

複数評価値を用いた検索のための Web リンク空間の可視化

宮原 秀友[†] 小林 亜樹^{††} 山岡 克式^{††} 酒井 善則^{††}

^{†,††} 東京工業大学 〒 152-8552 東京都目黒区大岡山 2-12-1

E-mail: [†]hide@net.ss.titech.ac.jp, ^{††}{koba,yamaoka,ys}@ss.titech.ac.jp

あらまし 筆者らは、Web リンク空間における近傍情報探索を行う PIRCS(パークス)を提案している。PIRCS では、Web サーバ側で検索の一部を行う構成とすることで検索処理の効率化を図る。本稿では、検索キーの入力や結果表示を含むインターフェースを担うクライアント部分について、直観的な検索結果の提示を実現する手法を提案し、また、試作システムを作成し、その効果について検証する。本提案手法は、システムに用意された複数の検索アルゴリズムによる各評価結果を多次元の表現軸へ投影する手法で、表現軸に、表示空間を作り出す x 軸、y 軸、z 軸、ノードの色軸、形状軸を実装し、評価結果に従ってコンテンツを表すノードを配置し、リンクを線で結び表示する。ノード位置とリンク表示により単独の検索アルゴリズムでは表現できないコンテンツ間の関係をも可視化し、目的コンテンツ発見を支援する。また、ユーザが検索アルゴリズムと表現軸を自由に選択できるため、目的に応じた直感的な理解を促す表示が出来る。

キーワード 情報検索、データの可視化、Web 利用技術

Visualization of the Web link space for the retrieval result using multiple scores

Hidetomo MIYAHARA[†], Aki KOBAYASHI^{††}, Katsunori YAMAOKA^{††}, and Yoshinori SAKAI^{††}

^{†,††}Faculty of Engineering, Tokyo Institute of Technology 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan

E-mail: [†]hide@net.ss.titech.ac.jp, ^{††}{koba,yamaoka,ys}@ss.titech.ac.jp

Abstract We proposed PIRCS for neighboring search in Web link space. In PIRCS, it intends to increase the efficiency of the search process by constructing web servers to take on a portion of the search. In this paper, we propose how to show retrieval results viscerally regarding the client side that provides the interfaces, including input of the retrieval keys and display of the retrieval results, as well as create experimental system and validate its effects. It is the technique of projecting each evaluation result by two or more reference algorithm prepared for the system to the expression axes of many dimensions. We implemented x-axis, y-axis and z-axis which make display space and color-axis and form-axis of a node as the expression axes. According to each evaluation result, the nodes which express contents are arranged, and the lines which express links are connected. Usually, independent retrieval algorithm cannot visualize the relation between contents. However the proposed method can visualize the relation between contents and supports the purpose contents discovery because of nodes position and links display. Moreover, since a user can choose the combination of retrieval algorithm and an expression axis freely, a user can get the display which can obtain an intuitive understanding according to the purpose.

Key words Information retrieval, Visualization of data, Web technology

1. はじめに

最も一般的な情報ツールとなった Web 上には、大量の情報が、リンクで結ばれたコンテンツの集合体として存在している。Web のリンクは、一般に人手によって関連するコンテンツを結びつけたものであり、関連情報がまとまっていることが多い。このようにまとまった一連のコンテンツ群のことは Web コミュ

ニティなどと呼ばれ、Web リンク空間の大域構造を示すものとして研究が進められている [1]–[5]。また、一般的な Yahoo! [6] や Google [7] といった検索エンジンによっても、求める情報が属するコミュニティの 1 ページを発見することは比較的容易である。

しかし人が情報を欲するとき、情報の大まかな分類を示すコミュニティに関する知識だけを得ても不十分である。たいいてい

は詳細な検索を行う必要があり、検索エンジンを用いる場合には、検索キーワードを追加していく行為などがこれに該当する。このことを Web リンク空間で考えると、関連情報の存在するコミュニティ内での絞り込み検索を行えば目的に近づけると言える。

しかし、現状のサーバ系の検索サービスすべてが、事前に収集した情報に関してサーバ上で検索処理を行い、結果を返すという処理系となっており、コミュニティ内のようなリンク近傍空間での詳細な検索には向いていない。具体的には、Web の超分散的性格から、頻繁に更新される Web コンテンツに対する検索のためには、事前に情報収集を行うアーキテクチャでは対応しきれない問題が指摘される。また、キーワードの追加による絞り込み検索であっても、Web リンク空間としては毎回ゼロからの検索であり、検索結果はあくまでグローバルな情報源からのフィルタ結果でしかないため、大量の情報が返されたりする点も問題である。

そこで、このようなリンク近傍空間での検索を、近傍情報探索として、Web 空間からの情報検索をより効率的に行うことを提唱している [8]。また、Fetuccino [9] と呼ばれるサービスも提案されており、Web 部分空間の把握と検索結果の提示を両立している。Fetuccino のように、近傍情報探索は、すべての処理をユーザの手元のコンピュータで行うことも可能である。しかし、探索範囲内に存在するすべてのコンテンツを手元に引き出す必要があるため、時間がかかるばかりか、通信という観点から考えると、サーバからすべての情報がクライアントに転送されており、検索によって効率化が図られたとは言えない。

筆者らは、近傍情報探索を効率よく行うプラットフォームとして、現在の Web アーキテクチャに拡張を加えた「協調型 Web アーキテクチャ」[10]–[12] を提案し、さらにこれを用いることで効率的なリンク空間での検索を目指す近傍情報探索システム PIRCS (パークス) を提案している。協調型 Web アーキテクチャでは、Web サーバを従来の単なるコンテンツ提供システムから自律分散型データベースシステムとしても利用できるよう、サーバ間でのコンテンツに関するメタ情報の交換機能を追加し、さらにそのメタ情報を利用し、論理近傍における検索キーとメタ情報との距離計算の機能を追加することで Web サーバにおける検索を可能とする。

