

空間情報ハブ抽出のためのウェブリンク解析手法

Web Link Analysis for Extracting Spatial Information Hub Pages

張 建偉[◇] 石川 佳治[♡] 北川 博之[♡]

Jianwei ZHANG Yoshiharu ISHIKAWA
Hiroyuki KITAGAWA

ウェブの爆発的な拡大から、大量のウェブページの中から有用な情報を抽出するためのウェブマイニングの研究が盛んに進められている。一方では、携帯機器や GPS の普及から、特定の地域に関するウェブ情報を提供する研究が進められている。本研究では、これらの動向を踏まえ、ウェブページ群からの空間情報ハブの抽出手法の開発を行う。本手法では、地理的な情報をもとに拡張されたウェブのリンク構造を解析することにより、着目している地域におけるページの評判や有用性を判定する。ウェブのグラフ構造を、着目している地理領域に関連した空間ノードと空間リンクにより拡張する点の特徴となっている。

Recently web mining that tries to find relevant information from the vast amount of web pages has attracted a lot of research interests. Besides, it is becoming an important task to provide information related to a user-specified geographic area. In this paper, we propose a method to extract spatial information hub pages. A spatial information hub is a webpage which is related to a specified geographic area and has much local information or many hyperlinks to local web pages. We employ geographic information to create spatial nodes and spatial links, and then conduct link analysis based on the extended link structure.

1. はじめに

ウェブの爆発的な拡大により、大量のウェブページの中から有用な情報を抽出する技術はより重要性を増している。そのための手法として、近年ウェブマイニング [1, 2] の研究が盛んに進められている。特に、ウェブページ間のリンク情報を用いるリンク解析は、評判の高いウェブページを特定するための重要な技術となっている。

一方では、携帯機器や GPS の普及などにより、位置に応じて適切な情報提供を行うためのサービスが現在重要となってきた。そのようなサービスの 1 つとして、指定された地点の周辺情報に関するウェブページ群をユーザに提示する、地域性を考慮した検索エンジンや、特定の地域に関するウェブ情報を提供する研究が進められている [3, 4, 5]。これらの研究では、ウェブページのコンテンツ解析やリンク解析の手法を用いて、地域性を有するウェブページ群の特定などを行っている。

これらの関連研究の動向を踏まえ、本研究ではウェブページ群からの空間情報ハブの抽出手法の開発を行っている。本手法の特

徴の一つは、ハイパーリンクで構成されるウェブのグラフ構造を、着目している地理領域に関連した空間ノードと空間リンクにより拡張する点にある。従来のウェブのリンク解析がウェブのリンク情報をもとにページの評判や有用性を判定していたのに対し、本手法では地理的な情報をもとに拡張されたウェブのリンク構造を解析する。これにより、着目している地域における評判や有用性という面も考慮してページを評価する。このような意味で、本研究では空間情報ハブを、

ある地域に関するウェブページや地理情報に関して有用な多くのリンク（ハイパーリンクおよび空間リンク）を張っているウェブページ

と定義する。ウェブ上からこのようなページを抽出することができれば、ある地域に関するポータルサイトとして活用することが可能となる。

2. 関連研究

2.1 ウェブマイニング

ウェブマイニング [1, 2] は、大別すると、ウェブページのコンテンツマイニング、リンク解析、ウェブサイトのログ解析の 3 つのアプローチに分けることができる。本研究では、リンク解析を中心に空間情報ハブの抽出を図る。リンク解析手法の代表例としては、Google で用いられている PageRank [6] や、ユーザが指定したトピックに関してハブとオーソリティのページを抽出する HITS [7] が挙げられる。また、[8] では、ウェブアーカイブからウェブコミュニティを抽出し、関連するウェブコミュニティ同士を辺で結んだグラフを作成する手法を提案されている。[8] がオーソリティ及びハブの集合をコミュニティとして扱ったのに対し、我々は地域性を考慮し、特に空間情報ハブページの抽出を目的とする。

ここでは特に本研究で拡張を図る HITS (Hypertext Induced Topic Selection) [7] について簡単に説明する。まず、前もってユーザが指定したキーワードにより数百ページ程度のウェブページを検索エンジンなどで抽出してルートセットとする。次に、ルートセット内のページからリンクされているページの集合とルートセット内のページをリンクしているページの集合を検索エンジンなどを利用して求める。これらのページ群から構成されるウェブの部分グラフを V とする。

