個々の Web ページを再発見することを支援している。しかし過去に行った調査行動の概要を思い出すには、履歴を総当りするか、ユーザ自身の記憶に頼るほか無い。また、実際の閲覧では Web ブラウザを複数起動して、並行的に Web ページを閲覧する場合があるため、閲覧履歴は時系列に関してより複雑な構造を持っていると考えられる。

本研究では「Web 閲覧履歴は検索サイトによる検索を結節点とする構造をもつ」と仮定し実際の調査活動における閲覧履歴の分析を試みた。本稿では分析によって得られた結果について報告し、それを基に閲覧履歴の要約手法についても考察する。

2. ログ分析

本研究で用いた Web 閲覧履歴は、与えられる課題に対して 30 分間 Web を使って調査する作業の履歴である。課題には以下のようなものがある。

- ・ 東京観光することになった。東京ディズニーランド、ナンジャタウン、のいずれかで遊ぶ計画を立てよ。
- ・ スキー旅行をすることになった。草津、蔵王、菅平のいずれかへ旅行する計画を立てよ。

同様の課題が合計 5 個あり、被験者はそのすべての課題に対して作業する。調査に使用できる検索エンジンは Goo, Google, Yahoo! の三つという制約を設けた。これは、検索エンジンによる検索結果ページを URL から識別するには、検索エンジンの URL を前定義する必要があるためである。尚、閲覧するページに関して制約は無い。

本研究では被験者 20 人に対して上記の作業を行ってもらい、その閲覧履歴 100 件に対して分析を行った。

2.1 ログについて

ブラウザに残る閲覧履歴は単純なアクセス履歴である場合が多いため、「ある時点でユーザがどのページを閲覧していたか」という情報を得ることは難しい。ある URL から次の URL へのアクセスの間は、前者の URL のページを見ていた、と仮定することは出来るが、前述のようにユーザは複数のブラウザを開く場合があり、その期間中ずっとそのページを見ていたとは限らない。

本研究では NTT サイバーソリューション研究所で開発中の行動検索システム「MemoryRetriever」[2]のログ(以下、

MR ログ)を用いた。このシステムの一部にユーザの PC 上での行動——Web 閲覧、メール閲覧、etc...——の履歴を収集する常駐型プログラムがあり、より粒度の細かい情報を得ることが可能である。Web 閲覧に関しては、あるページを表示させたブラウザウィンドウがアクティブになった場合でもログとして残るという利点がある。

2.2 分析手法

Web 閲覧履歴は URL をノード、ある URL から別の URL への移動をエッジとする有向グラフと見なすことが出来る。本研究では一件の課題ごとにこの有向グラフを作成し、観察することでユーザの行動の特徴を分析した。以下このグラフを Web 閲覧グラフと呼ぶ。Web 閲覧グラフでは、ノードにはページタイトル、閲覧時刻をラベルとして表示し、ノードのサイズは閲覧時間に比例して大きくなる。また、検索結果ページはサイズ固定の矩形で示し、それ以外は楕円形で示した。更にノードはホストごとに色分けされる[図 1]。尚、グラフの表示には Graphviz[3]を用いた。

2.3 パターン

観察の結果以下のような行動パターンが見られた。本稿では、頻繁に現れる小さなパターンをプリミティブとして定義する。プリミティブは Web 閲覧グラフの随所に現れるサブグラフ構造のことである。さらに検索サイトによる検索とその検索結果ページでの行動で場合分けし、各プリミティブがどのように出現するかを分析した。

●プリミティブ

A) リカランス(回帰)

あるページから出発して同じ経路、あるいは極短いサイクルを使って戻る構造を持つサブグラフ構造を指す。多くの場合、折り返すノードの滞留時間は長い。

B) スルー(通過)

特定のパスを立ち戻る事なく順に辿るサブグラフ構造を指す。各ノードの滞留時間のばらつきは小さい。

C) スター(結節点)

ある一つノードからリカランス、スルーが多数発生するサブグラフ構造を指す。スターの中心をハブと呼ぶ。ハブとなるノード自体の滞留時間は極短い。

●検索行動の行動での分類

(ア) 検索行動中に少数のサイトしか訪れない場合

検索結果から辿ったページをルートとするツリー構造をなす。多くの場合、ルートはサイトのトップページとなっており、各ブランチはリカランスとなるかスルーとなつて別の構造に接続する。各ブランチを構成するページの内容は似通った内容のノードで構成される。また、各ブランチにおける滞留時間のばらつきは小さい[図 1左]。