ここで、Web 検索システムにおけるインターフェースには、まず、データ集合としてのコンテンツの検索結果に基づく可視化と、実際の Web コンテンツ取得のためのナビゲーション機能が必須である。一般的な検索エンジンでは、それぞれ、固定的な検索評価に基づくリスト表示と、Web コンテンツへのリンクによって、最小限の機能を提供していると言える。しかし、Web コンテンツは単体で完結するのではなく、リンクによって前後の文脈を構成している例も多いため、インターフェースとして貧弱なものであると言わざるを得ない。また、Web リンク空間の可視化自体は、これまでも、Hyperbolic Tree [13]、H3 [14]、Cone Tree [15]、納豆ビュー [16] などに見られるが、これらは固定された配置方法で、検索結果としての表示は難しい。

PageRank [17] は、検索評価値に反映しようとした試みとし

て評価されるが、修辭的には同等のハイパーリンクであっても、その意味するところには大きな隔たりが見られるようになった現状では、有効に機能するとは言い難い。なぜなら、ブログ (Weblog) の普及によってさらに飛躍的にコンテンツ作成に関わる人口が増加した結果、Web コンテンツの一部であるリンクもまた急激に増加し、また、更新の頻度も飛躍的に高まったためである。残念ながら、ハイパーリンクの意味するところを正確に分類することは困難であるため、本研究では、ユーザがリンクの前後のコンテンツなどを直接的に見ることで、その判断をユーザ自身に委ねるアプローチを採る。

INFOVISER [18] のようなデータ可視化手法によって、ユーザは原始的な検索アルゴリズムを多数組み合わせることで、自己の情報要求に合った情報を引き出すことが容易になると考えられる。単純なキーワードへの適合度を複数キーワード毎に独立に得て、その結果を多次元表示することで、AND、OR、NOT などの論理演算と同等の集合を直観的に導き出すことが出来るばかりか、微妙な重み付けなどにもその画面のままで読み取ることが可能である。本稿では、Web コンテンツの検索に特化した表現軸について提案する。

さらに、雑多な Web コンテンツからの検索では、検索クエリや評価軸を適宜変化させながら、試行錯誤することで情報要求に見合ったコンテンツの取得へ向かえると考えられる。このような行為を支援する機能を、ナビゲーション支援機能と呼ぶ。このとき、検索エンジンにはせいぜい 2、3 個のキーワードしか入力されない例がほとんどであると言われていることから、多くの操作がマウスによって行えることが重要である。マウス動作によって、局所的にリンク構造を詳細に見たり、より多くの情報を表示したり、検索空間を移動できたりなど、Web 検索特有のインタラクティブな操作を導入することでナビゲーション支援を行う。

また、データ集合の検索には、データそのものの情報が得られなければならない。しかし、超分散型のデータソースである Web システムにおいて、処理のためにデータを取得する行為は、通信という観点から見た場合、検索による効率化が図られたとは言えない。そこで、本稿では、リンク空間の整合性をサーバレベルで維持することを提案している協調型 Web アーキテクチャに基づくサーバに、検索のための機能を追加し、サーバ側、クライアント側双方で、検索処理を分担することで、通信量も低減でき、ユーザへのレスポンスも改善するシステムについて提案し、試作する。

以降、2 章で筆者らが提案している拡張型 Web アーキテクチャと PIRCS について説明し、3 章で提案手法を述べる。その後、4 章で PIRCS のクライアントに提案手法を用いた実装を説明し、評価実験とその結果について述べる。

