

Webドキュメントの要約と携帯端末での閲覧方式

4T - 05

正賀 信寛[†] 角谷 和俊[‡] 上原 邦昭[‡]

[†] 神戸大学大学院自然科学研究科情報知能工学専攻

[‡] 神戸大学都市安全研究センター 都市情報システム分野

E-Mail: {shouga,sumiya,uehara}@ai.cs.kobe-u.ac.jp

1 はじめに

近年の携帯電話の普及により、携帯電話からも Web ドキュメントを閲覧したいという要求が高まってきている。携帯電話の特性として、ディスプレイが小さい、方向キー、数字キーで操作する、などがあげられる。したがって携帯電話での Web ドキュメントの閲覧はこれらの特性をふまえて行う必要がある。また、携帯電話で見ることを前提とした Web ドキュメントもあるものの、その数は一般のページに比べると少ない。本研究では、一般のページを携帯電話で閲覧するためのページの変換と要約について提案する。まず、複数の Web ページの集合である Web ドキュメントをリンク関係を基に分析し、ページを表示する順序を自動的に生成する方式について検討する。また、携帯電話でのページ閲覧について述べる。

2 ドキュメントの要約

Web 空間上の全ページを対象とし、リンク関係に基づいて Web ページの種類・重要度を判別する方法 [1] があるが、本研究では、異なるドメイン間でのリンク関係は考慮せず、同一ドメイン内のページ間でのリンク関係のみを対象とし、ドキュメントを要約する方式について提案する。

2.1 リンク解析

ユーザの指定したページを始点として、まずそのページのリンク先ページを抜き出し、その中から同一ドメイン内のページのみを取り出す。次に、得られたページの中でまだ調べていないページに対して同様の操作をおこなう。この手順を繰り返すことにより、ドキュメント内での各ページの構造を分析する。また、各ページごとにいくつかのページからリンクされているか (IN 値)、いくつかのページにリンクしているか (OUT 値) を求める。さらに、IN 値と OUT 値とを比較し、IN 値 $\leq$ OUT 値 となるページを Hub ページ、IN 値 $>$ OUT 値 となるページを Authority ページとする。最後に、より多くの Hub ページにリンクしている Hub ページ (Hub の Hub) を Top ページとする。

2.2 リンクの重み付け

あるページ P_j に入出入りしているリンクの重みは、図 1 のようにリンク元ページ P_i およびリンク先ページ P_k が Hub か Authority かの、4 つのパターンに分けて求める。

- (a) Hub ページ P_i からのリンクの重みは、 P_i の OUT 値、 $OUT(P_i)$ とする。
- (b) Hub ページ P_k へのリンクの重みは、 P_k の IN 値の逆数、 $\frac{1}{IN(P_k)}$ とする。
- (c) Authority ページ P_i からのリンクの重みは、 P_i の OUT 値の逆数、 $\frac{1}{OUT(P_i)}$ とする。
- (d) Authority ページ P_k へのリンクの重みは、 P_k の IN 値、 $IN(P_k)$ とする。

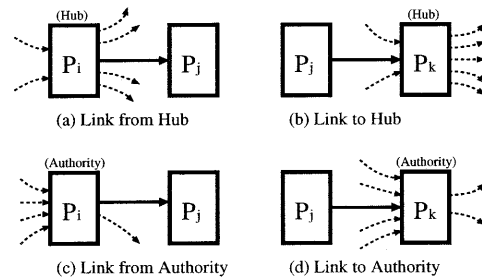


図 1: ページ P_j に入出入りするリンクの重み付け

2.3 ページの重み付け

通常、Web ページには図 2 に示すように複数のリンクが入り出している。図 2 において、ページ P_j に入ってくるリンクの重みを X_s ($s = 1, \dots, m$)、出ているリンクの重みを Y_t ($t = 1, \dots, n$) とすると、ページ P_j の重み W_j は、次式により求められる。

$$W_j = \sum_{s=1}^m X_s + \sum_{t=1}^n Y_t$$

3 閲覧順序の生成

2 章で述べたアルゴリズムを用いて Top ページと Authority ページの確定および各ページの重み付けが完了したら、続いて Top ページから Authority

⁰ Summarizing and Browsing Web Document for Mobile Environment, Nobuhiro SHOUGA, Kazutoshi SUMIYA and Kuniaki UEHARA, Kobe University

