

複数 Web コンテンツの多面的閲覧のための空間インタフェース

小谷 彬[†] 小山 聡^{††} 田中 克己^{††}

[†] 京都大学工学部情報学科

〒 606-8501 京都府京都市左京区吉田本町

^{††} 京都大学大学院情報学研究科社会情報学専攻

〒 606-8501 京都府京都市左京区吉田本町

E-mail: †kotani@dl.kuis.kyoto-u.ac.jp, ††{oyama,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 通常、ひとつの Web サイトの内容は多面的であり、複数の側面からの記述がある。しかし、現在の検索エンジンやポータルサイトのインタフェースでは、多数の Web サイトの様々な側面を比較しながら閲覧することは容易ではない。そこで我々は、多数の Web サイトの異なった側面を空間的に配置することで、Web の多面的な閲覧を可能にするインタフェースを提案する。ユーザは三次元のこのインタフェースの中でウォークスルーをしてブラウジングを行う。このインターフェースにおいて、ひとつのサイトはブロックと呼ばれる直方体で表わされ、ブロックの各々の面にそのサイトの各ページが割り当てられる。これにより、ひとつの Web サイトを空間的な角度によってさまざまな側面から閲覧することが可能となる。本論文では、多面的な閲覧を容易にするための複数 Web サイトの空間配置の指針と、自動配置を行う手法について述べる。

キーワード データの可視化 Web マイニング Web 利用技術

A Spatial Interface for Aspect Browsing of Multiple Web Contents

Akira KOTANI[†], Satoshi OYAMA^{††}, and Katsumi TANAKA^{††}

[†] School of Informatics, Kyoto University

Yoshidahonmati, Sakyou-ku, Kyoto, 606-8501 Japan

^{††} Department of Social Informatics, Graduate School of Informatics, Kyoto University

Yoshidahonmati, Sakyou-ku, Kyoto 606-8501 Japan

E-mail: †kotani@dl.kuis.kyoto-u.ac.jp, ††{oyama,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract Usually the content of one Web site is multi-pronged, and there is a description from different viewpoints. However, in the interface of present search engines or portal sites it is not easy to browse Web pages while comparing various sides of many Web sites. Then, we propose the interface that enables multi-pronged browsing of Web contents by spatially arranging different sides of many Web sites. Users browse Web sites and Web pages walking through in this three dimensional interface. Each site is represented as a hexahedron that is called a block, and each page of the site is assigned to each surface of the block in this interface. Our system enables us to browse various sides of one Web site from spatial angle. In this paper, we describe the guideline for the space arrangement of multiple Web sites to facilitate multi-pronged browsing and the technique for automatically arranging Web sites in three dimensional space.

Key words Data Visualization, Web Mining, Web Technology

1. はじめに

近年インターネットの普及が急速に広がったことにより、Web サイトの量も爆発的に増え続けている。それら Web サイトの中でもニュースサイトや大学のサイトなどに代表されるように、類似した構成を持つ複数の Web サイトも存在する。例えば、多

くの大学のサイトには入学案内のページや研究のページなどがある。また、多くの大学の研究室のサイトにはメンバーのページや論文のページがある。そういった Web サイトは、複数の側面からの記述があり、Web サイトの内容が多面的であると言える。同時に、共通の側面をもつ似た構成の複数の Web サイトが存在していると言える。しかしながら、ユーザがそれら共

通の側面を持つ複数の類似した Web サイトを、ある一定の側面から連続的に比較閲覧することや、複数サイト間の同側面のページを発見することは、現在の検索エンジンやポータルサイトのインタフェースでは容易ではない。

そこで我々は、多数の Web サイトの複数の異なった側面を空間的に配置することで、Web の多面的な閲覧を可能にするインタフェースを提案している。ユーザは三次元のこのインタフェースの中でウォークスルーをしてブラウジングを行う。具体的にはひとつのサイトをブロックと呼ばれる直方体で表し、直方体の各々の面に、そのサイトの各側面（各ページ）を貼り付けるものである。これにより、ひとつのサイトを空間的なアングルによってさまざまな側面から閲覧することが可能となる。各サイトの各側面の抽出には、アンカー文字列や見出しタグなどに着目してキーワードを抽出し、そのキーワードを元に側面を抽出する手法を提案している。ここではサイト内における各ページの重要度という概念を導入し、重要度を算出し側面に相当するページの選択に用いている。

また、ブロックとして表されるサイトを複数並べることにより多面的な閲覧を支援することを目的としている。その多面的な連続的比較閲覧を容易にするための空間配置の指針として、格子状配置を提案している。そしてよりよい空間配置を行うために、空間配置に対して指針を設定し、その指針に対する現在の配置の適合度を表す評価値を設定する。この評価値の高い配置すなわち空間配置指針に適った配置を求めるために、ある問題に対する最適解の探索法のひとつである遺伝的アルゴリズム (Genetic Algorithm) の考え方をを用いた。初期配置をランダムで与えて、配置に対する選択・交叉・突然変異の操作を設定し、これらを繰り返すことにより評価値の高い配置を探索し再配置するという自動配置手法を提案している。

本論文では、このインタフェースを実現するために、

- Web ページの側面抽出手法
- 空間配置の指針
- 自動配置手法

の 3 つに問題を大別し、それぞれに関して手法の提案・検討を行い、提案手法の有用性を検証するために部分的に実験を行った。それにより一定の有用性が認められ、またさらなる改善が必要な点も整理した。