2. 協調型 Web アーキテクチャと PIRCS

2.1 従来の Web アーキテクチャ

1 章で挙げた従来の検索システムの欠点は、従来の Web アーキテクチャにも原因がある。

従来の Web アーキテクチャは図 1 の様に、サーバ-クライア

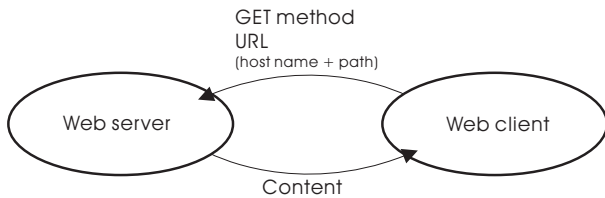


図1 Web アーキテクチャ

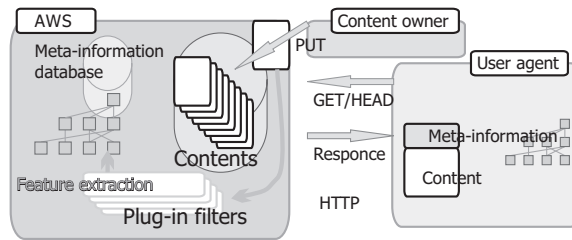


図3 検索処理

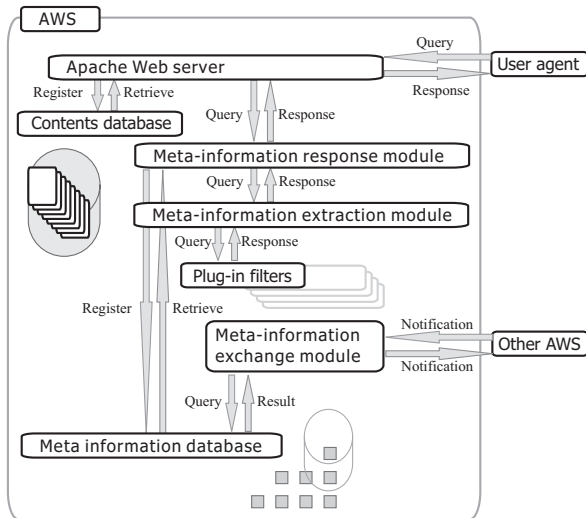


図2 Advanced Web Server

ント方式を採っており、クライアント(ブラウザ)からの要求に対し、Webサーバは自サーバ内に要求されたコンテンツがある時のみコンテンツを配信する。また、Webサーバは各々独立して存在しており、各々が独自に情報を配信している。コンテンツ同士はリンクで繋がれているが、Webサーバ同士が繋がっているわけではない。つまり、Webは本質的に分散型の情報配信アーキテクチャを持っており、コンテンツの位置や内容、相互の関係などを記述、制御する仕組みを持っておらず、検索を考慮した構造にはなっていない。

2.2 協調型 Web アーキテクチャ

従来の問題点を解決するため、筆者らは協調型 Web アーキテクチャ(CWA)を提案している。CWAでは、従来の Web アーキテクチャに加えて、リンクコンテンツのメタ情報収集、維持、提供、論理近傍における検索キーとメタ情報との距離計算といった機能が Webサーバに追加される。

このような機能を持たせた Webサーバを Advanced Web Server(AWS)と呼ぶ(図2)。

リンクコンテンツとは、リンクで結ばれたコンテンツを指し、そのメタ情報とは、そのリンクコンテンツの特徴を表す情報の総称である。そのうち、リンク構造を決定するリンクするコンテンツの URL などの情報を特にリンク情報と呼び、それ以外のリンクには関係しないコンテンツの要約などを検索情報と呼ぶ。

Webサーバがリンクコンテンツのメタ情報収集、維持、提供を行うことで、コンテンツを Webサーバが管理し、Webサーバ間でコンテンツ情報のやり取りが可能になる。つまり、Webサーバが協調しあい、コンテンツやその繋がりを管理すること

ができる。また、論理近傍における検索キーとメタ情報との距離計算を行えるようにすることで、Webサーバ自体が検索を行えるようになる。検索の範囲は、いち Webサーバが管理しているコンテンツが対象となり、一種のコミュニティ内検索となる。コンテンツやその繋がりを Webサーバが管理しているので、今現在の情報との整合性は常に取られることになり、また、リンク情報を使った検索が可能となる。

Webサーバ側検索における具体的な検索アルゴリズムそのものは CWA 本体では扱わず、様々な種類のコンテンツに適用できるように、プラグイン形式の検索フィルタを使用する。

2.3 PIRCS 上での近傍検索

PIRCS は AWS と対話するユーザクライアントからなるサーバクライアント方式の検索システムである。近傍検索を PIRCSで行うときは、図3のようにクライアントからの検索リクエストに対して、サーバ側が近傍コンテンツに対する検索を行った結果を応答する形式で実行される。サーバ側、クライアント側双方で、検索処理を分担することで、通信量も低減でき、更に、事前にサーバ側で検索処理が行われているので、クライアント側のみでコンテンツを集め処理を始めるより早く検索が行える。検索リクエストは、検索の起点となるコンテンツの URL(StartURL と呼ぶ)、検索フィルタ名、検索キー、および検索範囲の指定によって構成される。サーバである AWS では受信した検索リクエストに従って、検索フィルタに内蔵された検索アルゴリズムによって StartURL 近傍のコンテンツにおける検索処理を実行する。

このとき、近傍コンテンツの検索に必要なメタ情報は、協調型 Web アーキテクチャの仕組みによって、常に最新の情報にメンテナンスされているため、応答する検索結果においてリンクおよび内容が実際のコンテンツと整合しない状況は発生しない。また、startURL にユーザの Web コミュニティに関する知識を活用できる。検索範囲はこの startURL からの近傍コンテンツに限定されるため、適当な Web コミュニティにおける検索結果に近い、扱いやすい数の検索結果となる。なお、クライアント側で意図する検索範囲が、AWS が保持する範囲より広い場合には、クライアントは範囲を広げるよう、複数回、異なる startURL を与えて同様の検索をリクエストする。これらの複数回の検索結果を統合(合成)して最終的な検索結果とし、ユーザに提示する。

AWS における検索処理は、実際には各サーバにプラグインされた検索フィルタに委譲され、フィルタに内蔵されたアルゴリズムが実行し、結果を返す。フィルタはテキストだけではな

く、静止画や動画像、音楽ファイルなどそれぞれの形式、または複数の形式を処理することが可能なように設計される。また、検索アルゴリズムもそれぞれのフィルタによって異なるため、多数の評価軸を自由に設定できる。

3. 検索結果の可視化

3.1 概要

既存の検索結果の提示には、検索結果にパイパーリンクを十分に使っていない点と、複数の検索キーを用いた場合、検索結果に各検索キーがどのように反映されているかをユーザが理解できない点が問題点としてあげられる。Web コンテンツの評価点を算出するためにリンク情報を一部用いている例 (PageRank) があるが、ユーザにそのリンク関係を直観的に提示するインターフェースとして用いられているわけではない。リンクによってつながれたコンテンツ間には意味的なつながりが存在するため、ユーザにリンク関係を見せることで、ユーザが Web コンテンツ空間においてそのコンテンツが存在する位置を直観的に理解することを助けることが可能となる。また、既存の検索エンジンでは、複数の検索キーを用い and, or 検索を行った場合でも、独自の論理式、重み付けで検索結果を導き出し、提示する。しかし、その論理式、重み付けにユーザの嗜好を反映することは難しい。論理式や重み付けを使用して一つに検索結果をまとめるのではなく、すべての検索結果を同時に表示し、ユーザが判断し選択出来る結果表示が望ましい。関連研究で挙げた Web リンク空間の可視化や検索結果の可視化の研究では、第一の点は多少解決されているが、第二の点において適用することは不可能である。そこで、検索結果に付与される特徴的な情報を使って、検索結果に含まれる Web コンテンツを視覚的に配置し、リンク関係を表示する可視化手法を提案する。

