

WWWにおけるユーザの興味対象と閲覧時間の関係の調査

5L-6

清水 勇喜（静岡大学大学院理工学研究科）
s3053@cs.inf.shizuoka.ac.jp

1. はじめに

コンピュータユーザの興味情報をコンピュータシステムが知ることは、ユーザの立場にたったシステムを作成する上で重要である。興味情報を得ることができれば、興味のある分野、興味の無い分野のそれぞれの情報をユーザに提示したり、コンピュータシステムがその情報に従って動作することによって、執筆、推考、検索、発想などの知的活動を支援することが容易になる。また個人的な興味情報を利用した新しいコミュニケーションシステムの開発なども可能になる。

さて、このような興味情報をシステムが取得するにはどのようにすればよいただろうか。第一に考えられるのは、興味情報をユーザに入力してもらう方法である。この方法はシステムにとってはコストがかからず簡単である。しかし、一人のユーザがコンピュータに入力する情報の量は少なく、興味情報を抽出するには十分ではない。

また、近年のパソコンの社会への普及を考慮すると、これらの興味情報を得るために、ユーザに特別な操作を強いるのは望ましくない。

そこで次に、ユーザに出力される側の情報に着目してみる。スタンドアローンのコンピュータの場合はユーザの入力した情報、外部記憶媒体から供給される情報のみが出力される情報となり、興味情報を得るための情報量は多くない。一方、ネットワークが利用可能なコンピュータでは、ネットワークから大量の情報を出力することができる。しかもインターネットにはWWWによる情報が大量に存在し、その情報

は日々更新されている。これらの情報と連動する形で興味情報を抽出することができれば、ユーザの利便性も高い。

そこで私はWWWブラウザの閲覧時間からユーザの興味情報を抽出する手法を提案したい。

2. 経験則

ユーザの興味対象と、WWWブラウザでの閲覧時間との間にある経験則を以下に示す。この経験則はWWWのユーザへのインタビューと、私自身の経験から考察された仮説である。

1. ユーザの興味のないページの閲覧時間は短く、興味のあるページの閲覧時間は長い。
2. 情報量の多いページの閲覧時間は長く、情報量の少ないページの閲覧時間は短い。
3. ユーザはページの概要をつかむためにまず、斜め読みをする。よって項目2.の時間と情報量の関係は単純な比例関係にはならない。
4. 以前に見た内容と同じ内容のページの閲覧時間は短くなる。

3. 実験方法

経験則が実際に正しいかどうかを検証するために以下の手順で実験を行った。なお、紙面の都合により、以下では経験則1.の検証を行っている。（経験則1.以外の検証結果については文末のアドレスで公開している）

まず、学内から被験者を募集し実験に協力してもらった。ユーザである被験者の興味対象を実験者が知るために、被験者にはテーマに従ってレポートを書いてもらい、そのテーマにつ

いて WWW で調べてもらうことにした。この時、以下の 1.から 3.については、ビデオでディスプレイを撮影し、同時に、WWW へのアクセス履歴及びアクセス時間をプロキシサーバによって記録する(図 2)。なお、手順 3.は WWW を閲覧したことによるユーザーの興味の変化を知るために行う。

3.1. 実験手順

1. まず、各自自由に興味のあるテーマを決めてもらい、そのテーマに従ってレポートを書いてもらう。[20 分]
2. レポートのテーマに関して WWW で調べてもらう。[20 分]
3. もう一度、最初からテーマに従ってレポートを書いてもらう。[20 分]
4. 2.で撮影した WWW の閲覧時のビデオをみてもらい、a.大変興味があったページ、b.少し興味があったページ、c.興味がなかったページに分類してもらう。

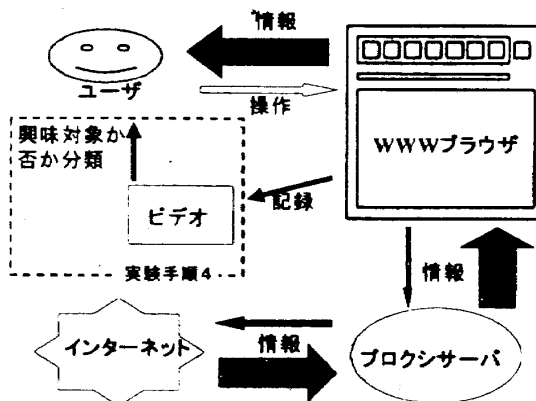


図 1 実験環境

3.2. 閲覧時間の測定方法

WWW ブラウザは同時に複数のウィンドウで閲覧ができる。しかし、ユーザーは画面の一番手前でアクティブになっているウィンドウを注視しているはずである。よって、アクティブウィンドウに WWW からの情報が表示されている時間を測定し、この時間を閲覧時間と定義する。

ブラウザの操作では「戻る」ボタンなどで過去に表示されたページを再度表示することが多い。この場合には、以前に表示していた時間に

加算する。

URL の入力やフォームへの文字の入力にはかなり時間がかかるため、URL やフォームの入力個所にカーソルが入っている時間は閲覧時間に含めないこととする。

3.3. 実験結果

現在までに 6 名の被験者に対して実験を行った。実験結果を図 2 に示す。

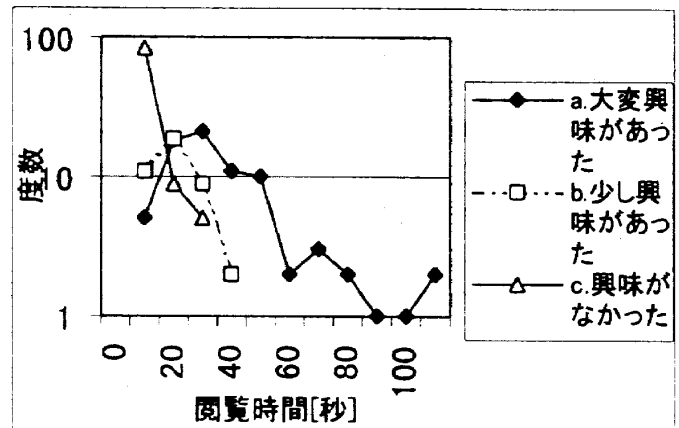


図 2 ユーザの興味別の閲覧時間の度数分布

なお閲覧時間の平均値はそれぞれ、a.40.7 秒、b.14.9 秒、c.5.4 秒であり、全体の平均値は 19.9 秒である。

図 2 では興味の数合いに応じて閲覧時間に明らかな偏りが見られる。興味の数合いが上がるにつれて閲覧時間も長くなっている。全体の平均値付近で興味のありなしを分けることが可能であり、a.の平均値以上の閲覧時間のページを興味のあるページ、c.の平均値以下の閲覧時間のページを興味のないページとみなすことが出来そうである。

4. 最後に

今後は本研究の結果を利用した応用システムを作成して、仮説の検証を進める予定である。閲覧時間は、OS のメッセージとプロキシサーバへのリクエストを監視し、自動的に取得する予定である。本研究の最新の状況は <http://www.pipi.net/study>にて公開する。より良いご提案などがあればお気軽にご連絡いただきたい。