最後に、インタフェース全体のプロトタイプを製作し、同様に有用性を検証し今後の方針を示した。

本論文では、第 2 章ではこれらを考察する上で必要となる基本的事項および関連研究について述べ、第 3 章ではイメージ図・構成要素などインタフェース全体の外観について述べ、第 4 章では Web ページの側面抽出手法について述べ提案手法による実験を行い、第 5 章では空間配置の指針を明らかにし、第 6 章では自動配置手法について述べ提案手法による実験を行い、第 7 章ではインタフェース全体のプロトタイプ作成とそれに対する考察を述べ、第 8 章で本論文の結論を述べている。

2. 関連研究

2.1 INFOTUBE

INFOTUBE [1] は実際の商店街の Web ページひとつひとつをセルとして扱い、それを三次元空間のチューブに貼ることで、複数 Web サイトの複数 Web ページを空間的に表示している。複数 Web コンテンツ表示の空間インタフェースという点で本研究と共通するが、本研究では複数コンテンツの一定の側面からの閲覧を支援するために配置に意味を持たせるが、INFOTUBE におけるセルの位置はランダムである。

2.2 デジタルシティ京都

デジタルシティ京都 [2] は、実際の都市を Web サイトの情報など埋め込んで仮想空間に構築したものであり、実時間性を持ち、実際の都市の情報に応じてリアルタイムに再構成される。Web サイトの情報を埋め込んだ空間インタフェースという点で本研究と共通するが、本研究では現実空間との対応ではなく、Web 空間を実空間的なインタフェースで表すことが目的である。また、デジタルシティ京都では、3D 京都として 3DML [3] を用いて京都の街を立体的に再現している。本システムの実装に際しても同様に 3DML を使用した。

2.3 City of news

City of news [4] は、K. Lynch が The Image of the City [5] で提唱した認知地図による都市計画に基づいた都市メタファで構成された Web 仮想空間で、ユーザの Web 探索などに応じて有機的にその構成を成長させたり、現代の新聞のレイアウトをマッピングしたり、ウェアラブルなインタフェースを用いてのウォーク・フライスルーができる。本研究でも、最終的には都市メタファで表された空間インタフェースを目的としているが、Web ページの側面という点に特に着目して空間配置を行う点で異なる。

2.4 Focus + Context Views of World-Wide Web Nodes

Focus + Context Views of World-Wide Web Nodes [6] は、Web 空間で「迷子」しないために、現在見ているページの Web 空間における位置を提示するシステムである。この中で、Web サイト内の Web ページに *importance* を設定し、*importance* の高いものをランドマークとすることにより、現在見ているページの空間位置の把握を支援している。本研究では Web サイト内の側面抽出において、各 Web ページの重要度を設定する際にこの考えを用いた。

3. インタフェース概観

3.1 イメージ図

提案する空間インタフェースは街をイメージしており、街内の建物が Web サイトに対応し、ユーザはこの街内をウォークスルーしながら Web ブラウジングを行う。図 1 がこのインタフェースのイメージ図である。

3.2 空間インタフェースの構成要素

提案する空間インタフェースにおいて、ひとつのサイトを立方体が 2 段積み上げられたひとつの直方体で表す。これを「ブ

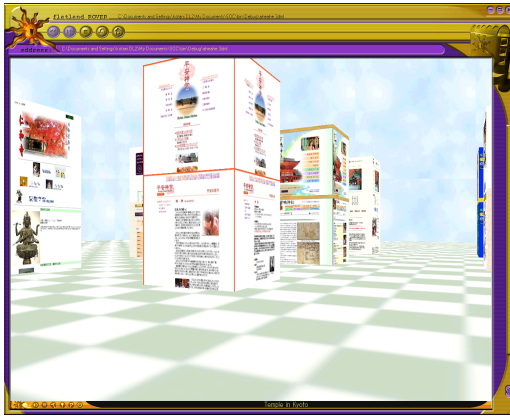


図1 インタフェース概観

ロック」と呼ぶことにする。

ブロックは直方体で表され、ひとつのサイトがひとつのブロックに対応する。上段の立方体にはサイトのトップページが割り当てられ、下段の立方体の上面と底面を除く4つの側面には、サイトの各側面が割り当てられる。

この空間インタフェース上に配置する複数の類似サイト集合を $S = \{s_1, s_2, \dots, s_n\}$ とする。さらに S の各要素 $s_k \in S$ に対して、そのサイトのページ群集合を $P_k = \{p_{k1}, p_{k2}, \dots, p_{km}\}$ とする。

空間インタフェース上に配置されるブロックの集合を $X = \{x_1, x_2, \dots, x_n\}$ とする。各ブロックに S の各要素であるサイト s_k を割り当てる。このとき以下のような写像で表される。

$$S \rightarrow X$$

ブロック x_k は直方体であり上底面以外の東西南北に面を持ち、これらの面を $x_k^n, x_k^e, x_k^w, x_k^s$ で表す。各面に、そのブロックに割り当てられているサイト s_k の P_k の各要素であるページ $p_{k1}, p_{k2}, \dots, p_{km}$ を割り当てる。面にはテクスチャとして該当ページのスクリーンショットを貼る。 $m < 4$ (elements number of $m : P_k$) のときは、同一のページが複数の面に割り当てられる。 $m > 4$ のときは、いずれの面にも割り当てられないページが存在する。これを p_{ki} を *null* に割り当てると表すこととする。このとき、以下のような写像で表される。

$$P_k \rightarrow \{x_k^n, x_k^e, x_k^w, x_k^s, null\}$$

3.3 問題

この空間インタフェースを実現する上において、3つの問題に大別できる。

- Web ページの側面抽出手法
- 空間配置の指針
- 自動配置手法

これらを以降の、それぞれ4, 5, 6章で述べる。

4. Web サイトの側面抽出手法

本章では Web サイトの側面抽出手法の検討と提案手法によ

る実験および評価を述べる。

4.1 トップページ利用

Web サイトのトップページに存在するリンク先のページを側面として、それらをページ群とする。Web サイトの構成は一定ではなく、必ずしもトップページにすべてのページへのリンクがあるとは限らず、適当なページを取得できないことも考えられる。