3.2 提案手法

本手法は、概要で述べた二つの問題点を解決し、且つ、ユーザが直観的に理解できるよう、自由度を持たせ、以下のように提案する。

- コンテンツをノードとして表現する。
- 複数の評価結果が与えられたとき、それぞれの検索意図を適切に表現する表現軸へと投影する。
- 表現軸には、表示省略機能として閾値による表示/非表示機能を持たせる。
- ノード間をリンクで接続する。
- コンテンツの概要をノード毎に表示する。
- あるノードから更に検索空間の移動、拡張が行える。

つまり、各検索アルゴリズム、検索キーで得られた評価値を、各々を表現する表現軸に投影し、表示する手法である。

表現軸には、まず、表示空間を作り出す、x 軸、y 軸、z 軸を用意し、得られた評価値によって、表示空間に各コンテンツを表すノードを配置する。既存の検索エンジンが検索結果を一つにまとめなければならない理由の一つとして、表示空間が 1 次元で、複数の結果を同時に表示することが不可能だということが挙げられるが、表示空間として 3 次元空間を用意すれば、最低でも 3 種類の評価値を同時に表示することが可能となる。ま

表 1 コンテンツと評価値

コンテンツ	評価値 1	評価値 2
A	10	20
B	20	50
C	30	40
D	40	10
E	50	30

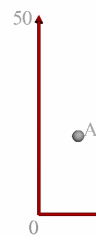


図 4 表示例 1

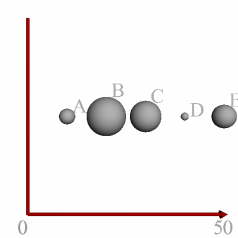


図 5 表示例 2

た、ノードの色や形状と言った表現軸も用意することにより、さらに次元数を上げて検索し、各検索結果を同時に表示することも可能となる。

例えば、表 1 のようにある二つの検索アルゴリズム、検索キーによって得られた二つの評価値 (0 から 50) を持つコンテンツ群があるとする。評価値 1 を x 軸に、評価値 2 を y 軸に投影し、コンテンツをノードとし各評価値ごとにプロットすると図 4 のように表せる。一方、評価値 1 を x 軸に、評価値 2 をコンテンツを表すノードの半径に投影すると図 5 のように表せる。これは簡単な例だが、このように同じ評価値を持つコンテンツ群でも投影する表現軸により結果表示が違ったものとなる。

ユーザは表現軸と検索アルゴリズム、検索キーの組み合わせを自由に決めることができるため、目的に応じた直観的な理解を促す表示が得られる。

また、結果と共に同時にリンクも表示させる。リンク構造を可視化することで、従来の検索エンジンの検索結果表示では表示しきれなかったコンテンツ間関係をも可視化し、更にナビゲーション支援としての機能を持たせることができる。

ユーザの求める複数の検索キーにおける評価値の相互関係や、それがどのようにリンク構造と関連しているかが一目でわかるという利点があり、かつ、この方式では、プロット位置は評価値から一意に定まるため、複雑な計算を要することがない。

3.3 検索シナリオ

検索結果として、コンテンツを複数の評価値を用い多次元にノードとして配置しリンクの繋がりを見せる方法が、Web リンク空間における近傍検索においてどのような表示を行い、検索に際しどのような効果を持つのかについて、いくつかの検索シナリオを用いて説明する。

まず、ユーザが 3 つの Keyword A, B, C に関連する情報を検索する場合について説明する。ユーザは Keyword A, B, C それぞれに対する評価値を使って多次元表示することで目的を達成しようとする。図 6 は、Keyword A, B, C にそれぞれに対する評価値 (0 から 100) を x, y, z の各軸に投影した空間である。ここに、各ノード及びその繋がりであるリンクを表示する。この場

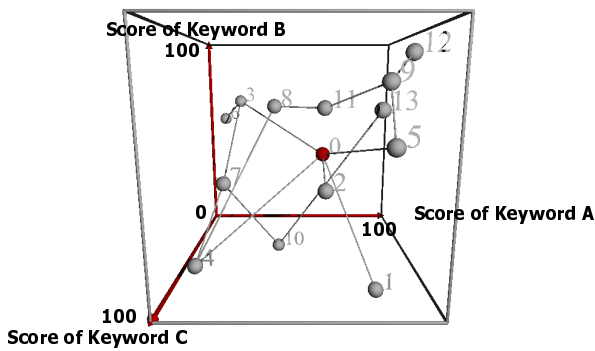


図6 直観的理解を促す論理演算検索結果の表示

合、Keyword A に対する評価値が高いノードほど右に配置されている。

コンテンツ 12 が、すべてのキーワードにもっとも関連性の高いコンテンツであることは、リスト表示型インターフェースと同様にわかる。加えて、リンクで繋がれているコンテンツは関連性が高いと考えられるため、リンクで繋がっている 5,9,11 のコンテンツに一連の情報があると読み取れる。一方、コンテンツ 13 は各評価値は高いが、周囲のコンテンツの評価値が低いため、関連性が低い可能性が高い。