HITS のアルゴリズムを図 1 に示す。各ページのハブ度（そのページが良いオーソリティのページをリンクしている指標）をベクトル h で、オーソリティ度（そのページが良いハブからリンクされている指標）をベクトル a で表す。

```

1  $1 := [1, \dots, 1] \in \mathcal{R}^{|V|}$  // スコアを初期化
2  $a_0 := h_0 := 1$ 
3  $t := 1$ 
4 repeat
5   foreach  $v \in V$  do // オーソリティ度の更新
6      $a_t(v) := \sum_{w \in \text{parent}[v]} h_{t-1}(w)$  // ハブ度の更新
7      $h_t(v) := \sum_{w \in \text{child}[v]} a_{t-1}(w)$ 
8   end // スコアの正規化
9    $a_t := a_t / \|a_t\|$ 
10   $h_t := h_t / \|h_t\|$ 
11   $t := t + 1$  // スコアが収束するまで
12 until  $\|a_t - a_{t-1}\| + \|h_t - h_{t-1}\| < \epsilon$ 
13 return  $(a_t, h_t)$ 

```

図 1: HITS のアルゴリズム

Fig. 1: HITS algorithm

2.2 地域性を考慮したウェブ情報の収集・探索

ウェブの中から特定の地域に関するページを抽出するための研究としてさまざまな手法が提案されている。[3] では、位置情報をウェブから収集する手法について述べている。[4] では、ウェブページ内に含まれる地名・組織名などの地理情報、ページ内の話題

[◇] 学生会員 筑波大学システム情報工学研究科
zjw@kde.cs.tsukuba.ac.jp

[♡] 正会員 筑波大学システム情報工学研究科/計算科学研究センター
{ishikawa.kitagawa}@cs.tsukuba.ac.jp

の偏在性、話題の注目度など、さまざまな要素を考慮して、ページのローカル度を与える手法を提案している。ローカル度は、そのページが地域密着型の情報を有しているかの判断に利用する。ローカル度のアプローチとは異なるが、本研究においてもウェブページがどの程度地域密着型の情報を表しているかというローカル性を考慮している。ローカル性が高いページやローカル性が高いページを多数リンクしているページをより高く評価するようにリンク解析処理を工夫している。

[5]では、地域情報サービスを提供するため、ウェブ空間を拡張するアプローチを述べている。通常のウェブページ間のリンク以外に地理空間上へのリンクを用いてウェブを拡張することで、地理空間を経由したウェブ空間のナビゲーションが可能となる。一方、本研究では、着目している地理情報を表す空間ノードと、空間ノード間および空間ノードとウェブページ間の空間リンクを導入することで、ウェブ空間を拡張する。[5]が拡張したウェブ空間をユーザに提供するナビゲーション機能に利用していたのに対し、本研究ではリンク解析によるウェブページからの情報抽出・知識発見に用いる点が大きく異なる。

3. 提案手法の概要

3.1 ウェブページ群からの空間情報の抽出

前処理として、ウェブデータの収集を行い、収集した各ウェブページの中から、空間情報の抽出を行い、その座標値を計算する。空間情報の抽出においては、正確性を重視して、正確な座標が特定できるような7桁の郵便番号などの情報を利用する。実際のウェブページには、これら以外にも地理情報に関連付け可能な多数の情報が含まれている。しかし、部分的な地名(例:「東京」「春日」)などは抽出が困難であり、着目している地理領域内に含まれる地理情報に対応する地点を正確に特定することが難しいため、抽出は行わない。以上の処理により、各ウェブページには一般に0個以上複数個の座標値が対応することになる。

3.2 ベースセットの構築

リンク解析処理は、ユーザから分析対象の地理領域が指定された時点で開始する。指定された地理領域に対し、その中に関連する座標値の少なくとも1つが含まれるようなページを収集し、これをルートセットとする。HITSと同様、このルートセットのページに関連するページを追加し、ベースセットを構築する。すなわち、ルートセットに含まれるページへのリンクを有するウェブページ、および、ルートセットに含まれるページがリンクしているウェブページの追加を行う。これにより、図2の左側のような、指定された地理領域に関連するページ群からなるウェブ空間のサブグラフを構築する。

