

Web サイト改良支援のための構造と履歴の視覚化システム*

熊谷 知久

東京大学 工学部 機械情報工学科

寺田 実

東京大学大学院 情報理工学系研究科

1 概要

Web サイトに対して、木構造¹に基づく構造画像、アクセスログの対話的閲覧、サイト改良案画像を提供してサイト管理者の管理作業を支援するツールを開発した。

視覚化画面におけるマウス操作やメニューからの選択によって、構造、アクセスログ、改良案を様々な角度から見るができるようになっている。

2 背景と目的

Web サイトの運営は、そのコンテンツを少しずつ更新、追加をしていくという形が一般的である。しかし、こういった部分的な修正を無計画に重ねていくと、サイトはその構造的性を失ってしまいかねない。これを避けるためにサイト管理者は常に自分のサイトの構造を把握しておき、その構造的性に沿った修正をする必要がある。また、サイト運営を別の人から引き継ぐ者はそのサイトの構造やアクセス状況を新たに把握しなくてはならない。

更に、サイトにおける閲覧者のリクエストはサーバにアクセスログとして記録されるが、時系列でリクエストを記録しただけのテキストのみで構成されるアクセスログをそのまま読んでも殆ど意味がない。

このため、サイト管理者にとって必要となるサイトの全体構造の把握支援、アクセスログの理解、活用支援、全体構造的性の再付与支援を行うために以下のような機能を持つシステムを開発した。

- 木構造に基づくサイト全体構造の図示
- サイト構造図上での閲覧者行動の図示
- 閲覧者行動に基づくサイト改良案の図示

3 関連研究

VISVIP[1] などグラフ構造を利用して閲覧者行動を視覚化する研究は数多くなされている。

Adobe 社の GoLive[2] など一部の商用 Web オールシングツールにおいては親子関係に基づく Web サイト全体構造の視覚化機能を備えている。しかし、閲覧者行動を視覚化する機能は持っていない。

*Visualization of Site Structure and Access Patterns to Support Web Site Restructuring, Tomohisa Kumagai and Minoru Terada, Department of Mechano-informatics, School of Engineering, University of Tokyo.

¹ 多くの Web デザインマニュアルにおいて推奨されている

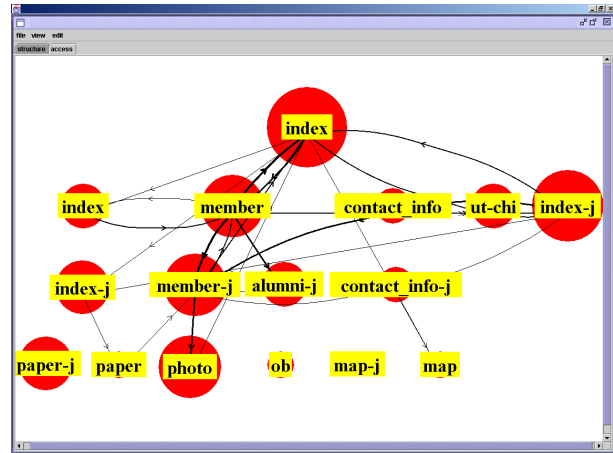


図 1: アクセスログの視覚化例 (ノードが Web ページ, エッジがページ間閲覧者移動, ノード背後の円の面積がページリクエスト数を示している.)

アクセスログ解析ソフトや Web サーバ運営会社が提供するアクセスログ解析結果は、統計量を示したり、それをグラフ化したものが殆んどである。しかし、中には Microsoft 社の Site Server[3] のように自身が Web サーバとして記録したアクセスログを構造と結び付いた形で視覚化するサーバプログラムもある。

4 システムの特徴

4.1 入力の手軽化

対象サイトの URL, またはファイルパスを入力するだけでサイト構造の画像を生成し、アクセスログについても、ファイルパスと見たい時期を指定するだけで閲覧者行動画像を生成することができる。4.3 節で述べる連続画像表示も、見たい期間を入力するだけで良い。

但し、アクセスログの形式によっては、IP アドレスなどの各要素を抽出する正規表現は新たに書く必要がある。

4.2 構造の視覚化

ページをノードとして木構造状に配置する。エッジとして、すべてのページ間リンクを表示するか、親子関係

を示す線のみを表示するかはユーザが選択できる。ユーザはノードをドラッグして見やすい位置に移動させることができる。また、マウスを近付けることによってノードを拡大したり、そのノードに繋がるエッジのみを表示させることも可能である。

