時間制約下でのWWW検索スケジューリング方法

3T－5

柏井 優希　灰原 清太郎　川越 恭二
立命館大学理工学部

1. はじめに

現在、インターネットを用いた情報収集が盛んに行われている。特にWWW上の情報量は多大であり、利用者はその規模の大きさのあまり、情報を得るために必要な時間がどれくらいなのか見当がつけにくい。そこで、時間に制約のある利用者がWWW上で検索を行なう場合にも、必要な情報をある時間内で収集できるようなスケジューリング方法を提案する。

2. 従来のWWW上での検索方法と問題点

利用者がWWW上で情報検索を行なうには、通常、検索のためのWebサービスを利用する。これらのサービスには、大きく分けて利用者が求めている情報のキーワードを用いて検索を行なうディレクトリ階層型と、利用者による語句の入力を用いて文字列検索を行なうサーチエンジン型が存在する。しかし検索結果の情報量が多大になることがあり、この場合利用者が必要な情報を得るために必要な仕事量は増大することになる。これを解決する方法の一つとして、従来、検索された情報の各々に価値をもたせてソートをする方法(情報価値によるソーティング)が用いられている。しかし、情報価値基準が万人にとって完璧であることはあり得ず、利用者が人間であるがために不満が起きるソーティングが発生する。このため、利用者はソーティングに従って求めるべき情報を探し続けることになる。

3. 情報認識時間を加味したスケジューリング

従来のWWW検索での問題点を改善するために、限られた時間の中で如何に効率よく必要な情報を得ることができるかに注目する。利用者の収集時間が限られているときには、目的に近い内容でなおかつより多くの種類のページを目に触れさせることが重要である。そこで、従来の情報価値によるソーティングに、情報認識に必要な時間を加味することを提案する。これをスケジューリングと呼ぶ。

3.1 情報認識時間

スケジューリングは、人間が一つのページからそのページの全ての情報を得ることが可能な時間(情報認識時間)を元に考えることにする。特定のページを読むのに必要な情報認識時間の正確な値を求めることは不可能である。しかし情報認識時間を推測する際に必要な情報の要素は、検索対象ページにおける

・文書量
・ページが持つリンク先の数
・gifなどのグラフィックの種類
・アプリケーションの必要数

などを使用する。

これらは、人間の時間拘束$f(x)$とマシンの時間拘束$g(y)$に分離できる。情報認識時間$T_{\inf}$を、次のような式で計算する。

$$T_{\inf} = f(x) + g(y)$$
$$f(x) = (\text{リンクの数} \times t_1 + \text{文書量} \times t_2) \times C_1$$
$$g(y) = (\text{読み込むファイル数} \times t_3) \times C_2$$

3.2 スケジューリングの方法

検索対象ページの情報価値の要素と情報認識時間の要素を考慮して以下のスケジューリングを実現する。

"A scheduling method for WWW searching on the consideration of time constraint"
Yuuki KASHIWAI, Seitarou HAIBARA and Kyoji KAWAGOE
Ritsumeikan University
1-1-1,Noji-higashi,Kusatsu,Japan

情報価値からスコア S_{inf} を定め、それに情報認識時間 T_{inf} を加味して TS_{inf} を算出し、その値でソートを行なった後、利用者が制約時間分 T_{max} だけ出力するというプロセスとなるが、TS_{inf} の求め方としてここでは 2 つ挙げる。

$$TS_{inf}{}^{1} = \frac{S_{inf}}{T_{inf}} \times C_3$$

$$TS_{inf}{}^{2} = \left(-\frac{S_{inf}}{T_{max}} T_{inf} + S_{inf} \right) \times C_4 \quad (T_{max} \geq T_{inf})$$

$$T_{max} < T_{inf} \text{のとき} TS_{inf}{}^{2} = 0$$

利用者が 30 分の時間 ($T_{max} = 30\text{min}$) で情報検索を行ないたいのであれば、30 分で最も効率の良い検索ができるスケジューリングが、これにより可能となる。この場合、TS_{inf} の値が大きいページから順に抽出し、情報認識時間の総和が 30 分を超えた時点で出力を中止する。

3.3 スケジューリングの具体例

スケジューリングの具体例を図 1 に示す。本稿では、情報価値の要素として、検索対象ページより、

・アクセス頻度
・更新時刻と現時刻の差
・更新頻度

の 3 点を抜き出している。

利用者の時間制約 T_{max} を 30 とし、TS_{inf} の求め方を $TS_{inf}{}^{1} = S_{inf}/T_{inf}$ と定めてこのスケジューリングを実行した結果を表 1 に示す。TS_{inf} の値の大きい順に、しかも、情報認識時間の総和が利用者の時間制約を超えた時まで抽出しているため、ページ a～e が利用者に対して出力されることになる。これにより、あるページが従来の情報価値によるソーティングにより価値が低いと判断されても、情報認識時間が少なくてすむならリストに出力される可能性は高くなる。

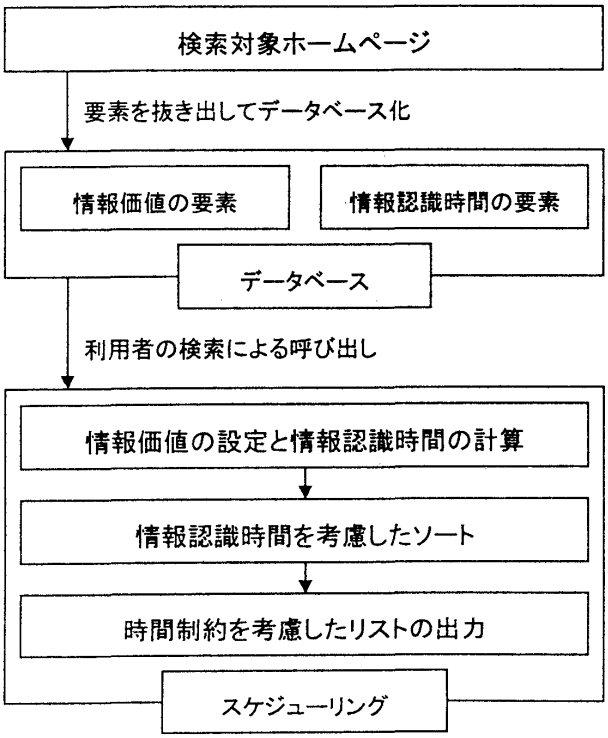

図 1　スケジューリングの具体例

表 1　スケジューリングの結果

ページ	S_{inf}	T_{inf}	$TS_{inf}{}^{1}$
http://a/	0.8	4	0.2
http://b/	0.96	8	0.12
http://c/	0.6	5	0.12
http://d/	0.4	4	0.1
http://e/	0.56	7	0.08
http://f/	0.12	3	0.04
http://g/	0.05	5	0.01
http://h/	0.024	3	0.008
http://i/	0.035	7	0.005

4. おわりに

今後、TS_{inf} の求め方や情報認識時間 T_{inf} の算出方法に関して、妥当性、有効性の面から検討していきたい。さらに、情報認識時間を求める要素数を増加することによって、正確な時間算出に発展していきたい。

参考文献

[1] 灰原清太郎 柏井優希 川越恭二「サイト評価情報を用いたWWW検索表示方法」本会第 58 回全国大会 3T-07 1999