4.2 メニューページ、サイトマップページの利用

ページ Web サイトには、メニューページやサイトマップページを持つものが多く、それらに列挙されているリンク先のページを側面として、それらをページ群とする。メニューページやサイトマップページは URL から推測できるものが多いと考えられる。メニューページやサイトマップページを持たないサイトもまた多く、存在しても URL から推測できない場合もある。

他の試みとしては、Web サイトの全ページを取得して内容・構造解析し一定の形式のサイトマップを自動で作成し、それをを用いる方法も考えられる。

4.3 キーワードによる側面の抽出

対象とする複数 Web サイトの全ページの内容を対象として解析しキーワードを抽出し、キーワードにより各ページを特徴付けて側面の抽出に用いる方法が考えられる。

手順は以下ようになる。

- (1) 対象とする複数 Web サイトの全ページから面となるキーワードを、タイトルタグや見出しタグなどをもとにピックアップする (キーワード候補)
- (2) ピックアップされたキーワード候補に対し、そのキーワードを含むページを持つサイトの割合が高い数個をキーワードとする。
- (3) 各ページの各サイト内における重要度を設定する。
- (4) 各サイトにおいて、キーワードを含むページのうち最もサイト内重要度が高いページを側面とする。
- (5) 採用する面の決め方は、後に述べる評価関数を用いた最適化、もしくは面候補をユーザに提示して選択させる。

以下でキーワードの抽出方法、各ページのサイト内における重要度の算出を述べる。

4.3.1 キーワードの抽出

キーワード候補ならびにキーワードは以下のように抽出する。 $p_{kx} \ p_{ky} \ p_{ki}$ この空間インタフェース上に配置する複数の類似サイト集合を $S = \{s_1, s_2, \dots, s_n\}$ とする。さらに S の各要素 $s_k \in S$ に対して、そのサイトのページ群集合を $P_k = \{p_{k1}, p_{k2}, \dots, p_{km}\}$ とする。そして P の全要素について、そのページへのアンカー文字列、タイトル文字列、meta タグ内の keywords 属性の文字列、ページ内の見出しタグ (<h>) および強調タグ (,) で囲まれた文字列を形態素解析し単語に分割して、それらのうちストップワードなどを除いたものをキーワード候補とする。

これらキーワード候補集合を $K = \{k_1, k_2, \dots, k_l\}$ とし、 k_i を含むページを持つサイトの割合をキーワードレベルとする。 $(0 < level(k_i) \leq 1)$ 。

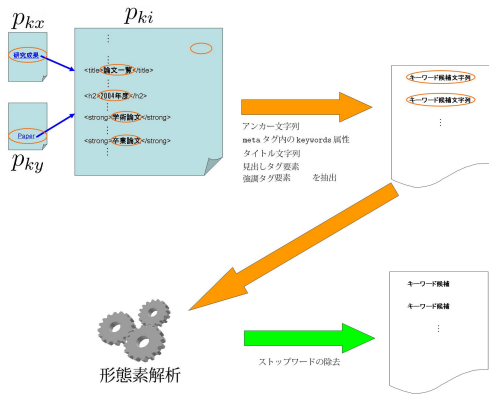


図 2 キーワード候補の抽出

キーワードレベルが大きい数個 (4 つ) のキーワード候補が、キーワードとなる。

4.3.2 各ページのサイト内における重要度

各ページに対して重要度 (*importance*) を設定する。重要度は *connectivity, access frequency, depth* の 3 つのパラメータから算出されるものとする。

(1) 接続度 *connectivity*

より重要なページは、多くのページからリンクが貼られており、また多くのページへのリンクがあると考えられる。

Indegree(n) : 該当ページから *n* ステップでたどり着けるページの数

Outdegree(n) : 該当ページへ *n* ステップでたどり着けるページの数

と表すこととする。

(2) アクセス頻度 *access frequency*

より重要なページは、Web サイトのアクセスカウンタにも表されるようにアクセス頻度が高くなると考えられる。これはログとして保存する。

(3) 深さ *depth*

より重要なページは、トップページからそのページまでのリンクステップ回数は小さくなると考えられる。

(4) *importance*

ページの重要度を算出する上において、まずリンク構造のみに注目した構造的な重要度 *connectivity importance* を算出する。

$$\begin{aligned} & \text{connectivity importance} \\ &= (\text{Indegree}(1) + \text{Outdegree}(1))w_1 \\ &+ (\text{Indegree}(2) + \text{Outdegree}(2))w_2 \\ & (w_1 + w_2 = 1) \end{aligned}$$

w_1, w_2 はそれぞれ重みパラメータで総和が 1 となるように設定する。2 ステップで移動できる範囲のリンク構造をもとに *connectivity importance* を算出する。そして、重要度 *importance* を次式で算出する。

$$\begin{aligned} \text{importance} &= \frac{\text{connectivity importance}}{\text{MAX}(\text{connectivity importance})}w_s \\ &+ \frac{A}{\text{MAX}(A)}w_a + \frac{1}{D}w_d \end{aligned}$$

$$(w_c + w_a + w_d = 1)$$

w_c, w_a, w_d はそれぞれ重みパラメータで総和が 1 となるように設定する。*importance* の値域は $0 \leq \text{importance} \leq 1$ となる。

