

# え～し～あ～る：探している Web ページを過去に閲覧した Web ページの中から推測推薦するシステム

ACR: A recommendation system which suggests web pages from personal web archives

小澤崇記 志築文太郎 田中二郎\*

**Summary.** 我々は、ユーザが探している Web ページを過去に閲覧した Web ページの中から推測し自動で推薦するシステム「え～し～あ～る」を実装した。本システムは、スレッドと閲覧パターンを検出し、今現在閲覧してきたページの閲覧パターンと同一の閲覧パターンを含むスレッドを検出する。その中から、ユーザが満足したページを自動で推薦することによって、ユーザは探しているページをすぐに得ることができるようになる。

## 1 はじめに

Cockburn らの調査 [1] によると、およそ 81% の Web ページ閲覧が「再訪問」である。この再訪問の中には意図せずに行ったものも多く含まれると考えられる。というのも、人は意図せずと同じ行動をとることが多く、このことは Web ページ (以下、ページ) の閲覧においても同様のことが言える。ページを探している最中、もしくは探し出した後に以前同じリンクを辿って閲覧していたことを思い出すことはないだろうか。このような時、過去に閲覧したページを保存し、その中から満足したページを自動推薦することによって、ユーザは探しているページを高い確率ですぐに得ることができるようになると考えた。

この考えに基づき、我々は探しているページを推測し推薦するシステム「え～し～あ～る」を実装した。システムが推測し推薦することで、ユーザはブラウジングの延長で探しているページを得ることができるようになる。

## 2 え～し～あ～る

本システムでは、ユーザが探しているページを 2 つの方法で推測する。現在閲覧中のページ P が再訪問であるならば、過去そのページを閲覧した後にリンクを辿り閲覧したページの内、満足したページを提示する。P が初見であるならば、過去に閲覧したページの中で現在閲覧中のものに類似したページの後にリンクを辿り閲覧したページの内、満足したページを提示する。さらに、後者の精度を上げるために、P を閲覧するまでに辿ったページも履歴を検索するキーとする。

図 1 に本システムによるページ推薦のイメージを

Copyright is held by the author(s).

\* Takaki Ozawa, Buntarou Shizuki and Jiro Tanaka, 筑波大学大学院 システム情報工学研究科 コンピュータサイエンス専攻

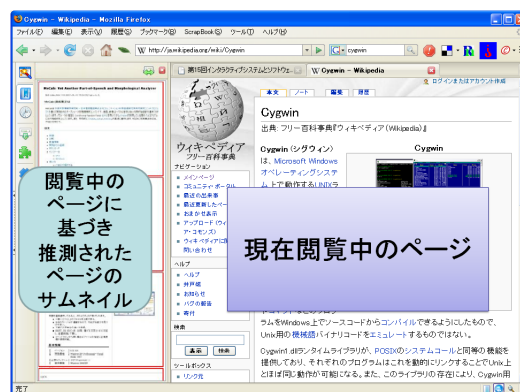


図 1. Web ページの推薦

示す。本システムは、推薦するページを Web ブラウザのサイドバーに提示する。これにより、ユーザは画面を切り替えることなくすぐにその場でページを得られるようになる。また、提示はページの内容が一目でわかるようにサムネイルを表示することで行う。ユーザは、サイドバーに探しているページがある場合にはそれをクリックするだけで閲覧でき、そうでない場合には無視して閲覧を続けられる。

## 3 推測アルゴリズム

### 3.1 スレッドの検出

閲覧履歴の中の Web ページを URL 等を基に解析し閲覧開始ページとそれ以外の 2 種類に分別する。閲覧開始ページとは、アドレスバーに URL を直接入力したり、ブックマークを選択したり等して閲覧したページである。それ以外のページとは、ブラウザの「戻る」ボタン、またはリンクを辿って閲覧したページである。図 2a は、ユーザが閲覧した全てのページ (閲覧履歴) とする。この閲覧履歴を 2 種類に分別した結果が図 2b である。黒のページが閲