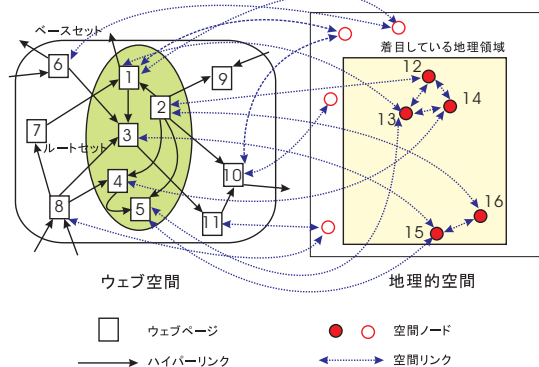


図2: 拡張ベースセットの例
Fig. 2: Example of an extended base set

HITSではユーザ指定のキーワードをもとにルートセットを構築するのに対し、本手法では空間的な条件をもとに構築する点に大きな違いがある。

3.3 空間ノード・空間リンクの生成

次に、空間ノードと空間リンクを生成する。空間ノードとは、ベースセット内のウェブページ中に含まれる住所・郵便番号などの地理的情報に対応するノードである。ただし、同じ表記(例:同一の郵便番号や住所表記)については空間ノードを1つだけ生成する。

次いで、空間ノードとその情報を含んでいたページ間に双方向のリンクを作成する。このリンクは、ウェブのハイパーリンクと異なり、空間的な関連性に基づくことから、空間リンクと呼ぶ。また、2つの空間ノード間の距離がある閾値以下である場合にも、お互いが近傍にあるものと考え、双方向の空間リンクを生成する。

本手法では、空間ノードと空間リンクを前節で説明したベースセットに追加することにより、ベースセットを拡張する。以下ではこれを拡張ベースセット(図2)と呼ぶ。あいまい性が生じない場合には、そのままベースセットと呼ぶ。これにより、たとえば2つのウェブページ間にハイパーリンクのつながりがない場合でも、位置が近い地理情報を含んでいる場合などにはお互いの間に関連性が生じることになる。このように拡張されたベースセットをもとにリンク解析を行うことで、本手法では地理的な情報も考慮してウェブページの評判や有用性を評価する。

以上のように構築された拡張ベースセットのグラフ構造は、ウェブ空間上における近さ(意味的な関連や組織・社会的な関連を反映)と地理空間上の近さを融合したものとなっている。

3.4 リンク解析処理

3.4.1 リンクの入出度数に基づく指標の導入

本手法では、図2のように、各ウェブページ内に現れるすべての空間情報に対応して空間ノードが存在し、そこへの空間リンクがあると考えられる。しかし、実際のリンク解析では、着目している地理領域内の空間ノード、および、それらへの空間リンクのみを分析対象とする。つまり、リンク解析の際には、仮想的に存在する空間ノード、空間リンクのうち、着目している地理領域に関連するもののみを実際に考慮する。ただし、複数の空間情報を含むウェブページの場合、それらの地理的位置が離れていたり、着目している領域外の空間情報が含まれる場合も多いため、本研究では、着目領域内への空間リンクを多数有しているページの重要性を高め、他の領域の情報を多く含むページの重要性を低くするために、ノードに重み付けを行うリンク解析手法[9]の考え方を拡張し、以下の2つの指標を定義する。

ノード v から出るリンク(ハイパーリンクおよび空間リンク)の総数を $outlinks(v)$ とし、そのうち拡張ベースセット内のノードを指すリンクの数を $effective_outlinks(v)$ とする。このとき、 $out_ratio(v)$ を、

$$out_ratio(v) = \frac{effective_outlinks(v) + 1}{outlinks(v) + 1} \quad (1)$$

と定義する。この値は、含んでいるリンクが拡張ベースセット内のノードを指すほど高くなり、拡張ベースセット外のノード(地理的な意味で関連が低いノード)を指すほど低くなるという性質がある。

たとえば、図2において、ノード1から出るリンクは5本があるが、そのうち拡張ベースセット内へのリンクは2本である。よって、ノード1については、

$$out_ratio(1) = \frac{2 + 1}{5 + 1} = \frac{1}{2} \quad (2)$$

となる。

同様に、ノード v に入るリンクの総数を $inlinks(v)$ とする。そのうち、拡張ベースセット内のノードからのリンク数を $effective_inlinks(v)$ とする。これらをもとに、 $in_ratio(v)$ を、

$$in_ratio(v) = \frac{effective_inlinks(v) + 1}{inlinks(v) + 1} \quad (3)$$

と定義する。

3.4.2 拡張した HITS によるリンク解析

本手法では、前節で提案した in/out_ratio を HITS のアルゴリズムに導入する。これにより、拡張ベースセット内のノードよりも外のノードと関連が深いノードのハブ度、オーソリティ度のスコアをより低く評価することができ、逆に、拡張ベースセット内のノードと密接に関連しているノードの評価を高くできると考えられる。