重要度を加味することにより、より重要なページを側面として選ぶことができる。

4.4 キーワードによる側面抽出の準備実験

キーワードによる側面抽出の準備実験を行った。11 の京都大学情報学研究所の研究室の Web サイトを対象とし、キーワード抽出、重要度の算出を行った。

表 1 に対象とした Web サイト一覧を挙げた。

研究室	Web サイトドメイン
社会情報モデル講座 情報図書館学分野	www.dl.kuis.kyoto-u.ac.jp
社会情報モデル講座 分散情報システム分野	www.isse.kuis.kyoto-u.ac.jp
社会情報モデル講座 情報教育環境分野	www.edu.soc.i.kyoto-u.ac.jp
社会情報ネットワーク講座 広域情報ネットワーク分野	www.lab7.kuis.kyoto-u.ac.jp
生物圏情報学講座	bre.soc.i.kyoto-u.ac.jp
知能情報ソフトウェア講座 知能情報応用論分野	www.ii.ist.i.kyoto-u.ac.jp
知能情報ソフトウェア講座 ソフトウェア基礎論分野	www.sato.kuis.kyoto-u.ac.jp
メディア応用講座 映像メディア分野	www.imel1.kuis.kyoto-u.ac.jp
コンピュータ工学講座 論理回路分野	www.lab2.kuis.kyoto-u.ac.jp
コンピュータ工学講座 計算機アーキテクチャー分野	www.lab3.kuis.kyoto-u.ac.jp
コンピュータ工学講座 計算機ソフトウェア分野	www.yuasa.kuis.kyoto-u.ac.jp

表 1 対象とした Web サイト

表 2 は、表 1 のサイト内のページに対してキーワード候補を抽出したものである。出現率はそのキーワード候補を持つページが存在するサイトの割合である。キーワード候補の中でも出現率が高いものがキーワードとして有効であると考えられる。

表 3 は、社会情報学専攻社会情報モデル講座情報図書館学分野 (www.dl.kuis.kyoto-u.ac.jp) の Web サイトのページに対し、重要度を算出し高い順に数ページを並べたものである。今回重要度の算出に *access frequency* は考慮していない。また $w_1 = w_2 = w_c = w_d = 0.5$ とした。

表 4 では、キーワード出現率の高かった「教授」という単語を含むページのうち、各サイト内で重要度が最も高かったものを挙げている。項目の適合には、当該研究室の教授のページである場合は、そうでない場合は \times を割り当てた。

表 5 では、キーワード出現率の高かった「メンバー」という単語を含むページのうち、各サイト内で重要度が最も高かったものを挙げている。項目の適合には、当該研究室のメンバー一覧のページである場合は、そうでない場合は \times を割り当てた。

例えば「member」と「メンバー」は同義だが別の単語として扱われてしまうので、類義語をまとめてキーワードとすると精度が上がるものと考えられる。

5. 空間配置の指針

5.1 格子状配置

図 4 のようにブロックを格子状に並べる。ユーザはこのブ

出現率	キーワード候補
1.00	テーマ
1.00	研究
1.00	リンク
1.00	情報
0.91	紹介
0.91	年度
0.91	講義
0.82	活動
0.82	発表
0.82	京都大学
0.82	論文
0.73	メンバー
0.73	概要
0.73	教授

表 2

重要度	ページ
1.00	/photo/photo.html
0.85	/paper.html
0.68	/member.html
0.58	/link.html
0.57	/access.html
0.57	/outline.html
0.44	/info.html
0.34	/infosystem.php
0.34	/UDM2005/

表 3

Web サイトドメイン	ページ	適合
www.dl.kuis.kyoto-u.ac.jp	/member.html	×
www.isse.kuis.kyoto-u.ac.jp	/usr/yahiko/yahiko.html	-
www.edu.soc.i.kyoto-u.ac.jp	-	-
www.lab7.kuis.kyoto-u.ac.jp	/ishida/indexj.htm	-
bre.soc.i.kyoto-u.ac.jp	/moriya/souran.htm	-
www.ii.ist.i.kyoto-u.ac.jp	/nishida/index-j.html	-
www.sato.kuis.kyoto-u.ac.jp	-	-
www.lab2.kuis.kyoto-u.ac.jp	/iwama/SOURAN/iwama.html	-
www.lab3.kuis.kyoto-u.ac.jp	/members/tomita/home.html	-
www.yuasa.kuis.kyoto-u.ac.jp	-	-

表 4 「教授」を含むページのうちサイト内重要度が最も高かったページ

Web サイトドメイン	ページ	適合
www.dl.kuis.kyoto-u.ac.jp	-	-
www.isse.kuis.kyoto-u.ac.jp	/kambmemb.html	-
www.edu.soc.i.kyoto-u.ac.jp	/members.php	-
www.lab7.kuis.kyoto-u.ac.jp	people_j.html	-
bre.soc.i.kyoto-u.ac.jp	-	-
www.ii.ist.i.kyoto-u.ac.jp	/member/index-jp.html	-
www.sato.kuis.kyoto-u.ac.jp	/member/index-j.html	-
www.lab2.kuis.kyoto-u.ac.jp	-	-
www.lab3.kuis.kyoto-u.ac.jp	/members/	-
www.yuasa.kuis.kyoto-u.ac.jp	/ylab.html	×

表 5 「メンバー」を含むページのうちサイト内重要度が最も高かったページ

ブロックの間をウォークスルーして閲覧を行うものとする。格子状に配置された各ブロックをその座標を用いて、 $x_{i,j} \in X$ とする。また、ブロック間を東西・南北方向に走るユーザの可移動直線路を「通り」と呼ぶことにする..

