

# ハイパーリンクコンテンツの速覧支援システム「ForeScope」 における先読み手法の提案と評価

深 津 真 二<sup>†</sup> 佐々木 努<sup>†</sup>  
大久保 雅且<sup>†</sup> 外 村 佳 伸<sup>†</sup>

本論文では、ハイパーリンクコンテンツの速覧支援システム「ForeScope」の先読み手法について述べ、本先読み手法の有用性を被験者実験をもとに評価した結果を述べる。本先読み手法では、利用者の閲覧操作、特に、マウスカーソルの移動軌跡をもとに、利用者が次に閲覧するであろうアンカーを含むエリア（先読みエリア）を予測し、その先読みエリア内のアンカーが示す Web ページを優先的に先読みする。評価実験の結果、本先読み手法を用いることで、Web ページの先頭にあるアンカーから順に先読みする手法に比べて、Hit 率（閲覧する Web ページが先読み済みであった確率）と適合率（先読みした Web ページのうち実際に閲覧された比率）がともに向上することを確認した。

## A Prefetching method for ForeScope, a Browsing system for hyperlinked contents, and Evaluation of the prefetching method

SHINJI FUKATSU,<sup>†</sup> TSUTOMU SASAKI,<sup>†</sup> MASAOKI OHKUBO<sup>†</sup>  
and YOSHINOBU TONOMURA<sup>†</sup>

In this paper, we present a prefetching method in our system for browsing hyperlinked contents efficiently, called ForeScope. The method predicts an area where a user will browse in the next seconds based on the user's browsing operation, and preferentially prefetch contents in the area. Experimental results shows the proposed method increase hit ratio and relevance ratio compared to a method which prefetch contents from top down.

### 1. はじめに

現在、Web ページ中には多数のハイパーリンクが設置されており、安形らが行った調査<sup>1)</sup>によると 1 ページ中のアンカー数は平均 18.4 個、Vers 社が行った調査によるとポータルサイトなどのトップページ中のアンカー数は平均 45 個<sup>2)</sup>であったことが報告されている。また、最近、Blog<sup>3),4)</sup> や Wiki<sup>5),6)</sup> が普及してきており、閲覧者が Web ページに任意にコメントを書き込めるコメント機能、相手の Web ページに自分の Web ページへのハイパーリンクを付与できるトラックバック機能などにより、Web ページ間のハイパーリンク構造が複雑になってきている。

そこで、ハイパーリンクで結ばれた Web ページの素早い閲覧を支援するシステムが各種開発されている。例えば、Zero-Click<sup>8)</sup> や AskJeeves、Girafa で

は、マウスカーソルをアンカー上に重ねることで、アンカーが示す Web ページや Web ページのサムネイル画像をポップアップ表示できる。また、InlineLink<sup>10)</sup> や Fluid annotations<sup>9)</sup>、WiseNut では、アンカーが示す Web ページをアンカー近傍に埋め込む形で表示できる。しかし、従来のシステムでは、Web 閲覧操作の単純化といった操作面の速覧支援を行うのみで、Web ページや Web ページのサムネイル画像は利用者からのアクセス発生後に取得する。その結果、実際に表示されるまでには待ち時間が生じ、Web 閲覧行為全体での速覧支援を考えた場合、この表示待ち時間を短縮することが重要になる。

本稿では、Web ページの表示待ち時間を短縮することを目的に、我々が開発するハイパーリンクコンテンツの速覧支援システム「ForeScope」<sup>7)</sup>における“先読み”手法を提案する。提案する先読み手法では、利用者の閲覧操作、特に、マウスカーソルの移動軌跡をもとに利用者が次に閲覧するであろうアンカーを含む

<sup>†</sup> NTTサイバーソリューション研究所  
NTT Cyber Solutions Laboratories  
<http://www.askjeeves.com/>  
<http://www.girafa.com/>

<http://www.wisenut.com/>

エリア（先読みエリア）を予測し、その先読みエリア内のアンカーが示す Web ページを予め取得する。そして、被験者実験の実験結果をもとに、本提案手法の効果と性能について述べる。