拡張した HITS のアルゴリズムを図 3 に示す。6 行目と 7 行目で上記の指標による重み付けを行っている点が相違点である。

```

1  $1 := [1, \dots, 1] \in \mathcal{R}^{|V|}$ 
2  $a_0 := h_0 := 1$ 
3  $t := 1$ 
4 repeat
5   foreach  $v \in V$  do
6      $a_t(v) := in\_ratio(v) \times \sum_{w \in parent[v]} h_{t-1}(w)$ 
7      $h_t(v) := out\_ratio(v) \times \sum_{w \in child[v]} a_{t-1}(w)$ 
8   end
9    $a_t := a_t / \|a_t\|$ 
10   $h_t := h_t / \|h_t\|$ 
11   $t := t + 1$ 
12 until  $\|a_t - a_{t-1}\| + \|h_t - h_{t-1}\| < \varepsilon$ 
13 return  $(a_t, h_t)$ 

```

図 3: 拡張した HITS のアルゴリズム

Fig. 3: Extended HITS algorithm

3.4.3 スコアの計算例

図 2 の拡張ベースセットに対するハブ度、オーソリティ度の計算結果を表 1 に示す。ハブ度が最も高いノードはノード 2 で、ノード 4, 5 がそれに続いている。ウェブページとして見た場合、ノード 5 は外向きのリンクがなくハブとはなり得ないが、本手法では良質の空間情報を有している（良質の空間リンクを多数有している）と評価されて、高いハブ度となっている。

表 1: 提案手法によるスコアの計算結果

Table 1: Hub/authority scores (proposed method)

ノード番号	ハブ度 (×100)	オーソリティ度 (×100)
1	8	19
2	67	7
3	2	14
4	27	33
5	8	56
6	2	0
7	6	1
8	4	0
9	0	22
10	0	6
11	1	0
12	18	47
13	48	21
14	32	30
15	32	4
16	4	32

比較のため、従来の HITS アルゴリズム (図 1) により、拡張前のベースセット (3.2 の処理で生成されたグラフ) を分析した場合のスコアを表 2 に示す。ハブ度について見た場合、ノード 2 が高いスコアとなっている点は提案手法と同様であるが、ノード 5 の評価が低く、逆にノード 8 の評価が高い点などが異なる。提案手法では、ノード 8 は空間情報との直接的な結びつきがないため低く評価されているが、従来の HITS では高く評価している。

4. 実験

4.1 実験の準備

実験に利用するのは国立情報学研究所により提供されている NTCIR-4 WEB タスク文書データ (NW100G-01) である。このデータは文書 11,038,720 件を収集し、79,699,256 個のリンク

表 2: 従来の HITS による計算結果

Table 2: Hub/authority scores (original HITS)

ノード番号	ハブ度 (×100)	オーソリティ度 (×100)
1	9	43
2	88	0
3	0	22
4	17	50
5	0	43
6	9	0
7	17	15
8	35	0
9	0	36
10	0	43
11	17	0

が含まれている。

前処理として、NW100G-01 データから各ページのメタ情報 (URL など)、ページの間のリンク情報などをデータベースに追加した。空間情報に関しては、7 桁の郵便番号を中心に情報の抽出と空間的な位置との対応付けを行い、データベースに登録を行った。

4.2 実験結果

4.2.1 in/out_ratio 導入に関する比較

第一の実験として、空間ノード間に空間リンクを張るかどうかの距離の閾値を 0.002 に限定し、着目領域を東京都豊島区池袋本町 (〒1700011) から距離が 0.015 の地域に指定する。なお、ここでの距離は、緯度、経度を x 座標、 y 座標と近似的にみなした座標値をもとに計算する。

提案手法によるハブ度の上位 6 件のリンク情報を表 3 に示す。 doc_id は文書の ID である。 s_link は文書中の全ての空間リンク数で、 eff_s_link はそのうち着目している地域内の空間ノードへの空間リンク数である。 w_link は文書中の全てのウェブリンク数で、 eff_w_link はそのうちベースセット内のウェブページへのウェブリンク数である。 out_ratio は全てのリンク (ウェブリンクと空間リンク) のうち有効なリンクの割合である。 eva は各ページに対して、良いハブページかどうかの評価である。なお、各ページの内容とリンク先のページの内容を見ることで評価した。は良いハブページを表し、 \times は良くないハブページを表す。のページはどちらとも言えないものである。