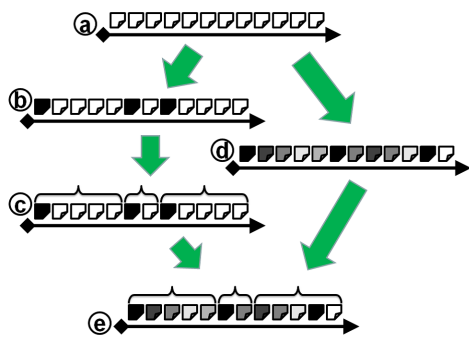


図 2. スレッドと閲覧パターンの解析

閲覧開始ページであり、白のページはそれ以外のページである。

閲覧開始ページから次の閲覧開始ページまでをスレッドと呼ぶことにする。図 2b では、黒のページから次の黒のページの直前のページまでがひとつのスレッドである。したがって、図 2 の閲覧履歴には図 2c に示すように 3 つのスレッドが含まれている。なお、一つのページから複数のリンクを同時に辿りそれらを読んだ場合、その全てのページも一つのスレッドとする。

### 3.2 現在閲覧中のページが再訪問の場合の推薦候補

今現在閲覧中のページと同じ URL のページを含むスレッドを検出する。そして、検出した各スレッド内の同一 URL のページ以降に閲覧したページを推薦候補とする。推薦候補のうち、満足度の高い Web ページを提示する。

### 3.3 現在閲覧中のページが初見の場合の推薦候補

まず、閲覧履歴に含まれる全てのページの内容を分析し、クラスタリングを行う。その結果とスレッドの二つを合わせることで閲覧パターンを解析する。すなわち、スレッド中の各ページの属しているクラスタを見て、どのクラスタからどのクラスタへ内容が移ったのかを調べる。

図 2d は図 2a をクラスタリングした結果であり、同一濃度のページは同一のクラスタに属するページを表す。図 2 の c と d を合わせることで、図 2 の閲覧履歴には図 2e に示すような 3 つの閲覧パターンがあるとするとする。

次に、現在スレッドを作成する。これは、閲覧開始ページから現在閲覧中のページに至るまでに閲覧したページのリストとする。そして、現在スレッドの閲覧パターンを、スレッド内の各ページを一番内容が類似しているクラスタに属させることで作成する。これをキーとして、閲覧履歴の中から同一の閲覧パターンを含むスレッドを検出する。

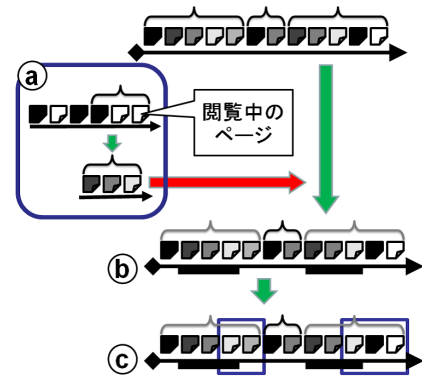


図 3. 同一の閲覧パターンを含むスレッドと満足したページ

図 3 に検出例を示す。この例では、現在スレッド (図 3a) は 3 ページからなる (閲覧しているページとその前の 2 ページ)。検出時にはこのスレッドの閲覧パターンが検索キーとなる。この例では、検出結果は図 3b のようになる。3 つあるスレッドのうち、かっこの色が薄いスレッドが現在スレッドと同一閲覧パターンが検出されたスレッドである。最後に、得られたスレッドの中で、検索キーとなった閲覧パターン以降のページを推薦候補とする。提示の際には推薦候補の中から満足度が高いページを提示する。

図 3b では線が太い部分が検索キーと同じ場所である。この場所以降に属するページ (図 3c の四角で囲まれた部分) が推薦候補である。

### 3.4 満足度の算出方法

満足度の算出には各ページの閲覧時間、閲覧回数、各ページの文字数を用いる。閲覧時間が長いほど、閲覧回数が多いほどユーザは満足したとする。ただし、ページに含まれる文字数が多い場合、内容に関係なく閲覧時間が長くなるので、文字数は満足度に反比例すると考える。

## 4 まとめ

過去に閲覧した Web ページの中から、探しているであろう Web ページを自動で推測し推薦するシステム「え〜し〜あ〜る」を示した。今後は、本システムの有効性を図るため、本システムを使用した場合と未使用の場合の満足する Web ページを見つけるまでのページ数等を検証する予定である。

## 参考文献

- [1] A. Cockburn and B. McKenzie. What Do Web Users Do? An Empirical Analysis of Web Use. Int. J. Human-Computer Studies, 54(6):903-922, June 2001.