## 2. ForeScope

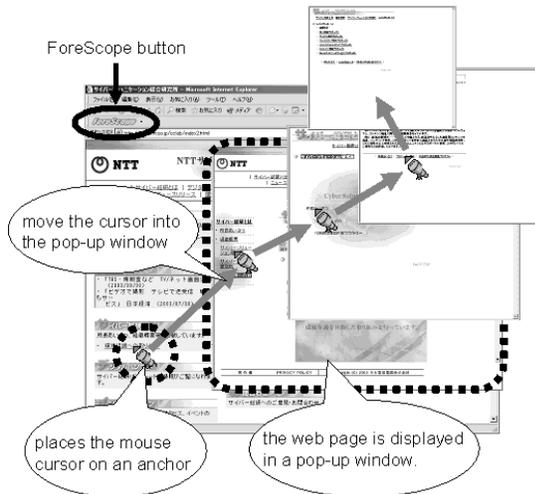
ForeScope では図 1(a) に示すように、マウスカーソルをアンカーに重ねる / 外すといった簡単な操作でアンカーが示す Web ページをポップアップ表示でき、さらには、そのポップアップウィンドウ内のアンカーにマウスカーソルを重ねることで、そのアンカーが示す Web ページをさらにポップアップ表示できる。また、図 1(b) に示すように、任意のポップアップウィンドウ / Web ブラウザへマウスカーソルを移動することで、先の任意の閲覧状態 / 初期状態に即座に戻れる。

このように、ForeScope を用いることで、ある Web ページを基点にハイパーリンクで結ばれた複数の Web ページを迅速かつシームレスに閲覧（基点不動型閲覧と呼ぶ）することが可能となる。なお、ForeScope は図 1(a) に示すように、Internet Explorer<sup>TM</sup> のツールバーに組み込まれ、利用者は任意のタイミングで ForeScope を起動できる。

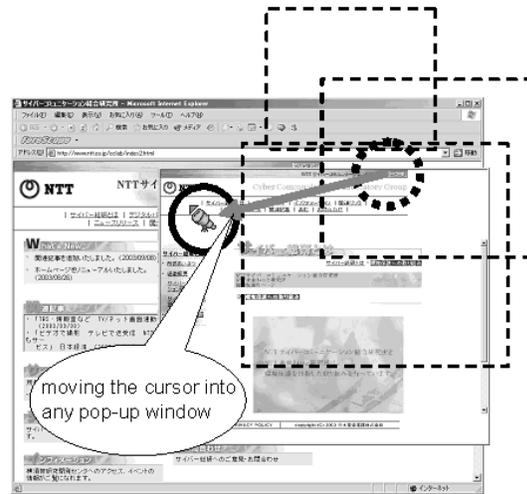
## 3. 関連研究

利用者からのアンカーに対するアクセスを推測し、アンカーが示す Web ページを予め取得する先読み手法が各種提案されている。例えば、知念等<sup>11),12)</sup> は、利用者が閲覧している Web ページの先頭にあるアンカーから順に先読みする手法を先読み代理サーバ上に実装し、先読み数が大きくなるに従い Hit 率は向上するが、ある値に収束することを示した。また、ie Ninja<sup>TM</sup> や極写 2001<sup>TM</sup> など、市販の先読みソフトウェアでは、アンカーが示す Web ページの種類やサイズ、タイプ（次ページへのアンカーなど）、内容（予め登録した語句を含むなど）をもとに優先順位を決定したり、利用者の閲覧パターンを統計的に分析し、アンカー毎に直進的に階層を辿りながら先読みする「直進型」とページ毎にアンカーが示す Web ページを順に先読みする「往復型」を自動的に選択する。

ここで、複数の Web ページを素早く閲覧、例えば、検索結果の各アンカーが示す Web ページを流し読みなど、する場合、利用者の Web 閲覧時の操作（マウスカーソルの移動、アンカーの選択・クリック、スクロー



(a) continuously browse multiple web pages



(b) go back to any previously browsed page

図 1 ForeScope

ルなど)は素早く、1つの Web ページを閲覧している時間も短い。そのため、従来の先読み手法のように、Web ページの表示時に先読み対象とするアンカーと優先順位を決定し、閲覧している Web ページの変化に合わせて変更する場合、先読みしている場所と利用者が閲覧している場所とに大きな隔たりが生じることが懸念される。そこで、本稿では、利用者の Web 閲覧時の操作に応じて動的に先読み対象と優先順位を決定し、利用者が次に閲覧するであろうアンカーを的確に推測する、利用者の先を読む先読み手法を提案する。