表 3: 提案手法によるハブ度上位 6 件のリンク情報 (豊島区)

Table 3: Link classification in top-ranked hubs (Toshima)

doc_id	s_link	eff_s_link	w_link	eff_w_link	out_ratio	eva
13039139	8	6	0	0	0.7777	
13513770	11	6	0	0	0.5833	
7160587	12	5	0	0	0.4615	
6641959	8	2	8	8	0.6470	\times
10565146	8	2	0	0	0.3333	\times
10585207	20	3	0	0	0.1904	\times

比較のため、空間ノードと空間リンクを生成するが、 in/out_ratio の指標を導入しない場合のハブ度上位 6 件のリンク情報を表 4 に示す。

表 4: in/out_ratio を導入しない場合のハブ度上位 6 件のリンク情報 (豊島区)

Table 4: Link classification in top-ranked hubs without in/out_ratio (Toshima)

doc_id	s_link	eff_s_link	w_link	eff_w_link	eva
10822643	62	4	26	26	\times
9321218	371	3	22	22	\times
11724304	209	4	31	31	\times
13039139	8	6	0	0	
13513770	11	6	0	0	
7160587	12	5	0	0	

表 4 で示されたように、 in/out_ratio の指標を導入しないと、ハブ度が上位になるページはウェブリンクと空間リンクを多く持つが、そのうち着目している地域内への空間リンクの割合が低い。即ち、着目している地域以外の空間情報を多く含むという欠点がある。

表 3 で示されたように、 in/out_ratio を導入することで、このよ

うなページはより低く評価される。

4.2.2 閾値の影響

第二の実験として、空間ノード間に空間リンクを張るかどうかの距離の閾値 τ を 0.005 に限定し、着目地域をつくば市天王台 (τ 3001253) から距離が 0.05 の地域に指定する。提案手法によるハブ度の上位 6 件のリンク情報を表 5 に示す。

表 5: 閾値 0.005 の場合のハブ度上位 6 件のリンク情報 (つくば)

Table 5: Link classification in top-ranked hubs(Tsukuba, $\tau = 0.005$)

doc_id	s_link	eff_s_link	w_link	eff_w_link	out_ratio	eva
6608658	0	0	630	45	0.0729	x
6191997	0	0	154	24	0.1612	x
9795990	0	0	92	17	0.1935	
7511229	0	0	30	9	0.3225	
13941808	1	1	10	10	1	
8254577	0	0	43	1	0.0454	x

空間リンク作成のための閾値 τ の影響を比較するために、同じく着目地域をつくば市天王台 (τ 3001253) から距離が 0.05 の地域に指定し、閾値 τ を 0.007 に増やした。提案手法によるハブ度の上位 6 件のリンク情報を表 6 に示す。

表 6: 閾値 0.007 の場合のハブ度上位 6 件のリンク情報 (つくば)

Table 6: Link classification in top-ranked hubs(Tsukuba, $\tau = 0.007$)

doc_id	s_link	eff_s_link	w_link	eff_w_link	out_ratio	eva
4968622	15	12	0	0	0.8125	
4968829	5	3	0	0	0.6667	
9921832	9	3	0	0	0.4	
12469677	52	10	44	44	0.5670	x
12343866	8	4	5	5	0.7143	x
7204947	9	3	5	5	0.6	

閾値 τ が小さい場合、空間リンクを持たないが有効なウェブリンクを有しているページ (つまり、リンク先のページが着目地域内の空間情報を持つページ) が上位となっている。閾値 τ を増やすことで、ウェブリンクを持たないが空間リンクを多数有しているページがより上位となっている。

なお、比較のため、ウェブリンクを用いず、空間リンクのみを用いた実験も行った (実験結果は省略)。ウェブリンクと空間リンクの双方を用いる手法とほぼ類似した結果を得たが、例えば表 5 の第 3 番目のページのような良いハブページは上位とはならなかった。このページはつくば市の研究機関、学校などのホームページのリンク集であり、それ自身は空間情報を含んでいないが、有効なウェブリンクを有している。すなわち、このページのリンク先のページは指定された地域内の空間情報を有している。単に空間リンクだけではこのようなページを良いハブとして抽出できないことから、本手法でウェブリンクと空間リンクを併用することの利点を確認できた。