次に、ユーザがある3つのキーワードを AND 検索や OR 検索などの論理式を用いて検索するシナリオを説明する。Keyword A、Keyword B、Keyword C のどれか1つを含むコンテンツを検索したい場合や、Keyword A と Keyword C を含んでいて、Keyword B を含んでいないコンテンツを検索したい場合など3つのキーワードをいろいろな組み合わせで検索したい場合、従来の検索エンジンでは、「Keyword A OR Keyword B OR Keyword C」、「Keyword A AND Keyword C AND NOT Keyword B」などと論理式を変えて検索を別々に行わなければならなかった。提案手法では3つのキーワードに対する検索結果を図6のように可視化できるため、Keyword A と Keyword C を含んでいて、Keyword B を含んでいないコンテンツはコンテンツ 1 であり、Keyword C のみを含んでいるコンテンツはコンテンツ 4 だということが同じ検索結果から読み取ることができ、AND 検索や OR 検索が同時に行えるという効果が得られる。

次に、近傍検索がユーザの事前知識を活かした検索を行える点と検索評価値による配置も行うことによる効果について述べる。

各コンテンツの Keyword に対する距離と StartURL コンテンツとの類似度を 0 から 100 までの評価値として得られるとする。Keyword に対する距離としての評価値を x 軸に、StartURL コンテンツとの類似度としての評価値を y 軸に投影し、コンテンツをノードとして 2 次元空間にプロットし、リンクによる結びつきを線として表現したのが図7の可視化である。ノードが上にあればあるほど startURL コンテンツとの類似度が高く、右にあればあるほど Keyword を多く含んでいることとなる。

近傍検索では、StartURL コンテンツが探したい情報の分野である場合に効率が高い。ここで、コンテンツ 5 が探したい情報分野にあり、その分野でキーワードを含むコンテンツを探そう

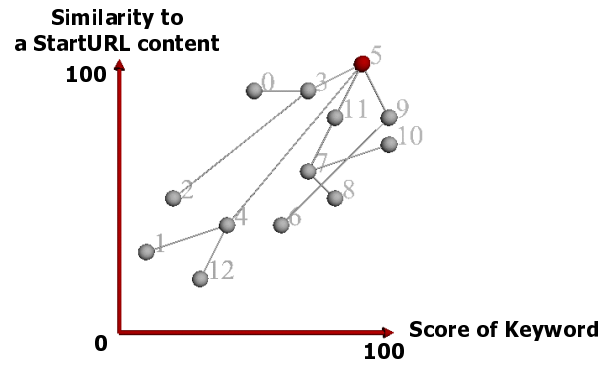


図7 近傍検索と検索評価値による配置による効果

としているとすると、ユーザはコンテンツ 9,10,11 が目的のコンテンツであると予測できる。

一方、コンテンツ 4 は StartURL コンテンツから 1 リンクで繋がっているが、StartURL コンテンツとの類似度が低い一連のコンテンツ (1,12) と繋がっており、目的の情報とは異なると判断できる。このように、近傍検索の結果を多次元空間にプロットし、更にリンク関係をも図示することで、近傍検索の結果を明確かつ効果的にユーザに提示することが出来る。

4. 試作システムと評価実験

4.1 概要

Web リンク空間における効率的な近傍情報探索システム PIRCS の一環として、クライアント側のユーザインターフェースに本提案手法を用いた試作システム作成した。本システムは、サーバ側に別途試作した協調型 Web アーキテクチャ (CWA) の Web サーバ (AWS) を配し、本提案手法によるクライアント側はこの AWS と通信を行い、検索結果の評価値やリンク情報を素早く収集する。

図8にクライアント側、およびサーバ側の構成を示す。サーバである AWS は apacheWeb サーバを核として、CWA により拡張された機能の実行は Java で記述されたシステムを呼び出す。リンク情報など多数のコンテンツに関する情報の格納には PostgreSQL を用いている。

クライアント側である本ユーザインターフェースを実装したブラウザは、左側の網掛け内のような構成を採る。ユーザからのリクエストを解釈し、サーバとは HTTP で通信を行い、検索評価値などのセットを受け取った後、多次元空間への配置処理等を行い、表示することでユーザに結果を示す。なお、以下単に本システムと呼ぶ場合、このクライアント側のみを指す。本システムは全体を Java でコーディングし、表示エンジンには Java3D を用いた。したがって、Java の実行環境上で動作する。

4.2 インターフェース

図9が試作した本システムのユーザインターフェース例である。画面左側の検索情報指定領域と右側の結果表示領域から構成される。検索情報指定は、左側の一番上に startURL を入力し、リンク空間の検索方向 (順方向、逆方向、双方向など) や深さ (何リンク先までを探索対象とするか) を入力する。その下に、表示空間軸と検索評価軸とを結びつけるための入力欄、