<http://www.ifour.co.jp/product/ien/>  
<http://202.83.137.29/webdata/catalog/s/sst/sste001.htm>

#### 4. 利用者の先を読む先読み手法

##### 4.1 概要

利用者の Web 閲覧状況を表す指標として、利用者からの Web 閲覧操作に伴う Web ページ中でのマウスカーソル位置の変化に着目し、マウスカーソルの移動軌跡をもとに動的に先読み対象と優先順位を決定することを考える。

本提案手法では、図 2(a) に示すように、マウスカーソルの移動軌跡をもとに利用者が次に閲覧するであろうアンカーを含むエリア（先読みエリア）を算出する。そして、図 2(b) に示すように、算出された先読みエリア内をマウスカーソルの移動方向を中心に放射状にサーチし、先読みエリア内のアンカーが示す Web ページを先読み対象とする。なお、図 2(b) 中の  $\psi_p$  はサーチ方向の変更角度、番号はサーチ順序を表す。以上の処理の結果、マウスカーソルの移動方向上で、現在のマウスカーソル近傍にあるアンカーが示す Web ページから先読みされるように優先順位を決定する。その後、利用者の Web 閲覧状況が変化、例えば、マウスカーソルが一定距離移動、マウスカーソルの移動方向が急変、アンカーのクリックによる Web ページの変化、一定時間経過など、に応じて先読みエリア、つまりは、先読み対象と優先順位を動的に変更する。

このように、本提案手法では、その時々で利用者が次に閲覧する可能性の高いアンカーのみを先読み対象とする。例えば、現在のマウスカーソルの位置から遠く、すぐに閲覧される可能性の低いアンカーは、そのアンカーが近づくまで先読み対象としない。その結果、先読み処理にかかるクライアントやネットワークへの負荷を軽減することが可能となる。

##### 4.2 先読みエリアの算出

本節では、図 3 を参照し、本提案手法の先読みエリア算出アルゴリズムを説明する。本提案手法では、

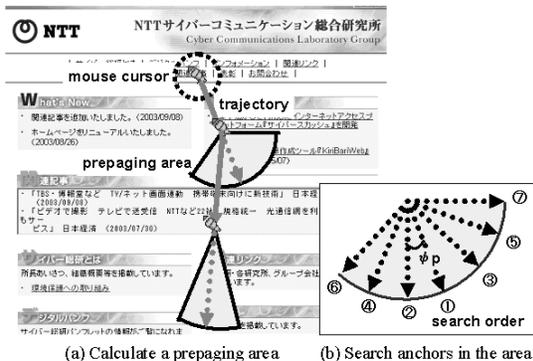


図 2 利用者の先を読む先読み手法

マウスカーソルの位置情報  $P_c(X_c, Y_c), P_c'(X_c', Y_c')$ 、移動方向  $\Theta_c$ 、移動速度  $V_c$  をもとに、先読みエリア（半径： $R_p$ 、角度： $\Phi_p$ 、面積： $S_p$ ）を算出する。この際、下記の式に示すように、先読みエリアの半径  $R_p$  を移動速度  $V_c$  に比例、先読みエリアの角度  $\Phi_p$  を移動速度  $V_c$  に反比例させる。

$$R_p = f(V_c)$$

$$\Phi_p = f(1/V_c)$$

$$V_c = \frac{\sqrt{(X_c - X_c')^2 + (Y_c - Y_c')^2}}{T_c - T_c'}$$

$T_c, T_c'$  はカーソル位置の取得時刻

$$\Theta_c = \arctan \frac{Y_c - Y_c'}{X_c - X_c'}$$

$$S_p = \pi R_p^2 \times \frac{\Phi_p}{360}$$

このようにすることで、マウスカーソルの移動速度が遅い、例えば、利用者がじっくりと Web ページを閲覧しているような場合、先読みエリアはマウスカーソルの位置を中心に大きく広がった扇形（半径：小、角度：大）となり、現在のマウスカーソル位置近傍の広い範囲に存在するアンカーが示す Web ページの先読みが可能となる。一方、マウスカーソルの移動速度が速い、例えば、利用者がさっと Web ページを閲覧しているような場合、先読みエリアはマウスカーソルの移動方向に縦長な鋭利な扇形（半径：大、角度：小）となり、現在のマウスカーソル位置から遠く離れた場所に存在するアンカーが示す Web ページの先読みが可能となる（図 2 参照）。このように、本提案手法を用いることで、利用者の閲覧速度に応じた先読みエリアの算出、つまり、先読みが可能となる。