5. まとめと今後の課題

本研究では、特定の地理領域に関し有用なリンクを張っている良質のウェブページである空間ハブを、ウェブのリンク解析によって求める手法を提案した。また、実験により、 in/out_ratio の有用性や、空間リンク作成のための閾値設定の影響を見た。

以下、今後の課題を挙げる。

1. 提案手法の改良

- in/out_ratio 定義式の工夫: 今回の提案手法では、有効なウェブリンクと空間リンクの割合を一つの指標にまとめ計算したが、それらを分離してウェブリンクの in/out_ratio 、空間リンクの in/out_ratio とすることも考えられる。
- リンクの重み付け: 今回の提案手法では、ウェブリンクと空間リンクを同様に扱ったが、ウェブリンクと空間リンクに異なる重みをつけることも考えられる。
- 閾値の決定: 今回の実験では、空間リンク作成のための閾値は実験的に設定したが、適切な値を自動的に設

定する方式の開発が求められる。例えば、空間ノード間の空間リンク数が一定となるように閾値を選ぶ方式が考えられる。

- HITS 以外のアルゴリズム: 今回の提案手法では、HITS をベースとしたが、HITS 以外のアルゴリズムを拡張することも検討したい。
- データの収集: 今回の実験では、NW100G-01 データを利用したが、着目する地域について、このデータの中に必ずしも良い空間ハブがないこともあった。また、リンク先のページが必ずしもデータセットに含まれていないため、精度の低い解析となることも考えられる。データを有効に収集するための特定の領域に関する空間情報のクロウリングも今後の課題である。

【謝辞】

本研究の一部は、日本学術振興会科学研究費基盤研究 (C)(2)(16500048)、旭硝子財団研究助成、および CREST「自律連合型基盤システムの構築」による。

【文献】

- S. Chakrabarti, *Mining the Web: Analysis of Hypertext and Semi Structured Data*, Morgan Kaufmann, 2002.
- P. Baldi, P. Frasconi, and P. Smyth, *Modeling the Internet and the Web: Probabilistic Methods and Algorithms*, Wiley, 2003.
- 横路誠司, 高橋克巳, 三浦信幸, 島健一, 位置指向の情報の収集, 構造化および検索手法, 情報処理学会論文誌, Vol. 41, No. 7, pp. 1987-1998 (2000).
- C. Matsumoto, Q. Ma, and K. Tanaka, Web Information Retrieval Based on the Localness Degree, *Proc. DEXA 2002*, LNCS 2453, pp. 172-181 (2002).
- 平松薫, 石田亨, 地域情報サービスのための拡張 Web 空間, 情報処理学会論文誌: データベース, Vol. 41, No. SIG 6(TOD 7), pp. 81-90 (2000).
- S. Brin and L. Page, The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, Vol. 30, pp. 1-7 (1998).
- J.M. Kleinberg, Authoritative Sources in a Hyperlinked Environment. *JACM*, Vol. 46, No. 5, pp. 604-632 (1999).
- 豊田正史, 吉田聡, 喜連川優, ウェブコミュニティチャート 膨大なウェブページを関連する話題を通して閲覧可能にするツール, 電子情報通信学会論文誌 D-I, Vol. J87-D-I, No. 2, pp. 256-265 (2004).
- K. Bharat and M.R. Henzinger, Improved Algorithms for Topic Distillation in a Hyperlinked Environment, *Proc. SIGIR*, pp. 104-111 (1998).

張 建偉 Jianwei ZHANG

筑波大学システム情報工学研究科在学中。Web とデータベースの研究に従事。日本データベース学会、情報処理学会学生会員。

石川 佳治 Yoshiharu ISHIKAWA

筑波大学システム情報工学研究科/計算科学研究センター助教授。データベース、データ工学、情報検索などに興味を持つ。日本データベース学会、情報処理学会、電子情報通信学会、人工知能学会、ACM、IEEE CS 各会員。

北川 博之 Hiroyuki KITAGAWA

筑波大学システム情報工学研究科/計算科学研究センター教授。異種情報源統合、データマイニング、文書データベース、WWW の高度利用などの研究に従事。日本データベース学会、情報処理学会、電子情報通信学会、日本ソフトウェア科学会、ACM、IEEE CS 各会員。著書に「データベースシステム」(昭晃堂) など。