5.2 配置に対する評価関数の設定

初期状態として、ブロックへのサイトの割り当てと、ブロックの各面へのページの割り当てをランダムに行う。そして、現在の配置に対する評価値

$$v = h(X)$$

を定める。そして第 6 章で述べる操作を繰り返し、 v が閾値以上になるか一定回数繰り返しを行うと、繰り返しを終了し配置を決定するものとする。配置指針により評価関数 $h(X)$ の設定が異なる。次に 2 つの配置指針とその場合の評価関数を述べる。

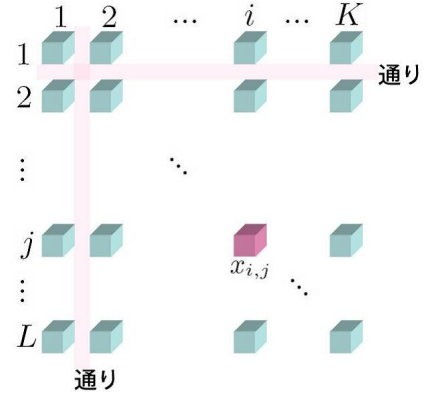


図 3 ブロックの格子状配置

5.3 配置指針

評価値を求める際に、各ページの特徴ベクトルを設定する。 $x_{i,j}^n, x_{i,j}^e, x_{i,j}^w, x_{i,j}^s$ に割り当てられているページの特徴ベクトルを、 $x_{i,j}^n, x_{i,j}^e, x_{i,j}^w, x_{i,j}^s$ で表すものとする。

特徴ベクトルは、TF/IDF 法など単語出現頻度に応じて生成するという方法が考えられる。

5.3.1 平行移動時の連続的比較閲覧を考慮した配置指針

この指針では、「同じ『通り』の同じ側には、複数ページの一定の側面が連続して並ぶ。」を目指している。

この指針においては、隣り合うブロックの同じ側の面の類似度の総和が評価関数として考えられる。評価値は次式で表される。

$$v = \sum_{j=1}^L \sum_{i=1}^{K-1} \{sim(x_{i,j}^n, x_{i+1,j}^n) + sim(x_{i,j}^s, x_{i+1,j}^s)\} + \sum_{i=1}^K \sum_{j=1}^{L-1} \{sim(x_{i,j}^w, x_{i,j+1}^w) + sim(x_{i,j}^e, x_{i,j+1}^e)\}$$

ただし $sim(x, y)$ は、Web ページの特徴ベクトル x と y の類似度である。

この指針においては、複数サイトの連続的比較閲覧はユーザの移動が前提となっている。

5.3.2 視界における連続的比較閲覧を考慮した配置指針

この指針では、「いま見えている範囲（視界）にある複数の面は共通の側面である（類似性がある）」を目指している。

ウォークスルーの最中の典型的な視界状況を 3 つ挙げ、それに基づき評価関数を設定する。

通りと平行に向いているとき (case1)

このとき共通の側面（類似性）を持つべき面は、見えている面積から優先順位をつけて列挙すると、

(1) 前後左右隣の建物どうしの同じ側の面

(2) ひとつの建物について対面以外の面

(3) 前後左右隣の建物どうしの向かいの面

ブロックに正対しているとき (case2)

このとき共通の側面（類似性）を持つべき面は、見えている面積から優先順位をつけて列挙すると、

(1) 斜め隣の建物どうしの同じ側の面
(2) ひとつの建物について対面以外の面
通りに対して 45°に向いているとき (case3)
このとき共通の側面 (類似性) を持つべき面は、見えている面積から優先順位をつけて列挙すると、

- (1) ひとつの建物について対面以外の面
- (2) 前後左右隣の建物どうしの向かいの面

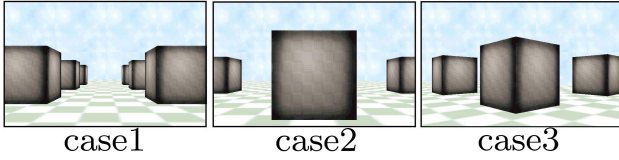


図 4 典型的な視界状況

評価関数

以上より評価関数として、

- 前後左右隣の建物どうしの同じ側の面
- ひとつの建物について対面以外の面
- 前後左右隣の建物どうしの向かいの面
- 斜め隣の建物どうしの同じ側の面

の類似度の総和が考えられる。

各々の評価値を定式化すると、

- 前後左右隣の建物どうしの同じ側の面

$$v = \sum_{j=1}^L \sum_{i=1}^{K-1} \{ \text{sim}(\mathbf{x}_{i,j}^n, \mathbf{x}_{i+1,j}^n) + \text{sim}(\mathbf{x}_{i,j}^s, \mathbf{x}_{i+1,j}^s) \} + \sum_{i=1}^K \sum_{j=1}^{L-1} \{ \text{sim}(\mathbf{x}_{i,j}^w, \mathbf{x}_{i,j+1}^w) + \text{sim}(\mathbf{x}_{i,j}^e, \mathbf{x}_{i,j+1}^e) \}$$

- ひとつの建物について対面以外の面

$$v = \sum_{j=1}^L \sum_{i=1}^K \{ \text{sim}(\mathbf{x}_{i,j}^n, \mathbf{x}_{i,j}^e) + \text{sim}(\mathbf{x}_{i,j}^n, \mathbf{x}_{i,j}^w) + \text{sim}(\mathbf{x}_{i,j}^n, \mathbf{x}_{i,j}^s) \} + \sum_{j=1}^L \sum_{i=1}^K \{ \text{sim}(\mathbf{x}_{i,j}^e, \mathbf{x}_{i,j}^w) + \text{sim}(\mathbf{x}_{i,j}^e, \mathbf{x}_{i,j}^s) + \text{sim}(\mathbf{x}_{i,j}^w, \mathbf{x}_{i,j}^n) \} + \sum_{j=1}^L \sum_{i=1}^K \{ \text{sim}(\mathbf{x}_{i,j}^w, \mathbf{x}_{i,j}^s) + \text{sim}(\mathbf{x}_{i,j}^w, \mathbf{x}_{i,j}^n) + \text{sim}(\mathbf{x}_{i,j}^w, \mathbf{x}_{i,j}^e) \} + \sum_{j=1}^L \sum_{i=1}^K \{ \text{sim}(\mathbf{x}_{i,j}^s, \mathbf{x}_{i,j}^n) + \text{sim}(\mathbf{x}_{i,j}^s, \mathbf{x}_{i,j}^e) + \text{sim}(\mathbf{x}_{i,j}^s, \mathbf{x}_{i,j}^w) \}$$