(イ) 検索行動中に複数のサイトに訪れる場合

(ア)の場合と異なり、検索結果をルートとするツリー構造となるが、(ア)に比して完全なツリー構造にはならないことが多い。またルート以下の各ブランチは大きなサ

† 東京電機大学大学院工学研究科

‡ NTT サイバーソリューション研究所

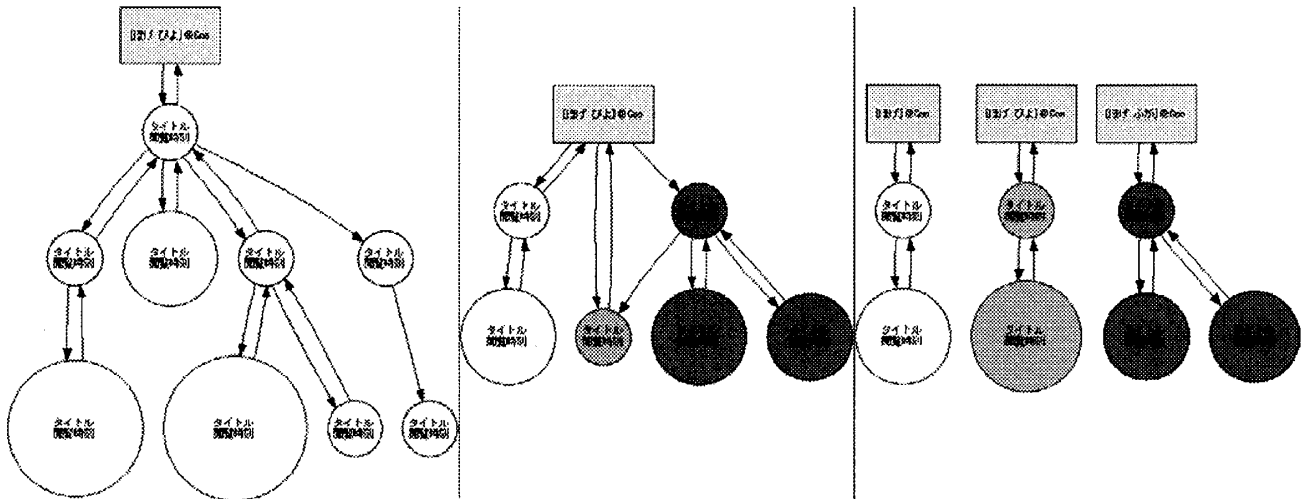


図1 パターン(ア)、パターン(イ)、パターン(ウ)

グラフを構成することは少ない[図1中央]。この場合の各ブランチはリカランスである場合が多い。

(ウ) 多数の検索行動が表れる場合

全体を通して多数の検索を行う場合がある。各検索は(イ)のようになる場合が多い[図2右]。

3. 考察

3.1 パターンに関する考察

2.3 節に示すプリミティブには Web 閲覧における普遍的な行動が現れている。リカランスは、数ページのリンクを辿りブラウザの「戻る」で戻ってくる場合等である。折り返し点で滞留時間が長くなるのも自然なことで、即ち、辿った先のページでユーザの望む情報の有無に関して吟味するために滞留時間が長くなると思われる。

スルーは単純に、辿った先で折り返すことなく、別の行動につながった場合で、リカランスはスルーの特殊な場合であるといえる。スターは閲覧しているサイト構造に依存している。スターのハブはリンク集のようなページになっている場合が多い。

(ア)もサイト構造に依存する。今回の課題では「東京ディズニーランド」等といった、比較的大規模で整理の行き届いているサイトに対する調査が多いため、単一サイト内の行動がきれいなツリー構造をなし、またブランチ内で内容が一定しているものと思われる。サイトによってはこのようなツリーにならない場合も見られた。

(イ)(ウ)では、一つのサイトに訪れる期間、あるいはページ数が少なくなっている。これは検索結果から数ページたどって、直ちに次のページに訪れる為である(所謂ナナム読み)。(ウ)では検索語群の近い検索行動をマージすれば(イ)と等しくなるものと思われる。

3.2 要約手法に関する考察

発見したパターンに基づいて Web 閲覧履歴を要約する手法について考察する。スルー、スター、(イ)におけるツリーの各ブランチを構成するページは似通った内容のページである場合が多いと先に述べた。調査における Web 閲覧履歴では、検索サイトによる検索を中心とした行動はユーザにとっては意味的なまとまりを持っていると考えることが出来る。これらから以下のような要約手法が考えられる。まず、履歴のある検索から次の検索までに見たページ

群ごとに分割する。次に検索に用いた検索語群の近いページ群同士をマージする。この際、あまりにも古い検索行動をマージさせないために、時間的制約を設けるべきである。このページ群をトピックと呼ぶ。次にトピックから前述の Web 閲覧グラフを作成し、密集したハブをルートに持つサブグラフや、多くのノード数を持つスルーといった構造をサブトピックとして分割する。これによって、サブトピックはトピックあるいはサブトピックにネストした構造になる。巨視的にはトピックをノードとする有向グラフとなり、時系列における興味の変遷を追跡することが出来、微視的にはサブトピック、ページをノードとする有向グラフとなり、具体的に閲覧したページを追跡することも出来ると思われる。

4. まとめ

今回の分析では、MR ログから作成した Web 閲覧グラフを観察し、主観的に前述までのようなパターンがあることを見出した。今後は、3.3 節に挙げる行動パターンの統計的な裏づけを行うつもりである。

Web 閲覧においては興味の移り変わり——トピックドリフトが起こるものであるが、今回のログは明確な課題を与えられたものであるためか、それに相当する行動パターンは見出せなかった。また、同じウィンドウ内での移動、ウィンドウをまたがる移動に関する行動パターンを見出せなかった。第1章で述べた様に複数のブラウザによる並行的な閲覧行動も何らかのパターンを持つことが予想される。

今後は、明確な課題を与えられたものではない長期にわたる閲覧履歴も交えて、引き続き分析を続けたい。

参考文献

- [1] Google デスクトップについて
<http://desktop.google.com/ja/about.html>
- [2] T. Morita, T. Hidaka, T. Kura, K. Ooura, Y. Kato, "Desktop search system based on the Action-Oriented algorithm," Proc. APSITT2005, A-5-3, pp.204-207, 2005
- [3] Graphviz <http://www.graphviz.org/>