

携帯情報端末における閲覧パターンを用いた Web サイトの自動ナビゲーション

6Y-02

粥川 隆信* 角谷 和俊** 田中 克己**
 *京都大学工学部 **京都大学情報学研究科

1.はじめに

本研究では、携帯情報端末での Web サイト閲覧において、ある Web サイトでの閲覧履歴を解析して得られる“閲覧パターン”を利用して、未訪問のサイトから、以前に訪れた Web サイトで閲覧した内容と類似する内容を持つ Web ページ群を自動的に収集して提示するシステムを提案する。本研究で提案するシステムは、ユーザーにある程度の Web 閲覧を手動で行わせることで閲覧を望む内容についての情報を受け取り、その情報をもとに未訪問のサイトにおいて望む内容と類似する内容を持つ Web ページを自動的に提示する。

2.動機

携帯情報端末による Web サイト閲覧は、デバイスが小さく操作性が良くないため、多数の Web サイトを閲覧して情報を得る、という目的には向かない。一方で、インターネット上で情報収集を行う場合には複数の Web サイトを閲覧するケースが多く、しかもそれらの Web サイトは互いに類似する内容を持つことが多い。これは複数の Web サイトを閲覧することにより情報の補完あるいは比較をしようという目的があるからである。例えば、

- ・ ニュースサイトの定期的のチェック
- ・ 旅行代理店の Web サイトを閲覧することによるツアーコースの違いのチェック

などである。

そこで、1つの Web サイトを閲覧するだけで、他の類似する内容を持つサイトを自動的に提示するシステムがあれば、デバイスが小さく操作性がよくない携帯情報端末による複数の Web サイトの閲覧を快適に行うことができるのでは、と考えたのである。例えばある新聞社 A の Web サイトにおいて、スポーツカテゴリの野球の x について

ページ、政治カテゴリの小泉首相についてのページを閲覧した後に、別の新聞社 B の Web サイトを指定すると、システムが B 社の Web サイトからスポーツカテゴリの野球の x についてページと政治カテゴリの小泉首相についてのページを探し出して自動的に提示する、などである。この際に問題となるのは、サイトごとに話題のカテゴリ分類やリンク構造などが異なることである。

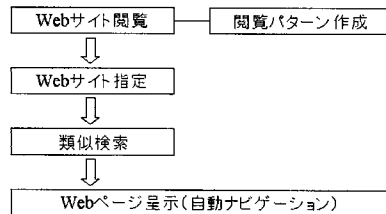


図 1 システム動作の概要

3.閲覧パターンと自動ナビゲーション

閲覧パターンとは、ある Web サイトでの閲覧履歴に含まれる 1つの Web ページについて、それを解析し内容を表現するために作成される“履歴要素”の集合である。

3.1 履歴要素の表現

以下の表現を基本とする。

¥. keystr1. keystr2...keystrN

全ての履歴要素は¥で始まる。¥は Web サイトのトップページを指す履歴要素である。その後 keystr1, keystr2, ..., keystrN とページの内容を端的に表す単語が続く。履歴要素は後に述べるように検索の問い合わせに用いられるため、ページ・サイト特有の単語や固有名詞は取り除かれる場合もある。

3.2 閲覧パターンの生成・表現

閲覧パターンは Web サイトを手動ナビゲーションする過程において生成される。手動ナビゲーション中に訪れた Web ページを逐次解析することで履歴要素を生成し、重複を許さない集合を作る。この集合が閲覧パターンである。閲覧パターンには順序に意味がある場合もあれば意味がない場合もある。なぜなら、異なる Web サイト間では互いに類似する内容のページが存在しても、リン

“Automatic Navigation of Web Sites using Browsing Patterns for Mobile Terminals”

Takanobu Kayugawa*, Kazutoshi Sumiya**,
 Katsumi Tanaka**

*Faculty of Engineering, Kyoto University

**Graduate school of Informatics, Kyoto University

ク構造・ディレクトリ構造・カテゴリ分類などが異なる場合があり、それらの Web サイトにおいて類似する内容を同じ順序で閲覧することが可能とは限らないし、求める情報によっては、例えば旅行代理店のツアーの日程が挙げられるが、Web ページ閲覧の順序が重要な意味を持つ場合があるからである。これは以下のように表現する。

