

| | |
|--------------|---|
| Title | 先読み代理サーバを用いたWWW情報探索支援 |
| Author(s) | 新井, 孝之 |
| Citation | |
| Issue Date | 2002-03 |
| Type | Thesis or Dissertation |
| Text version | author |
| URL | http://hdl.handle.net/10119/1523 |
| Rights | |
| Description | Supervisor: 白井 清昭, 情報科学研究科, 修士 |

先読み代理サーバを用いた WWW 情報探索支援

新井 孝之

北陸先端科学技術大学院大学 情報科学研究科

2002 年 2 月 15 日

キーワード: Web ブラウジング, 先読み代理サーバ, 要約, ユーザプロファイル.

現在 Web 上には膨大な情報が存在するため, Web から有益な情報を探し出す作業に多くの時間がかかる場合も多い. 通常, Web ページには多数のリンクが付けられており, ユーザはリンク先にジャンプしたり, また戻ってきたりして Web 探索を行う. しかし, この動作は非常に煩わしい. したがって, リンク先に頻繁にジャンプするイベントを減らすことは, Web 探索の効率を上げるのに役立つと考える. 本研究では, WWW 情報探索において, ユーザが Web ページ上のリンクにマウスカーソルを置くと, リンク先の Web ページの要約が見えるようにすることにより, どのリンク先が重要かをユーザが実際に閲覧する前に判断できるシステムを構築する. 近年, Web のように, 電子化されたテキストが大量に利用可能になっていることから, テキスト自動要約は脚光を浴びてきている. 本研究では, 要約生成技術を WWW 情報探索支援に用いることを考える. システムは, プラットフォームに依存しない実装方法の一つである先読み代理サーバとして実装する. リンク先ページの要約は, リンク先ページを事前に読み込み (先読みし), ユーザから要約の要求がくる前に作成しておく. しかし, 通常 Web ページ上にはリンクが複数あるため, すべてのリンク先ページに対して同時に先読みを行うことは, 効率や迅速性の観点から望ましくない. 本研究では, どのリンク先から要約を作成するかの順序付けを Web ページ中のアンカーテキストとユーザプロファイルから決定する. これによって, ユーザの要約の要求に迅速に応えることができると考える.

提案する WWW 情報探索支援システムの大まかな処理の流れについて説明する.

1. ユーザは通常の Web ブラウジングと同様に好きな Web ページを閲覧する. システムは, ユーザが要求した Web ページを解析し, 要約が表示できるように JavaScript を追加する. また, そのページに存在するすべてのハイパーリンクについて, リンク先のテキストファイルを読み込み, 要約を作成してキャッシュに保存する. ユーザが要約を要求したら, キャッシュに保存されている要約を取り出すことによってリンク先の要約を高速に表示できる.

2. ユーザが好きなアンカーの上にマウスポインタを置くと、JavaScript はリンク先の要約をツールチップまたはサブウィンドウに表示させる。要約を表示することにより、ユーザはリンク先の情報が有用であるかどうかを判断することができる。これにより、ユーザがあまり関心のないページを表示させてから元のページに戻る操作を少なくすることが期待できる。
3. ユーザがアンカーからマウスポインタを離すと、JavaScript はツールチップを消す。
4. ユーザが実際にリンクをクリックするなどして Web ページの要求をしたら、キャッシュに保存されている情報を取り出すことによってリンク先のページを高速に表示できる。キャッシュに保存されていないならば、システムは Web ページをインターネット上から入手する。

ユーザが WWW 情報探索を行っている間、上記の操作を繰り返す。本研究では、要約の表示方法として、ツールチップとサブウィンドウの 2 つを用いる。

本研究では、要約を生成する手法として、HTML のタグを利用したり、Web ページの位置情報を手がかりとした手法は有効ではないと考え、テキストを対象とした既存の要約技術を利用する。具体的には、TF に基づく重要文抽出法によって要約を作成した。

先読みはネットワークに対して多大なトラフィックを発生させるため、Web ページ中のすべてのアンカーについて同時に先読みを行うことは望ましくない。従って、同時に先読みする数を制限し、先読みする順序を決定し、その順序に従って逐次的に先読みを行う。先読みの順序を決定する手法としては、ユーザの興味（ユーザプロファイル）とアンカーテキストから決定する。ユーザにとって興味のある単語がアンカーテキストにより多く含まれていれば、ユーザは、そのアンカーを他と比較して見たいであろうと考える。システムは、ユーザが閲覧した Web ページの単語をユーザプロファイルとして保存しておき、アンカーテキストの単語とマッチングさせて先読みの順序を決定する。

本研究で提案する WWW 情報探索支援システムのプロトタイプシステムを作成し、評価実験を行った。その結果、以下のようなことがわかった。

- リンク先の要約を見ることが WWW 情報探索支援に有効であることが示された。
- 本研究で作成したプロトタイプシステムでは、先読みの効果は認められた。
- しかし、要約の先読みのヒット率は平均 27.6% であり、改善の余地がある。
- リンク先の情報の表示方法としては、先行研究の多くで用いられているツールチップ表示より、サブウィンドウ表示の方が良いという結果が得られた。