また、先読みエリアの面積  $S_p$  をクライアントの処理速度やネットワークの通信速度など、利用者の使用環境毎に設定される定数であると考え、この面積成分  $S_p$  により先読みエリアの算出に制限を設ける。これ

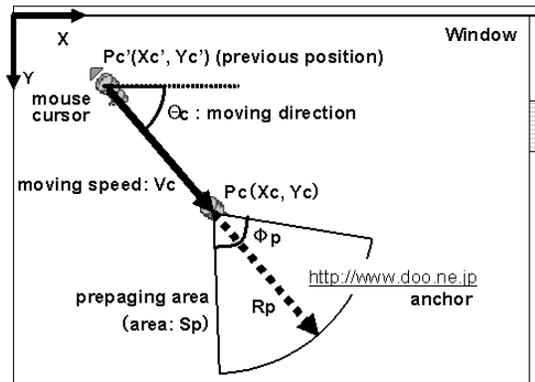


図 3 先読みエリアの算出

により、各使用環境に適した先読みを行えるだけでなく、Web ページのアンカー密度に応じた先読みを実現できる。例えば、アンカー数が密な場所では先読み対象とするアンカー数が増加し、アンカー数が疎な場所では先読みエリア対象とするアンカー数が減少する。

## 5. 評価実験

### 5.1 実験概要

提案する先読み手法の効果と性能、特に、実際の利用シーンでの有用性を評価することを目的に実験を行った。被験者はメールや Web、文章入力など、日常的にコンピュータやインターネットを利用している一般ユーザ 8 名（男：2、女：6、20～30 代前半）で、ノート PC とデスクトップ PC を各 4 台（CPU:Celeron 600MHz, メモリ:128MB～CPU:Pentium4 3.2GHz, メモリ:1023MB）用いて実験を行った。なお、各 PC の OS は WindowsXP SP2、Web ブラウザは Internet Explorer<sup>TM</sup> Ver.6 と共通で、ノート PC の場合でもホイールマウスの使用を求めた。なお、実験時のネットワークスピードは平均 2Mbps (0.90Mbps～3.48Mbps) であった。

また、実験に先立ち、被験者には「ForeScope」(先読みはなし)を用いて自由に Web 閲覧してもらい、「ForeScope」の操作方法を習得してもらった。

### 5.2 実験 1：先読み手法の比較評価

#### 5.2.1 実験方法

被験者には、提案手法 [proposed] と比較対照として準備した、Web ページの先頭にあるアンカーから順に先読みする手法 [normal] の各先読み手法が実装された「ForeScope」を用いて、10 分間、指定する Web サイト (goo、Yahoo!、OKWeb など) を中心に、任意の Web ページを自由に閲覧してもらった。この際、被験者を 2 つのグループに分け、一方は提案手法 [proposed] を先、他方は比較対象の先読み手法 [normal] を先に用いて実験を行い、後日、順番を変えて同様の実験を行った。

本実験で用いた提案手法 [proposed] における先読みエリアの算出式を以下に示す。具体的には、マウスカーソルの移動速度  $V_c$  : 100pixel/s (筆者の使用環境におけるホイール 1 操作あたりのスクロール量) の場合に、先読みエリアの半径  $R_p$  : 200pixel/s、角度  $\Phi_p$  : 80° となる。