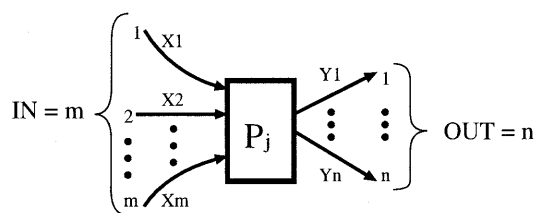


図 2: ページ P_j の重み付け

ページへの最短パスを動的に生成 [2] する。実際は、Authority ページから順にリンク元ページを辿り、最も少ない回数のページ移動で Top ページへと辿れるパスを見つけ出す。もし、同じ回数で複数のパスが見つかった場合は、途中に通っているページの重みの和をパスごとに求め、和の大きい方のパスを選択する。また、リンク元ページを辿っていく過程でパス内で既に辿ったページを再び訪れた場合は、そのパスを消して他のパスを調べる。

以上の手順により、Top ページを根とし、Authority ページを葉とした木構造が得られる。この木構造に対して、Top ページを始点としページの重さにより、深さ優先探索を行うことで閲覧順序が生成される。

4 携帯端末での閲覧

4.1 ページの要約

携帯電話上で一般の Web ドキュメントを閲覧する場合は、小さい画面でも閲覧しやすいようにページのタグに注目しページを要約していく。具体的には次のようなタグがあげられる。

図

図の代わりに図へのリンクを表示する。リンクを辿ることにより図が表示される。

表 <TABLE>, <TR>, <TH>, <TD>

ディスプレイにはテーブルの縮小図を表示しておき、各セルには数字を振り分けておく。方向キー、数字キーを用いてセルの数字を指定し、そのセルと周辺のセルを表示する。

アンカー <A>

アンカーには番号を割り当てて、数字キーを押すことで対応するページへ移動できるようになっている。実際のページ移動の前に移動先のページのタイトル、最頻出単語、画像数、アンカー数などの情報を先読みしてユーザに提示する。

フレーム <FRAMESET>, <FRAME>

フレームはその特徴から、ひとつのフレームが Hub ページとなっていると考えられる。<FRAMESET> が記述されているページと Hub ページとを同一のものとして扱う必要がある。

4.2 閲覧手順

本手法を用いて携帯電話で Web ドキュメントを閲覧するときの手順を以下に示す。

まず、閲覧したいページの URL を入力する。この URL を始点とし、同一ドメイン内でリンク解析を行う。解析の結果得られたリンク構造に、2, 3 章において説明したアルゴリズムを適用することで閲覧順序が得られる。例えば、図 3 (a) のような構造が生成された場合は、図 3 (b) の閲覧順序、

$$A \Rightarrow E \Rightarrow H \Rightarrow I \Rightarrow B \Rightarrow G \Rightarrow C$$

が得られる。この順序に従い各ページを閲覧する過程で、4.1 節のページ内要約や、音声によるページの読み上げ [3] を行う。

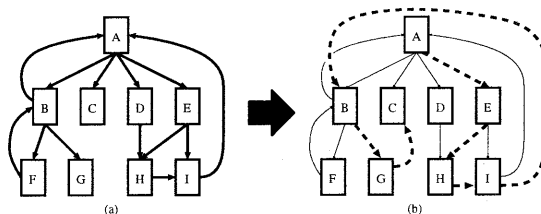


図 3: Web ドキュメントの例

5 おわりに

本論文では、携帯端末を使って Web ドキュメントを閲覧する際に、効率よく閲覧できるようにドキュメントを要約し、自動的に閲覧順序を生成する方式について提案した。また、携帯端末のディスプレイの大きさとボタン操作を考慮し、内容把握が容易に行えるようなページの要約について説明した。今後の課題として、実際のページを用いて検証を行うことなどがあげられる。

謝辞

本研究の一部は、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」(プロジェクト番号 JSPS-RFTF97P00501)、および文部省科学研究費「マルチメディアコンテンツの放送型配信に関する研究」(課題番号 12780308) による。ここに記して謝意を表します。

参考文献

- [1] Jon M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *ACM-SIAM Symposium on Discrete Algorithms*, 1998, <http://www.cs.cornell.edu/home/kleinber/auth.ps>
- [2] 角谷和俊, 正賀信寛, 上原邦昭. WebSkimming: WWW ページ群の動的要約による閲覧支援. 情報処理学会第 60 回全国大会論文集 (3), pp. 137-138. 情報処理学会, (2000).
- [3] 惣田一幸, 角谷和俊, 上原邦昭. 携帯情報端末における音声を用いた Web ナビゲーション. 情報処理学会研究報告 2000-DBS-122, pp. 323-330. (2000).