

## WWWにおける履歴情報に基づく探索支援の考察

4L-9

川口 晃司\* 斎藤 一\*\* 前田 隆\*

北海道情報大学大学院\* 北海道大学工学部研究科\*\*

### 1. はじめに

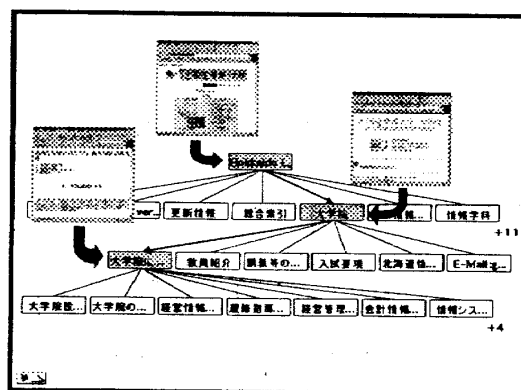
今日インターネットのWWWにおけるハイパーテキスト構造の情報は急速に増えており、その中を迷うこと無く、ユーザが興味のあるページを効率よくブラウジングすることは困難になってきている。そのような問題を解決する方法としてユーザの辿ってきたページとそのページに含まれるリンク情報を可視化し、ユーザの興味関心領域を推定することによってユーザに興味のあると思われるリンクを知らせる探索支援について考察する。

### 2. リンク情報の可視化

WWWをブラウジングする際に多くのユーザは自分が今までどのページを辿ってきたのか、どのページを今ブラウジングしているのかを頭の中に描いているものと思われる。これは短時間のブラウジングや記憶能力の優れたユーザにとってはたやすいことではあるが、ブラウジングが長くなるにつれて困難なこととなり、WWW内で迷子になることがある。このようにWWW上で迷わないようにするためにはユーザの辿ってきたページとそのページに含まれるリンク情報を可視化し、ユーザがブラウジングしたページの履歴と現在どのページをブラウジングしているのかを示すようにする。[1] このようにすることでユーザはブラウジングしてきたページの全体像を客観的に把握できWWW内で迷うことが少なくなる

### 3. 可視化の方法

ユーザの辿ってきたページを可視化するに当たって使用する情報は、そのページをブラウジングするために辿ったリンク文字、そのページのタイトル、そしてそのページ内に存在する辿らなかったリンク情報である。単にブラウジングしたページだけではなく、そのページに含まれる辿らなかったリンク情報も使用することでブラウジングしている（した）ページがどういうリンク情報を伴っているのか、どういうリンクの選択肢があるのかが分かるようになる。そして、各ページとリンク情報をノードとして表し、辿ったページを矢線で結ぶ。（図1参照）



(図1 可視化表現)

### 4. 興味関心領域の推定

履歴情報を可視化することで自分の現在位置を客観的に把握できたとしても、ユーザのメタ認知能力が低いことによって、多数のリンクの選択肢を前に何処にリンクしたら良いのかを決められなくなる可能性がある。[2] これを防ぐためにユーザの興味関心領域を推定し、興味のあると思われるリンクを勧める探索支援について考察する。