$$R_p = 2V_c, \Phi_p = \frac{320}{(V_c/50)^2}$$

$$S_p = \pi(2 \cdot 100)^2 \times \frac{80}{360} = \text{一定}$$

### 5.2.2 結果と考察

ForeScope を用いて閲覧 (アンカーが示す Web ページをポップアップ表示) した回数 “閲覧数”、ForeScope を用いて閲覧した Web ページが先読み済みであった回数 “先読み済数”、先読みした Web ページの総数 “先読み数”、利用者が閲覧した Web ページ中のアンカー数の累計 “アンカー数” を測定した。表 1 に各種測定データを被験者数で割った平均を示す。また、測定データをもとに、閲覧する Web ページが先読み済みであった確率 “Hit 率” (= 先読み済数/閲覧数) と先読みした Web ページのうち実際に閲覧された比率 “適合率” (= 先読み済数/先読み数) を算出した結果を図 4 に示す。なお、図 4 中の各ラベル下部に片側有意水準 5% にて t 検定を行った結果を示す。

実験の結果、提案手法 [proposed] を用いた場合、Web ページの先頭にあるアンカーから順に先読みする手法 [normal] に比べて、有意差は認められなかったものの Hit 率は向上 (約 1.24 倍) し、適合率は有意に向上 (約 1.89 倍) することが示された。この際、提案手法 [proposed] での先読み数は、先読み手法 [normal] での先読み数の約 3/4 であった。以上のことから、本提案手法を用いることで、閲覧される可能性の高いアンカーが示す Web ページを効率的に先読みできることが分かった。

また、先読み済みであった Web ページを分析した結果、ニュースサイトやポータルサイトのトップページなど頻繁にアクセスされた Web ページを閲覧、他

表 1 各種測定データ

先読み手法	閲覧数	先読み済数	先読み数	アンカー数
[proposed]	50.44	7.06	123.50	2865.6
[normal]	45.75	4.13	164.13	2564.8

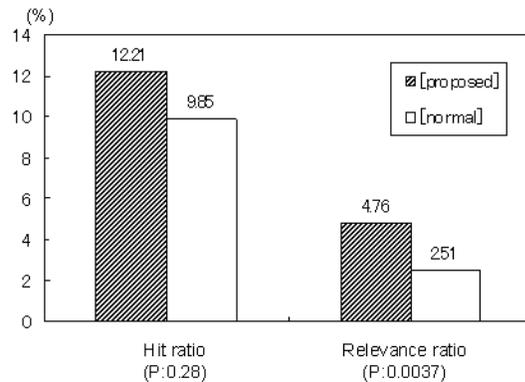


図 4 Hit 率と適合率 (実験 1)

には、ニュースサイトやポータルサイトのトップページなど、アンカーが整列して表示されており、各アンカーを順番に閲覧していたような場合であったことが分かった。ここで、先読みが的中した理由として、前者は、事前に閲覧されていることでクライアントにキャッシュが存在し、アクセスが速かった結果と考えられる。それに対し、後者は、本提案手法の先読みエリアの算出、つまりは、先読み対象の推測が的確に行われた結果と考えられ、ある Web ページ中の複数のアンカーが示す Web ページを連続して閲覧するような場合に、提案する先読み手法は効果的に働くことが分かった。

ここで、実験の結果、提案手法 [proposed]、先読み手法 [normal] の両方で、Hit 率と適合率がともに低くなったのは、被験者の 1 ページ当たりの平均閲覧時間が提案手法 [proposed] で 10.87 秒、先読み手法 [normal] で 13.62 秒と短く、その結果、1 ページあたりの平均先読み数が提案手法 [proposed] で 3.23 個、先読み手法 [normal] で 5.12 個と少なかったことが原因と考えられる。なお、NetRating 社の調査<sup>13)</sup>によると、ブロードバンドユーザの 1 ページ当たりの平均閲覧時間 39.65 秒である。実際、被験者の先読みログを分析した結果、あるアンカーの示す Web ページが先読み対象として登録されるものの、そのアンカーへのアクセスの方が早く、先読みが間に合わないことがしばしば発生していた。また、本実験では被験者に ForeScope の操作方法を予め習得してもらっていた結果、マウスカーソルの移動速度が速く、緩急も激しかったため、本提案手法における先読み範囲の算出にズレが生じたと考える。

### 5.3 実験 2：先読みエリアの形状が与える影響

#### 5.3.1 実験方法

提案する先読み手法において、算出される先読みエリアの面積は全て同一であるが、先読みエリアの形状が異なる 3 つの手法 [base]、[long] (先読みエリアが縦長)、[wide] (先読みエリアが幅広) を準備し、先読みエリアの形状が先読み効果に影響を与えるかどうかを調べた。各手法における先読みエリアの算出式を以下に示す。

