アクセスログを利用したアクセスパターンの解析の研究

98H134 福井 隆幹

1 はじめに

今日、学内ネットワークを構築している教育施設は多く、情報化に伴い学内ネットワークを利用し学生に対して教材として Web コンテンツを作成する教育者もいる。

Web コンテンツの作成者にとって、閲覧者が製作者 (ここでは教育者がコンテンツを製作すると考えて製作者 = 教育者と考える) の意図する順番に沿ってコンテンツを閲覧することは重要な問題である。例えば、基本問題から応用問題へと製作者がアクセスしてもらうように考えていたとしても、閲覧者が応用問題から基本問題へと逆の順番に閲覧した場合、応用問題が理解しづらいことによってページを読み進めることを途中で諦めてしまうということが起こる可能性がある。

そこで、Web コンテンツの製作者は閲覧者がどのような順番でアクセスしているかを知り、製作者の意図しないアクセスが多い場合にはコンテンツを改良する必要がある。アクセスを調べる方法として、アクセス数を数えるアクセスカウンタがあるが、この場合には単純にアクセスされた回数のみを表示するためにどのようにアクセスしたかという順番までは解らない。しかし、Web コンテンツのアクセスの際に Web サーバに保存されるアクセスログを利用する方法がある。この方法はアクセスログを解析することによってアクセスパターン (ページをアクセスした順番) を抽出し、抽出したアクセスパターンと製作者が意図している閲覧するべき順番とを比べることによって製作者の意図通り閲覧されているか解る。意図通りでない場合には、抽出したパターンを参考にしてコンテンツの改良を行うことが出来る。以上を踏まえて、本研究ではアクセスログの解析手法を確立する。

2 研究内容

上記の通り、閲覧者のクライアントマシンが Web コンテンツのアクセスの際に Web サーバのディスクに保存されるアクセスログを利用する。

従来のアクセスログ解析ソフトでは各ページのアクセスの回数、時間別アクセス数、アクセス元ごとの回数を表示するものが多く、アクセスパターンを表示する、かつ閲覧者のアクセスパターンと製作者の意図するアクセスパターンを比較するソフトが多くは存在しなかった。そこで、Web コンテンツ製作者を支援する為のアクセスログ解析手法を確立した。

この手法では Web コンテンツ製作者が意図しているアクセスパターンをファイルに書いておき、アクセスログを解析した結果とを比較する。図??はページをアクセスしている様子で、円はページを表し、矢印が閲覧者のアクセスパターンを表している。Web コンテンツ製作者の意図通りのアクセスパターンを (A) とする。アクセスパターンの途中でアクセスパターンから外れてアクセスしたものを (B、C) とする。このとき、式 (1) の割合が大きければ、意図通りのアクセスをする閲覧者が多いことになるが、低くなるにつれてコンテンツのガイドが悪いということが解り、改良する目安となる。

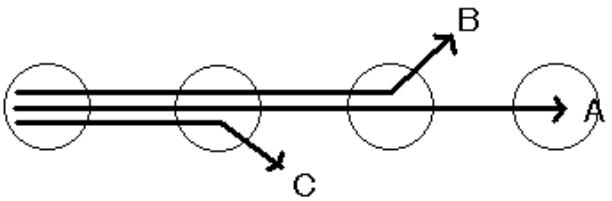

図 1: ページをアクセスしている様子

$$\frac{A}{A+B+C} \times \ 100(\%) \qquad (1)$$

また、閲覧者のアクセスパターンの統計を取ることによって、どの程度の閲覧者が、どのページから Web コンテンツ製作者が意図しているアクセスパターンから外れたか、またはどのページから沿って閲覧したかを調べることが出来る。それによって、改良すべきコンテンツがどこかも解る。

3 まとめ

本研究で従来のアクセス解析ソフトを利用し、Web コンテンツ製作者が行なってきた各ページの閲覧数だけで Web コンテンツを改良する方法より、アクセスパターンを提示することにより、Web コンテンツ製作者の負担が軽減されると期待できる。しかし、改良する際に推奨するパターンの提示やコンテンツの自動改良があれば、Web コンテンツ製作にとって更なる負担の軽減が期待できる。これについては今後の課題である。