- 前後左右隣の建物どうしの向かいの面

$$v = \sum_{j=1}^L \sum_{i=1}^{K-1} \{ \text{sim}(\mathbf{x}_{i,j}^e, \mathbf{x}_{i+1,j}^w) \} + \sum_{j=1}^L \sum_{i=2}^K \{ \text{sim}(\mathbf{x}_{i,j}^w, \mathbf{x}_{i-1,j}^e) \} + \sum_{i=1}^K \sum_{j=1}^{L-1} \{ \text{sim}(\mathbf{x}_{i,j}^s, \mathbf{x}_{i,j+1}^n) \} + \sum_{i=1}^K \sum_{j=2}^L \{ \text{sim}(\mathbf{x}_{i,j}^n, \mathbf{x}_{i,j-1}^s) \}$$

- 斜め隣の建物どうしの同じ側の面

$$v = \sum_{j=2}^L \sum_{i=1}^{K-1} \{ \text{sim}(\mathbf{x}_{i,j}^n, \mathbf{x}_{i+1,j-1}^n) + \text{sim}(\mathbf{x}_{i,j}^e, \mathbf{x}_{i+1,j-1}^e) + \text{sim}(\mathbf{x}_{i,j}^w, \mathbf{x}_{i+1,j-1}^w) + \text{sim}(\mathbf{x}_{i,j}^s, \mathbf{x}_{i+1,j-1}^s) \} + \sum_{j=1}^{L-1} \sum_{i=1}^{K-1} \{ \text{sim}(\mathbf{x}_{i,j}^n, \mathbf{x}_{i+1,j+1}^n) + \text{sim}(\mathbf{x}_{i,j}^e, \mathbf{x}_{i+1,j+1}^e) + \text{sim}(\mathbf{x}_{i,j}^w, \mathbf{x}_{i+1,j+1}^w) + \text{sim}(\mathbf{x}_{i,j}^s, \mathbf{x}_{i+1,j+1}^s) \} + \sum_{j=2}^L \sum_{i=2}^K \{ \text{sim}(\mathbf{x}_{i,j}^n, \mathbf{x}_{i-1,j-1}^n) + \text{sim}(\mathbf{x}_{i,j}^e, \mathbf{x}_{i-1,j-1}^e) + \text{sim}(\mathbf{x}_{i,j}^w, \mathbf{x}_{i-1,j-1}^w) + \text{sim}(\mathbf{x}_{i,j}^s, \mathbf{x}_{i-1,j-1}^s) \} + \sum_{j=1}^{L-1} \sum_{i=2}^K \{ \text{sim}(\mathbf{x}_{i,j}^n, \mathbf{x}_{i-1,j+1}^n) + \text{sim}(\mathbf{x}_{i,j}^e, \mathbf{x}_{i-1,j+1}^e) + \text{sim}(\mathbf{x}_{i,j}^w, \mathbf{x}_{i-1,j+1}^w) + \text{sim}(\mathbf{x}_{i,j}^s, \mathbf{x}_{i-1,j+1}^s) \}$$

この指針においては、複数サイトの連続的比較閲覧は視野の移動が前提となっている。インタフェースにおいて視界は実装手法により大きく変わり得ると考えられるので、より一般的に妥当な評価関数の設定が必要である。

6. 自動配置の手法

6.1 操作

与えられた初期配置から配置を変更して現在の配置に対する評価値を高める。配置を変更するための3つの操作を述べる。

6.1.1 ページの置換

ページの置換は、ブロック $x_{i,j}$ の $x_{i,j}^n, x_{i,j}^e, x_{i,j}^w, x_{i,j}^s$ に割り当てられているページの組み合わせを変更するものである。

$$\{x_{i,j}^n, x_{i,j}^e, x_{i,j}^w, x_{i,j}^s\} \Leftrightarrow \{x_{i,j}'^n, x_{i,j}'^e, x_{i,j}'^w, x_{i,j}'^s\}$$

ブロック $x_{i,j}$ にサイト s_k が割り当てられているとき、そのページ群は P_k であり、 $x_{i,j}^n, x_{i,j}^e, x_{i,j}^w, x_{i,j}^s$ へのページの割り当て方は、重複を許すのであれば $|P_k|^4$ 通り、重複を許さないのであれば $|P_k|P_4$ 通りとなる。

ページの置換操作は遺伝的アルゴリズムの選択に相当する。

6.1.2 サイトの置換

サイトの置換は、2つのブロックの位置を交換するものである。

$$x_{i,j} \Leftrightarrow x_{i',j'}$$

ブロック集合 X の要素数を $|X|$ で表すと、ある時点におけるサイトの置換を行う2つのブロックの選び方は、 $|X|C_2$ 通りである。いずれの操作も、前述の評価関数では変更のあったブロックに関する部分の再計算だけをすればよく、全体を再計算する必要はない。

サイトの置換操作は遺伝的アルゴリズムの交叉に相当する。

6.1.3 サイトの突然変異

局所解に陥るのを防ぐために、遺伝的アルゴリズムでは突然変異が行われる。この自動配置手法においては、ブロックに対してページ置換を試み、最も $v_{current}$ が小さくなるページの割り当て方を選択する。