我々の提案する興味関心領域の推定方法は以下とうりである。

・ユーザの辿ったリンク情報を同じリンク文字ご

On Supporting of Information Search Process

based on Browsing History for WWW

\* Kouji KAWAGUCHI, Takashi MAEDA

Graduate School of Hokkaido Information University

59-2 NishiNopporo, Ebetsu, Hokkaido, Japan

\*\* Hajime SAITOH

Graduate school of Engineering Hokkaido University

Kita13, Nishi8, Kita-ku, Sapporo, Hokkaido Japan

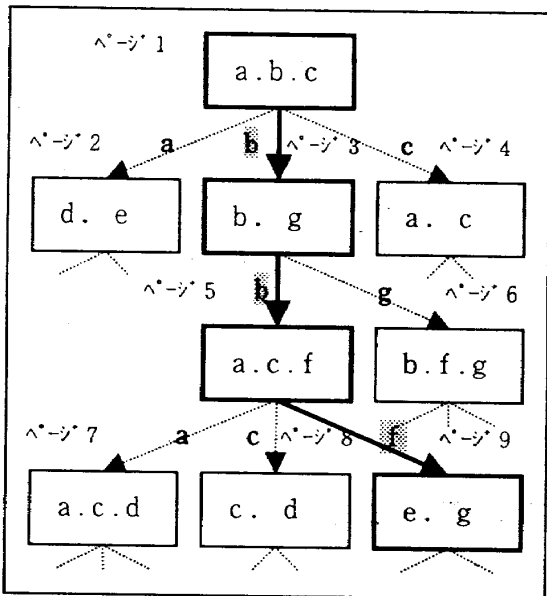
とにカウントし、ユーザの辿ったリンク情報ベクトルを求める。このベクトルを  $F$  とする。

- 各ページに含まれるリンク情報を同じリンク文字ごとにカウントし、各ページのリンク情報ベクトルを求める。このベクトルを  $N_i$  ( $i=1,2 \dots n$ ) とする。

- ベクトル  $F$  と  $N_i$  の内積を計算する

- 内積の計算結果より、数値の大きいものから順次ユーザの興味関心の高いページとする。

簡単なハイパーテキストの例 (図 2) を使って説明する。図 2 はページが 1 から 9 までであるハイパーテキスト構造になっており、各ページには a から g までのリンク情報が任意の数存在し、実矢線がユーザの辿ったリンクである。ユーザはページ 1 の "b" のリンク、ページ 3 の "b"、ページ 5 の "f" のリンクを辿っている。



(図 2 簡単なハイパーテキストモデル例)

この時のベクトル  $F$ 、 $N1 \sim N9$  は以下の通りである。

- ユーザの辿ったリンクのベクトル  $F$  (0,2,0,0,0,1,0)
- 各ページのリンク文字のベクトル
  - $N1$  (1,1,1,0,0,0,0) : ページ 1
  - $N2$  (0,0,0,1,1,0,0) : ページ 2
  - $N3$  (0,1,0,0,0,0,1) : ページ 3
  - $N4$  (1,0,1,0,0,0,0) : ページ 4

$N5$  (1,0,1,0,0,1,0) : ページ 5

$N6$  (0,1,0,0,0,1,1) : ページ 6

$N7$  (1,0,1,1,0,0,0) : ページ 7

$N8$  (0,0,1,1,0,0,0) : ページ 8

$N9$  (0,0,0,0,1,0,1) : ページ 9

$F$  と  $N1$  から  $N9$  までの内積を計算すると、

$$F \cdot N1 = 2, \quad F \cdot N2 = 0, \quad F \cdot N3 = 2$$

$$F \cdot N4 = 0, \quad F \cdot N5 = 1, \quad F \cdot N6 = 3$$

$$F \cdot N7 = 0, \quad F \cdot N8 = 0, \quad F \cdot N9 = 0$$

となり、この中でページ 1, 3, 5, 6 に当たる内積の値がそれぞれ、2, 2, 1, 3 で、ユーザの辿ってきたページが含まれており、提案方法の妥当性を示している。値の最も大きいものはページ 6 の "3" であり、このページ 6 をユーザの興味があるページとし、ユーザにブラウジングすることを勧める。このようにユーザの興味関心領域を推定し、興味のあると思われるページを勧めることで、ユーザに WWW 探索目的を明確にさせ、ブラウジング効率を上げさせる。

### 5. おわりに

今回は仮定の簡単な例を用いて WWW における探索支援の方法を提案したが、今後は実際の WWW 上のコンテンツについて実験すると共に、ブラウジングした時間やリンク情報のクラスタリングをすることで、より精度の高い興味関心領域の推定ができるものと思われ、実際のシステムを構築していきたい。

### 参考文献

- [1] 川口晃司、斎藤一、前田隆「WWWでの探索支援エージェントについての考察」
- [2] Robert E. Horn 著、松原光治 訳「ハイパーテキスト情報整理学」pp. 54-55 (1995)
- [3] 柏原昭博、佐竹義智、豊田順一「ハイパーメディア教材における履歴の可視化と知識整理支援」
- [4] 大久保雅且、杉崎正之、井上孝史、田中一男「www 検索ログに基づく情報ニーズの抽出」