(element1, ..., elementN) (順序に意味有り)

{element1, ..., elementN} (順序に意味無し)

また、一部の履歴要素の集合だけ順序が関係ある場合あるいはその逆、ということもある。その場合、例えば以下のような表現となる。

{element1, (element2, ..., elementX), ..., elementY}

例えば、 $C \rightarrow C1, C \rightarrow C2, C \rightarrow C3, \dots$ というリンクがあって、 $C \rightarrow C1 \rightarrow C \rightarrow C2 \rightarrow C \rightarrow C3 \dots$ というナビゲーションを行った場合に生成されるのは $(Y.C, Y.C.C1, Y.C.C2, Y.C.C3, \dots)$ 、または $(Y.C, (Y.C.C1+Y.C.C2+Y.C.C3+\dots))$ という閲覧パターンのいずれかである。

3.3 検索方法

閲覧パターンに含まれるそれぞれの履歴要素を問い合わせとして検索を行う。検索方法は 2 通りある。Y.A.B という履歴要素と例として説明すると、

- ・ 単語 A と単語 B が共に含まれるページを検索する。
- ・ 単語 A が含まれるページからリンクが張られている、それと同じあるいは下のディレクトリにある単語 B を含むページを検索する。

3.4 自動ナビゲーション

閲覧パターンに含まれる履歴要素を問い合わせとする類似検索を別の Web サイトにおいて行うことで、閲覧パターン作成元の閲覧履歴に含まれるページの内容と類似する内容のページ群を、検索対象となった Web サイトから自動的に収集し呈示することを可能にする。この自動的に Web ページ群を収集し呈示することが「自動ナビゲーション」である。

4. プロトタイプシステム

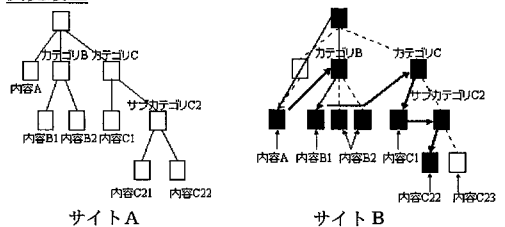
履歴要素を構成する単語は次のようにして Web ページから抽出する。

1. 現在のページにいたる直前のページでクリックされたリンク文字列中の名詞
 2. title タグに囲まれた文字列中の名詞
 3. その他ページ内で重要と考えられる名詞
- 3のような名詞を求めるアルゴリズムは現在検

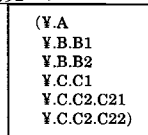
討中であるが、タグ解析や tf/idf 法を応用した方式を考えている。

システムの動作は以下のようなものである。まず Web サイトを手動で閲覧する (サンプルナビゲーション)。サンプルナビゲーションを終えると、その閲覧履歴を元にシステムが閲覧パターンを作成する。この後自動ナビゲーションを行うサイトのトップページを指定すると、システムが閲覧パターンに従って呈示すべきページを収集し自動的にユーザーに見せる。システム動作のイメージを図 2 に示す。

閲覧履歴



閲覧パターン



■ 表示されるページ
□ 表示されないページ

図 2 閲覧パターンを用いたナビゲーション例

5. おわりに

Web サイトの閲覧履歴から閲覧パターンを生成し、それを利用して類似検索をすることで未知の Web サイトの自動ナビゲーションが可能な方式を提案した。

今後、履歴要素の表現及び意味するところの拡張や、履歴要素を構成する単語の抽象化、複数の閲覧パターンをマージして新たな閲覧パターンを構成するなどが考えられる。

謝辞

本研究の一部は、(株) NTT ドコモと京都大学との民間共同研究による。

参考文献

- [1] 奥村 穂高, 田中 二郎, “重ね併せノードによる Web のキーワード検索”, 情報処理学会 第 60 回全国大会, 2000.
- [2] 福島 俊一, 石黒 義英, 喜田 弘司, 山田 洋志, 松田 勝志, “目的特化型 WWW サーチェンジンの開発”, 情報処理学会 第 61 回全国大会, 2000.