被験者は実験 1 の被験者と同一で、実験 1 の終了後、各先読み手法が実装された「ForeScope」を用いて、10 分間、指定する Web サイトを中心に、任意の Web ページを自由に閲覧してもらった。この際、被験者は全員、[base]、[long]、[wide] の順で実験を行い、後日も同様の順番で実験を行った。

$$[base] R_p = 2 \cdot V_c, \Phi_p = \frac{320}{(V_c/50)^2}$$

$$[long] R_p = 4 \cdot V_c, \Phi_p = \frac{80}{(V_c/50)^2}$$

$$[wide] R_p = 1.4 \cdot V_c, \Phi_p = \frac{640}{(V_c/50)^2}$$

$$S_p = \pi(2 \cdot 100)^2 \times \frac{80}{360} = \text{一定}$$

#### 5.3.2 結果と考察

表 2 に各先読み手法ごとに各種測定データを被験者数で割った平均を示す。また、各先読み手法ごとに Hit 率と先読み効率を算出した結果を図 5 に示す。なお、[base]-[long]、[base]-[wide]、[wide]-[long] のそれぞれの組合せにおいて、片側有意水準 5% にて t-検定を行った結果、Hit 率、先読み効率ともに有意差は確認されなかった。

実験の結果、先読みエリアの形状の違いによらず、Hit 率はほぼ同じ結果となった。これは、各先読み手法での 1 ページ当たりの平均先読み数が [base] で 3.40 個、[long] で 3.59 個、[wide] で 3.00 個と大差なく、どれも少数であったことを考えると、マウスの移動方向上でかつ近傍のアンカーが示す Web ページのみが先読みされ、先読みエリアの形状による違いが影響しなかったものと考えられる。このことより、今回の実験においては、本提案手法で提案している、利用者の閲覧速度に応じた先読みエリアの算出、それによる先読みエリアの形状の変化はあまり効果がなかったことが示された。

表 2 各種測定データ

先読み手法	閲覧数	先読み済数	先読み数	アンカー数
[base]	38.25	4.63	96.63	2393.3
[long]	35.19	3.69	88.38	2253.7
[wide]	47.31	4.56	95.75	2722.8

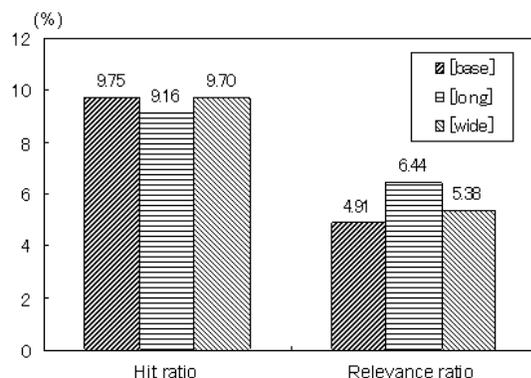


図 5 Hit 率と適合率 (実験 2)

次に、適合率については、各先読み手法間に有意差は認められなかったが、ある被験者の [long] における適合率が 54.5% と非常に高かったため、[long] で先読み効率が少し高くなる結果となった。そこで、この被験者の [long] での操作ログを分析した結果、ポップアップ表示を 1 階層で終わることが多く、階層的にポップアップ表示することが少なかったことが分かった。また、各先読み手法において適合率が高かった被験者の操作ログを分析した結果、同様の傾向が確認された。つまり、適合率が高かった被験者は複数の Web ページを移動し階層的に閲覧するのではなく、ある Web ページ中の複数のアンカーが示す Web ページを連続して閲覧していた結果、本提案手法の先読みが効果的に働いたと考えられる。なお、実験 1 では、提案手法 [proposed] と先読み手法 [normal] のどちらの場合でも、被験者全員が適度に階層的なポップアップ表示を行っていた結果、上記のような傾向は確認されなかった。