ユーザはサイト構造全体を一度に見ることができるようになり、ユーザが構造を把握する手助けになる。また、サイトを再構成することが容易になる。

4.3 閲覧者行動の視覚化

4.2 節と同様にページをノードとして木構造状に配置する。エッジは閲覧者移動とし、その太さには移動数を反映する。サイト内だけでなく外部からの流入が多い経路も示す。画面上で指定したノードへ至る経路、または指定したノードから出る経路を限定して表示することができる。ユーザは構造と共に閲覧者移動経路を見ることで、移動経路の出元と行き先のノードの示すページのサイト構造全体における位置付けや出元と行き先のページの親子関係、兄弟関係などの構造的な関係を直感的に知ることができる。

そして、閲覧者移動数以外の種々の閲覧者行動量²をユーザが選択した方式³で視覚化する。更に、これらを月、日、時間別に表し、連続的に表示することもできる。

4.4 改良案の視覚化

閲覧者行動量それぞれに対して、これを満たさなければ改良した方がよいという条件を設定し、条件を満たさない閲覧者行動量に関わる部分をサイト構造画像上に強調して表示する。その多くは閾値によって設定される条件であり、この値をユーザは設定することができる。ユーザは構造と共に改良案を見ることにより、改良案の採否や改良手法の検討などを効果的に行うことができる。

例えば、リクエストが極端に少ないページを知ることにより、そのページを消去することや、逆にデザインにおいてそのページへのリンクを強調することができる。また、相関度が高いページの組であるにもかかわらず、ページ間にリンクがなければリンクを加えることができる。

4.5 その他の機能

ミニブラウザと簡易エディタを備え、選択されたファイルを編集できる。また、全体の構造のうち一部だけを示す部分表示機能、ファイルからの読み込み機能、拡大縮小機能、ページ表現形式の選択機能、エッジ形状の選択機能、連続画像再生制御機能などを備える。

² ページリクエスト数、閲覧者のページ間移動数、閲覧者消費時間、ページ A を見たあと、ページ B を見る確率であるページ AB の相関度、ページのサイトの出口または入口に使われた回数など。

³ 例えばページに関する量ならば、ノード横にその高さに量が反映された柱を示す方式やノード背後にその面積に量が反映された円を示す方式など。

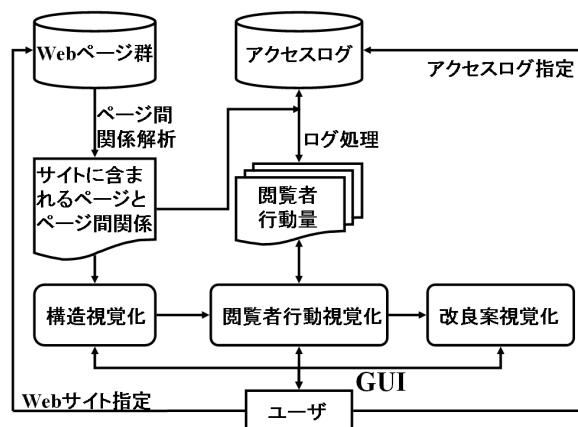


図 2: システム構成図

5 システムの構成

ユーザは対象サイトの URL、またはファイルパスを指定する。URL 指定の場合には、サイトからファイルのダウンロードを行う。サイトを構成するファイルからリンクタグの抽出を行い、ページ間のリンク関係を出力する。また、全体のリンク関係から親子関係を導く。

その親子関係をもとに木構造としてサイトを視覚化する。また、アクセスログから算出された閲覧者行動量も視覚化する。更に、4.4 節で述べた改良案も画像として提示する。

ページ間関係解析とアクセスログ処理は Perl で、視覚化部分は Java で実装している。URL で指定したときのダウンロードは Wget^[4]で行っている。

6 評価

現在のところ、まだ客観的な評価は行っていない。様々な画像による表現手法の評価と、ユーザに対する効果の評価を行う予定である。

参考文献

- [1] J. Cugini and J. Scholtz, "VISVIP: 3D Visualization of Paths through Web Site", Proceedings of the International Workshop on Web-Based Information Visualization (WebVis'99), pp. 259-263, 1999.
- [2] Adobe GoLive
<http://www.adobe.com/products/golive/main.html>.
- [3] Microsoft Site Server
<http://www.microsoft.com/siteserver/>
- [4] Wget
<http://www.gnu.org/software/wget/wget.html>