6.2 遺伝的アルゴリズム

遺伝的アルゴリズムによる自動配置の流れは以下のとおりになる。

- (1) 初期配置をランダムで設定。
- (2) 現在の配置に対して、評価値 $v_{current}$ を算出。
- (3) $v_{current}$ が閾値 V を超えると、終了配置を決定する。
- (4) 選択

すべてのブロックに対し、ページの置換操作を行い、最も $v_{current}$ が大きくなるページの割り当て方を選択する。ただし、ひとつのブロックに対して独立で $v_{current}$ が最大となるページの割り当て方は決定できないので、現在のブロックの配置に対して、すべてのブロックに対し、ページの置換操作を行い、最も $v_{current}$ が大きくなるページの割り当て方を選択という操作を、収束するまで繰り返す。

- (5) 交叉

すべてのブロックをペアにして、それぞれのペアに対して交叉確率 P_c に基づいて置換操作を試みるかを決定する。置換操作を試みる前の配置の評価値を v 、実行した後の配置の評価値を v' とするとき、 $v < v'$ ならば操作を実際に行う。

(6) 突然変異

すべてのブロックに対して突然変異確率 P_m に基づいて突然変異操作を実行する。

(7) 3 に戻る。

6.3 準備実験・評価

自動配置手法による、配置に対する評価値の推移を実験した。対象とする Web サイトは、任意に選んだ京都の寺社の 16 サイトとし、その側面となるページ群はトップページに存在するリンク先のページとした。また配置指針は、5.3.1 の「平行移動時の連続的比較閲覧を考慮した配置指針」とした。ページ間の類似度は単語の出現頻度を TF/IDF 法で重み付けてページ毎に特徴ベクトルを生成し、ページ間の類似度をコサイン類似度とした。日本語ページのテキストから単語を抽出するために、奈良先端科学技術大学院大学自然言語処理学講座にて開発された日本語形態素解析器である茶筌 [7] を用いた。交叉確率 P_c および突然変異確率 P_m はそれぞれ、 $P_c = 0.7, P_m = 0.1$ に設定した。

6.3.1 実験結果

3つの初期配置に対し、選択 交叉 突然変異のループを 250 回繰り返しその評価値の推移を 50 回ずつ調べた。図 7 が初期配置 A、図 8 が初期配置 B に対して行ったものである。

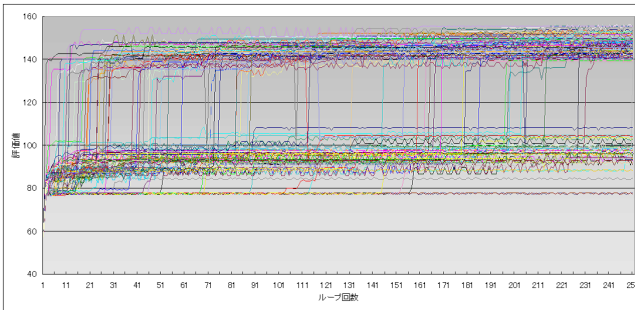


図 5 ループ回数と評価値の推移 (初期配置 A)

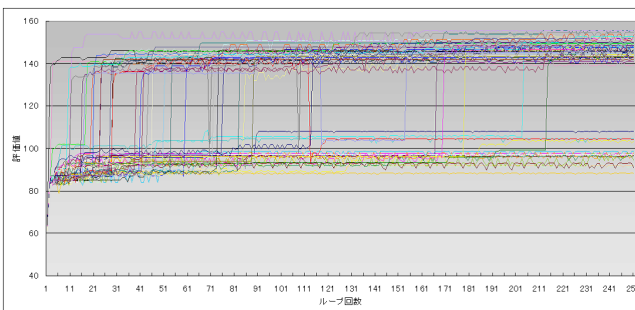


図 6 ループ回数と評価値の推移 (初期配置 B)

6.3.2 考察

まず、異なる初期配置に対しても、おおむね同じような値に収束することが分かった。また、初期配置 A,B とともに、評価値

が 90 前後と 150 前後の 2 通りに収束している。90 前後の方は局所解に陥っていると考えられる。突然変異確率 P_m の調整や、突然変異操作の見直しが必要と考えられる。

そこで、突然変異確率 P_m を $P_m = 0.2$ あるいは $P_m = 0.5$ などに設定して同様の実験を行ったが、やはり同様に 2 つの解に収束した。突然変異操作を、すべてのブロックをペアにして、それぞれのペアに対して突然変異確率 P_m に基づいて置換操作を実行する、という操作に変更して同様の実験を行っても、やはり同様に 2 つの解に収束した。局所解に陥る可能性をより低くするために、突然変異操作の検討が必要である。

図 9,10 は初期配置 A,B に対する 250 回ループ後の評価値の分布である。比較検討のために、突然変異操作を行わない場合 ($P_m = 0$) で同様の実験を行った場合の 250 回ループ後の評価値の分布が図 11,12 である。突然変異操作を行わない場合は、行った場合に比べて局所解に陥る可能性が高くなっている。これにより、突然変異操作により局所解に陥る可能性が低くなっていると言える。

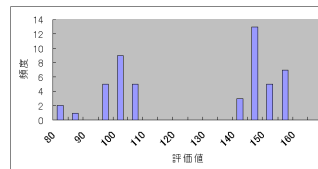


図 7 A:突然変異あり

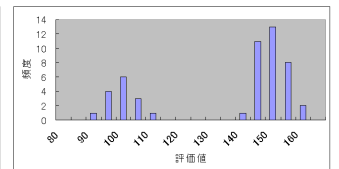


図 8 B:突然変異あり

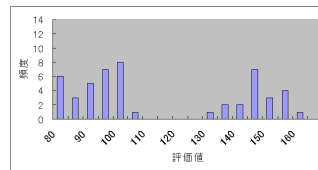


図 9 A:突然変異なし

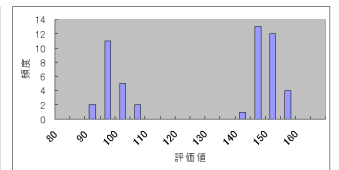


図 10 B:突然変異なし

7. プロトタイプ実装および考察

以上を踏まえて空間インタフェースを実際に構築し、この空間インタフェース内をウォークスルーしてブラウジングを行った結果を考察する。

7.1 実装



図 11 ウォークスルー中の様子

プロトタイプシステムの実装に当たっては 3DML [3] を使用した。3DML はアメリカ Flatland 社が開発した三次元空間構築言語であり、ユーザはプラグインを用いて、Web から利用することができる。