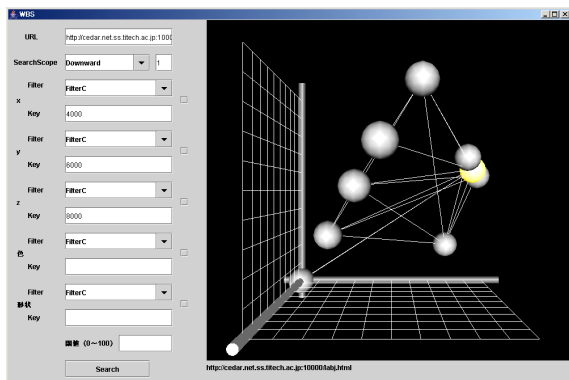


図9 試作ブラウザ

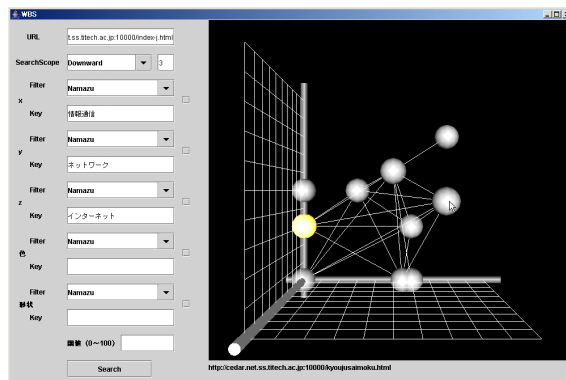


図11 検索結果表示

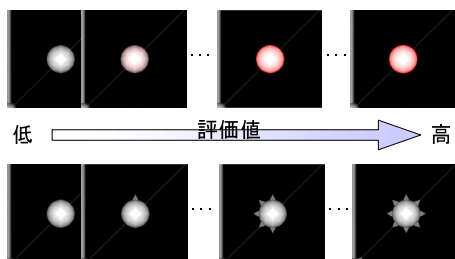


図10 ノードの色と形状

および検索評価軸で指定した検索フィルタによる検索のための検索キー入力欄が設けられている。

右側の結果表示は、画面左奥に評価値の原点が存在し、軸を表す棒が表示されている。したがって、画面右上手前にすべての評価軸で評価値の高いコンテンツがノードとして配置される。

画面左側の検索情報指定領域で指定する、画面上への表現軸としては、一般的な3次元空間を構成するx, y, z軸以外に、ノードの色調が評価値に従い変化することで評価値を表現するノードの色軸と、ノードの形状が評価値に従い変化することで評価値を表現するノードの形状軸を導入し、計5次元での表現ができるようにした。ノードの色軸は、投影した評価値が高ければ高いほど赤みが増すように、ノードの形状軸は、投影した評価値が高ければ高いほどノードに生えるトゲの数が増すようになっている。

4.3 検索フィルタ

検索フィルタには、サンプル用としてSize, Extension, Namazuを用意した。Sizeは、URLで指定したコンテンツのサイズ(バイト数)とキーで指定した数値との近さを評価値とする。Extensionは、URLで指定したコンテンツの拡張子とキーとの一致度を評価値とする。Namazuは、全文検索を行うフィルタで、キーに対する適合度を評価値とする。

4.4 検索例

図9は、x, y, z軸に別々の検索結果を投影した図である。奥の黄色いノードがstartURLで、そこからのリンク構造が検索評価値による配置と共に示されている。ユーザは、この配置の様子をみて、複数の検索評価値におけるノードの位置づけだけでなく、Webにおいて重要なリンク構造をも一度に把握することが可能である。

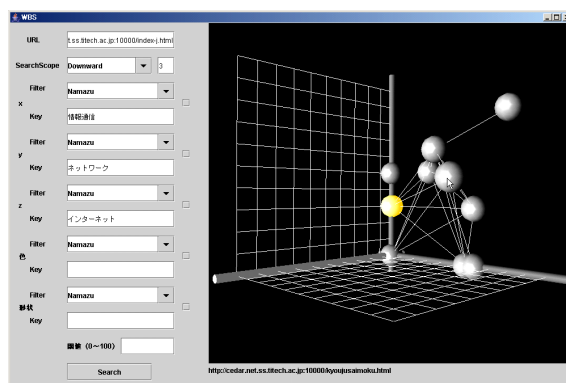


図12 検索結果表示(回転時)

4.5 評価実験

これまで述べたような提案手法の効果が、実際に実現されているかを評価するため、実際のインターネット上のデータを用い、実験とそれによるアンケートを行った。

4.6 評価実験と結果

以下の筋書きで検索をかけた。

情報通信やネットワーク、インターネットに関する研究を行っている研究室を探している。東京工業大学集積システム専攻のURLを事前にしており、そこから近傍検索を行った。x軸に、“情報通信”で全文検索を行った結果を、y軸に、“ネットワーク”で全文検索を行った結果を、z軸に、“インターネット”で全文検索を行った結果をそれぞれ投影した。

その結果、図11, 12で示される検索結果が得られた。検索結果表示でマウスオーバーしているノードは、教授細目書が書かれているコンテンツで、講義に関する説明がなされており、残念ながら求めていた情報ではない。最も右上に配置されているノードは、“植松・松本研究室”で、研究分野は情報通信システム講座ネットワーク構成分野であり、求めていた情報であると言える。最もリンクが集中しているノードは、組織構成が書かれているコンテンツで、各研究室の紹介と、その研究室へのリンクが張られており、求めている情報に関して重要なコンテンツであると考えられる。