なお、実験 2 では、Hit 率と適合率が実験 1 に比べて低い結果となった。これは、実験 2 では、先読み対象の検出に漏れが生じないように、先読みエリア内のサーチ間隔を実験 1 の半分に設定した。そのため、先読み対象検出にかかる負荷が増加し、先読み数が実験 1 に比べて減少した結果と考える。

## 6. おわりに

利用者の Web 閲覧操作に伴うマウスカーソルの移動軌跡をもとに、利用者が次に閲覧するであろう先読みエリアを算出し、その先読みエリア内のアンカーが示す Web ページを優先的に先読みする先読み手法を提案した。そして、評価実験の結果、Web ページの先頭にあるアンカーから順に先読みする手法に比べ、本提案手法を用いることで、Hit 率が 1.24 倍、適合率が 1.89 倍向上することを確認した。さらには、提案する先読み手法において、先読みエリアの形状が先読み効果に与える影響を調べた。しかしながら、今回の実験では先読み数が全体的に少なかった結果、先読みエリアの形状による違いは確認されなかった。

今後の課題としては、今回の実験結果を踏まえ、例えば、先読み処理をマルチスレッド化し、先読み数を増加することで、Hit 率と適合率の向上を図ることが挙げられる。他にも、クライアントの処理速度やネットワークの通信速度とマウスカーソルの移動速度との関係、さらには、マウスカーソルの移動加速度に応じて、現在のマウスカーソルの近傍から優先的に先読みするのではなく、先読みエリアの中ほどから先読み、

または、先読みエリアの端から逆方向に先読みするなど、利用者の操作状況をより踏まえた先読み手法を開発することが考えられる。

## 参 考 文 献

- 1) 安形輝, 野末道子, 久野高志, 池内淳, 石田栄美, 上田修一: Web ページの実態調査と自動判定, 2002 年度日本図書館情報学会春季研究大会発表要綱, pp. 11-14 (2002).
- 2) ヴェルス社: WE B 統計 (2000).  
<http://www.vers.co.jp/stat/statmain.html>.
- 3) 武田英明, 大向一輝: Weblog の現在と展望セマンティック Web およびソーシャルネットワーキングの基盤として, 情報処理, Vol. 45, No. 6, pp. 586-593 (2004).
- 4) ARTIFACT 人工事実: Weblog ツールリスト (2004). <http://artifact-jp.com/weblog/>.
- 5) Leuf, B., Cunningham, W. y.: Wiki Way コラボレーションツール Wiki, ソフトバンクパブリッシング株式会社 (2002).
- 6) Wiki 遊びのお砂場: WikiLab/Links (2004).  
<http://jm.vis.ne.jp/sbox/index.cgi?WikiLab>
- 7) 深津真二, 阿久津明人, 外村佳伸: ForeScope: 階層化コンテンツの閲覧支援システム, 情報処理学会研究報告 2003-HI-102, pp. 7-12 (2000).
- 8) Tomoyuki, N., Suguru, S. and Manabu, O.: Zero-Click: a System to Support Web Browsing, *In proceedings of the 11th International World Wide Web Conference* (2002).
- 9) Bouvin, N., Zellweger, P., Gronbak, K. and Mackinlay, J.: Fluid Annotations through Open Hypermedia: Using and Extending Emerging Web Standards, *In proceedings of the 11th International World Wide Web Conference*, pp. 160-171 (2002).
- 10) Miura, M., Shizuki, B. and Tanaka, J.: InlineLink: Inline Expansion Link Methods in Hypertext Browsing, *In proceedings of International Conference on Internet Computing Vol.II*, pp. 653-659 (2001).
- 11) 知念賢一, 山口英: WWW 先読み代理サーバにおける先読み対象決定戦略, 情報処理学会研究報告 96-DPS-77, pp. 55-60 (1996).
- 12) 知念賢一, 井上博之, 岡山聖彦, 山口英: WWW におけるハイブリッド先読み代理サーバの設計と実装, 信学会論文誌 D-I, Vol. J80-D-I, No. 11, pp. 907-915 (1997).
- 13) Nielsen//NetRatings: 2003 年 12 月の月間インターネット利用動向調査結果 (2004).  
[http://csp.netratings.co.jp/nnr/PDF/126\\_2004release-j-final.pdf](http://csp.netratings.co.jp/nnr/PDF/126_2004release-j-final.pdf).