立方体を 2 つ積み上げてひとつのブロック (Web サイト) を構成し、上の立方体には全面にトップページのスクリーンショットが貼られ、下の立方体の 4 つの側面には Web サイトの各々異なるページが割り当てられ、スクリーンショットが貼られている。

7.2 考 察

空間的移動による Web 閲覧

既存のブラウザでは、リンクアンカーをクリックして別ページへ移動して Web 閲覧を行うが、このインタフェースでは空間的に移動するのみで別ページの内容を参照することができる。ある一定の複数 Web サイトの集まりを「街」としてひとまとまりにして、ひとつの「街」内での移動に限らず、複数の「街」の間での移動による Web 閲覧スタイルも考えられる。

視覚的な印象

閲覧者が通りに立って Web 情報を見たときに、「類似したものが並んでいる」と感じにくい。内容的には類似したものを選んでいますが、それを一瞥しただけでは分からない。一瞥しただけで類似したものが分かるというような見せ方が必要である。例えば、ある通りの主題を抽出して主題を記した看板メタファを置いたりすることが考えられる。

閲覧スタイル

「同じ『通り』の同じ側には、複数ページの一定の側面が連続して並ぶ」に基づいた空間配置では、閲覧者は直進方向と直交する向きを見ながら移動しないと連続的閲覧ができず、これは不自然である。

ブロック

現在でブロックの高さはすべて均等で意味付けがなされていない。例えば、サイトに対して重要度を設定し、その重要度にもとにブロックの高さを変化させることが考えられる。また、ブロックを直方体に限定せずさらには格子状配置も含めて、より適切な空間配置の検討が必要である。

街メタファ

現在では規則的にブロックが林立しているだけである。K. Lynch が The Image of the City [5] で提唱した認知地図にある、Landmark や District などの概念をこのインタフェースに導入して、Web 空間の街メタファによる視覚化が行えないか検討したい。

受動的閲覧支援

このインタフェースでは能動的な比較閲覧のみならず、明確な目的を持たずとも流れるように関連ページの閲覧を行う受動的閲覧インタフェースにもなりうる。

可視性

通常のブラウザに比べ、表示面積が狭いためページによっては内容が読み取れない場合がある。ユーザの現在位置からの距離によって表示情報の詳細度制御を行う方法などが考えられる。

8. 結 論

本研究ではひとつの Web サイトをブロックと呼ばれる直方体で表し、ブロックの各々の面にそのサイトの各ページを貼り付けることにより、ひとつの Web サイトを空間的なアングルによってさまざまな側面からの閲覧を可能にするためのインタフェースを提案した。また、それらを複数並べて複数 Web サイトの多面的な閲覧・連続的比較閲覧を支援する空間配置の指針と、自動配置を行うための遺伝的アルゴリズムを用いた手法についても述べた。そして、システム全体のプロトタイプを作成しインタフェースとしての問題点を提起し考察した。

今後は空間配置のための基本的技術として、Web サイトにおける側面の抽出方法、空間配置の指針についてさらに検討し、より妥当な手法・指針が必要であると考えている。そのために、実装された空間インタフェース全体についての評価方法についても検討も必要であると考えている。さらには、インタフェース全体としての改善が必要であると考えている。今回は空間配置の基本的方針として格子状配置のみを検討したが、必ずしもこれが最適であるとは言えないので、ユーザの閲覧スタイルを研究してどのような空間配置が閲覧スタイルにフィットするのかが検討・考察したい。

また、ユーザのアクションやプリファレンスに基づいて配置構成を動的に変化させることも考えられる。さらには複数ユーザによるインタフェースの同時使用を想定した web 散策支援機構なども検討したい。

謝 辞

本研究の一部は《知的資産》文部科学省科学技術振興費プロジェクト「異メディア・アーカイブの横断的検索・統合ソフトウェア開発」(代表: 田中克己) および、平成 16 年度科研費特定領域研究 (2) 「Web の意味構造発見に基づく新しい Web 検索サービス方式に関する研究」(課題番号: 16016247, 代表: 田中克己) および、21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」および、平成 16 年度科研費若手研究 (B) 「参照の同一性判定に基づく複数 Web ページの検索閲覧方式の研究」(課題番号: 16700097, 代表: 小山 聡) によるものです。ここに記して謝意を表すものとします。

文 献

- [1] INFOTUBE
<http://www.plannet-arch.com/information/tube-jp.htm>.
- [2] T.Ishida: "Digital city kyoto: Social information infrastructure for everyday life", Communications of the ACM (CACM), Vol. 45 (2002).
- [3] Flatland Online
<http://www.flatland.com/>.
- [4] F. Sparacino, G. Davenport and A. Pentland: "City of news", KOS, No. 179-180 (2000).
- [5] K.Lynch: "The Image of the City", The MIT Press (1960).
- [6] S.Mukherjea and Y.Hara: "Focus + context views of worldwide web nodes", UK Conference on Hypertext, pp. 187-196 (1997).
- [7] 形態素解析システム茶釜
<http://chasen.naist.jp/hiki/ChaSen/>.