得られた検索結果を見て、アンケートに答えてもらった。被験者は5人であり、全員提案手法を既に知っている。

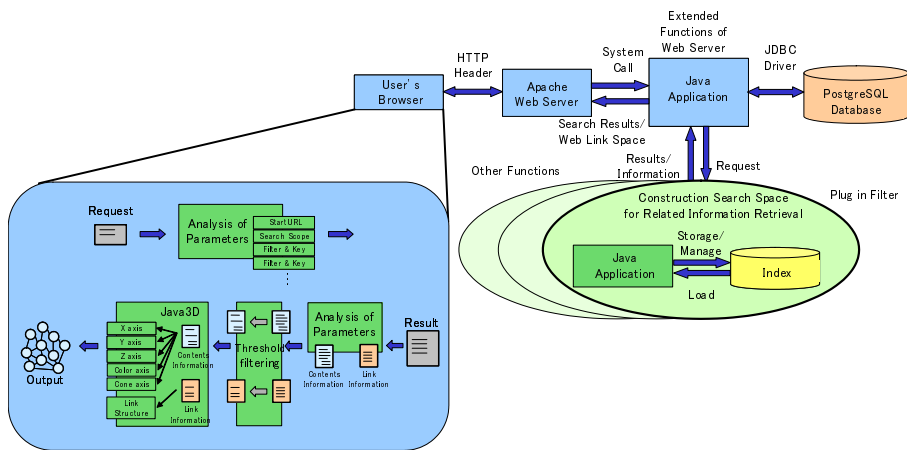


図 8 試作システムの構成

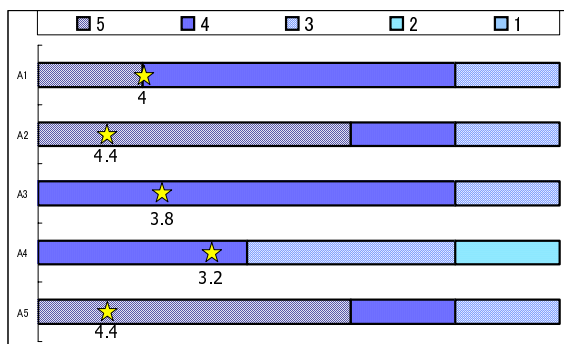


図 13 アンケート A の結果

4.6.1 アンケート A

提案手法 (図 11, 12) に関して A1 から A5 までの質問に, 5 段階評価で答えてもらった. 質問の内容は,

- A1 3次元表示はどの程度有効だったと思うか?
- A2 評価値による配置方法は, どの程度理解しやすかったか?
- A3 リンク関係はどの程度理解しやすかったか?
- A4 外見的なノードの連結状態から, 文章の内容や重要度を判断できたか?
- A5 移動, 回転, ズーム操作は, 表示上の混雑などで見づらい部分をみるために役だったか?

で, 基準は,

- 5: 有効である.
- 4: どちらかといえば, 有効である.
- 3: どちらともいえない.
- 2: どちらかといえば, 有効でない.
- 1: 有効でない.

である. アンケートの結果を図 13 に示す.

4.6.2 アンケート B

次に, 提案手法 (図 11, 12) と単純なリスト表示を見比べてもらい, B1 から B3 までの質問に答えてもらった.

B1 キーワードを含むコンテンツで, 沢山のコンテンツから参照されているコンテンツを見つけるのはどちらが簡単か?

(1) 沢山のコンテンツからなら, 提案方式の方が見やすい. しかし, 密集しすぎると, リスト表示の方が見やすいのでは.

(2) どちらとも言えない.

(3) 提案手法. 沢山のコンテンツから参照されているという状況が視覚的にすぐにわかるから.

(4) 提案手法.

(5) 提案手法. リストでは, 実際にページを開かなければならない手間がある.

B2 どれが求めている情報が, どちらが理解しやすいか?

(1) 絞り込みは提案方式の方が有効であると思う. しかし, 具体的なものを一つ探したいときは, サイトについて説明のあるリスト表示の方が探しやすいのでは.

(2) 提案方式.

(3) 単純なリスト表示. 人間の理解しやすい言葉でコンテンツが説明されているから. 提案手法だと, 各軸が何を意味しているのか等を頭の中で思い浮かべながら理解しなければならないから.

(4) タイトルが出ているのでリスト表示の方が分かりやすい.

(5)じっくり調べたい時は提案手法. 素早く調べたい時はリスト表示. 情報量の多さとスピードがトレードオフとなっている.

B3 論理演算検索を行いたい時, どちらが便利か?

(1) 提案方式だと思う.

(2) 提案方式.

(3) 提案手法. 自分の重視する項目に対して視覚的にソートすることができているから.

(4) キーワード数が少なければ提案手法の方が便利だと思うが, 7,8 個キーワードを使う場合はリスト表示の方が便利だと思う.

(5) 提案手法. どのキーワードが効いているのかがすぐ分かる.

4.7 実験結果の考察

今回の実験では, 実装した表現軸のうち3次元空間を作り出す x, y, z 軸を使い, 三つのキーワードを用いた全文検索の検索結果として表示を得た. アンケート A における質問 A1, A2 の結果から, 3次元表示や評価値による3次元空間への配置方法

が有効であることが示された。また、質問 A3 の結果から、リンクを線として表示することで、リンク関係の理解が容易になることが示された。質問 A4 の結果からは、外見的なノードの連結状態だけでは、文章の内容や重要度を判断するのは容易でないことが示された。コンテンツはノードとしてのみ表示される。ノードの色軸や形状軸に評価値を投影すれば重要度は理解できると思われるが、ノード単体だけでは文章の内容は表現しきれないので、別途考える必要がある。アンケートとは別に書いてもらった感想より、タイトルと要約などコンテンツに関する情報を表示する機能が求められている。質問 A5 の結果から、操作の有効性と必要性が示された。

アンケート B における質問 B1, B3 の結果から、例外があるものの提案手法の方が有用であることが示され、研究の目的が果たされていると言える。質問 B2 の結果からは、質問 A4 の結果からも示されたノード単体を表示するだけでなく、タイトルなど説明が必要であることが示されている。また、ユーザが自ら判断できるよう複数の検索評価値を別々に表示することを提案しているが、逆にユーザが判断しなければならない情報が増えたというマイナス面が指摘された。B2(5) の回答のように、リスト表示と提案手法による表示を切り替えて使えるように実装することも考慮しなければならない。

自由文記述による感想では、リンクに関して「リンクの多さから直観的に重要度がわかりやすい」、「リンクの段数が線の太さなどに反映されるとより解釈しやすい」、「リンクの向きで参照している/されていることが分かると良い」といった多数の回答が得られた。これはリンクを表示することへの関心の高さの表れであり、提案手法ではリンクは線のみでしか表示しなかったが、よりリンクの意味を解釈しやすい表示方法を考えるべきである。また、ノード自体に関する回答はなかったが、「ノードをマウスオーバーしたときにタイトルと要約が出る機能が必要」という回答を得た。単純なリスト表示を包含する意味でも、この機能は必要であると考えられる。ただし、タイトルと要約が結果表示の邪魔にならぬよう、同じ空間に表示するのであれば配置を上手く考えなければならない。「密集しすぎると、リスト表示の方が見やすいのでは」という回答に対しては、まだ実装は行えてないが、fisheye view [19] の技術を使い、密集部分の詳細を広げて表示することで解決できると考えられる。

他に、検索結果表示には関係ないが、「表示画面が殺伐としていたので、もう少し柔らかな雰囲気欲しい」という回答も得られた。ユーザインターフェースとしてユーザフレンドリーな表示は必要であるため、その点も考慮しなければならない。

以上のことより、今後求められる機能として、リンクの表示方法の工夫やマウスオーバーしているノードからコンテンツ情報が引き出せるような仕組みなどが挙げられる。また、今回はある条件下での検索結果としての表示のみを評価したが、今後はさらに幅広い検索シーンにおいて、多くの実データにおける実験を行い、評価をしていきたい。

5. おわりに

本稿では、近傍情報探索システム PIRCS の一環としてクラ

アントに適応でき、ナビゲーション支援としての Web リンク空間の可視化と複数キーワードによる検索結果の同時表示を可能とし、ユーザが直観的に理解できる表示方法を提案した。シナリオ例によって、リンクを表示することによるナビゲーション支援としての利点や複数キーワードによる検索結果の同時表示の利点を説明した。さらに、試作システムを構築し、評価実験を行い、ある条件下での検索における提案手法の有効性について述べた。今後の課題としては、今回の評価実験では使用しなかったノードの色軸や形状軸といった表現軸の評価、他に表示に有効な表現軸の考察、システムとしてユーザから求められている機能の洗い出しと実装などが挙げられる。

文 献

- [1] 久我昌崇, 中所属武司, “Web コミュニティの知識に基づく情報検索手法の評価”, 情報処理学会研究報告 DBS123-6, IPSJ, 2001.
- [2] 豊田正史, “WWW における関連コミュニティ群の発見”, 情報処理学会研究報告 DBS122-40, IPSJ, 2000.
- [3] 今藤紀子, 喜連川優, “ウェブグラフにおける密サブグラフの抽出とウェブコミュニティ”, 信学技報, DE2002-53, July, 2002.
- [4] 豊田正史, 喜連川優, “ウェブコミュニティの発展過程抽出手法”, 信学技報, DE2002-7, May, 2002.
- [5] 福地健太郎, 豊田正史, 喜連川優, “Web Community Browser: 大規模 Web コミュニティチャートの可視化”, DEWS2002 論文集, A1-4, 2002.
- [6] <http://www.yahoo.co.jp>
- [7] <http://www.google.com>
- [8] U. Manber, M. Smith, B. Gopal, “Webglimpse: Combining browsing and searching”, In Proc. of 1997 Usenix Technical Conference, 1997.
- [9] I. Ben-Shaul, M. Herscovici, M. Jacovi, Y. Maarek, D. Pelleg, M. Shtalham, V. Soroka, and Sigalit Ur, “Adding support for dynamic and focused search with fetuccino”, WWW8/Computer Networks, Vol.31, No.11-16, pp.1653-1665, 1999.
- [10] Kobayashi Aki, Yamaoka Katsunori, Sakai Yoshinori, “Cooperative Web Architecture for Search and Navigation Assistance”, Proc. ICWI2002 IADIS WWW/Internet 2002, pp.726-729 (2002)
- [11] Kobayashi Aki, Miyahara Hidetomo, Tochihara Kaito, Wang Hengjiang, Yamaoka Katsunori, Sakai Yoshinori, “PIRCS: A Link Context Based Search on The Web”, 2003 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, pp.514-517 (2003)
- [12] Aki Kobayashi, Kuangmin Tan, Katsunori Yamaoka, Yoshinori Sakai, “Relevant information retrieval for cooperative Web architecture”, In Proc. of IADIS International Conference WWW/Internet 2004, pp.1125-1128, Oct. 2004.
- [13] Lamping, J., Rao, R., Pirolli, P., “A focus+context technique based on hyperbolic geometry for visualizing large hierarchies”, Proc. ACM CHI'95, 1995.
- [14] Tamara Munzner, “H3: Laying out large directed graphs in 3d hyperbolic space”, IEEE InfoVis'97, pp. 2-10, 1997.
- [15] Robertson, G.G., Card, S.K., Mackinlay, J.D., “Cone Trees: Animated 3-D visualization of hierarchical information”, Proc. ACM CHI'91, pp.189-194, 1991.
- [16] 塩澤秀和, 西山晴彦, 松下温, “「納豆ビュー」の対話的な情報視覚化における位置づけ”, 情報処理学会論文誌, vol.38 No.11, pp.2331-2342, Nov. 1997.
- [17] Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd, “The PageRank Citation Ranking: Bringing Order to the Web”, Stanford Digital Library Technologies Project, 1998.
- [18] 磯部成二, 黒川清, 塩原寿子, 飯塚哲也, “視覚化による多次元データ分析システム: INFOVISER”, 情報処理学会論文誌, vol.40 No.5, pp.2417-2428, May 1999.
- [19] Furnas, G.W., “Generalized fisheye views”, Proc. ACM CHI'86, pp.16-23, 1986