

共同企業体結成のための意思決定支援に関する研究

関西大学大学院 中村健二*1

関西大学 田中成典*2

関西大学 古田 均*2

By Kenji NAKAMURA*1, Shigenori TANAKA*2, Hitoshi FURUTA*2

現在、大規模工事においては、質の高い施工を実現するために複数の企業が共同企業体を結成し対応するケースが増加している。しかし、共同企業体の結成には、様々な企業に関しての詳細なバックボーンやバックグラウンドを把握しておく必要がある。そこで、本研究では、Web上の情報からリンク構造解析と自然言語処理の技術を利用して、企業間の関係情報を取得することで企業間連携に必要な情報を獲得し、あらゆる角度からの共同企業体結成時の意思決定を支援することを目指す。そして、本提案手法で抽出された関係情報の有用性を確認するために、特定工事共同企業体の結成実績と比較実験を行った。比較実験の結果から本提案手法の可能性を検証した。

【キーワード】共同企業体、意思決定支援、Web情報、リンク構造解析

1. はじめに

近年、土木行政の需要の増大と技術開発の進展に伴い、土木業務は複雑かつ多様化の一途を辿っている。大規模な事業に対して、単独の企業で対応することは困難であり、円滑な業務の遂行と効率的な運営には、複数の職種に対する高い技術力が必要となる。そのため、国土交通省では、特定の建設工事において高い技術力が必要な場合、共同企業体（JV：Joint Venture）の結成を義務付けている¹⁾。共同企業体とは、大規模な事業において、単独企業では資金力、技術力や労働力を補うことが困難な場合や大規模な事業においてリスクを複数の企業で分散する必要がある場合に、複数の企業が共同で事業に参画することを目的として形成する事業組織体のことである。このことにより、各企業は、得意分野を活かして事業に取りかかることができる。さらに、共同で事業を実施することにより、企業間の信頼度が高い場合は新たな知識の獲得にも有効な方法となる^{2),3)}。

共同企業体には、年間を通して結成される経常共同企業体と特定の工事ごとに結成される特定工事共同企業体がある。経常共同企業体は、優良な中小・中

堅業者が継続的な協業関係を確保することにより、その経営力と施工力を強化することを目的として結成される。一方、特定工事共同企業体は、経験や融資力の増大および危険の分散を図ることを目的としている。さらに、特定工事共同企業体では、地方業者の育成や入札条件の対応のために結成されることがあり、対象工事に応じて柔軟に共同企業体が組み立てられることが多くある。

宮崎県では、共同企業体を結成する場合の留意点や手順を示したマニュアル⁴⁾を公開しており、企業間の連携を支援する活動を行っている。この事例が示すように、公的機関が企業間の連携を推進しており、企業間連携の重要性が高まっている。そうした背景の中、各企業は、自社の専門分野を活かして短期間に安全かつ確実な事業を実施するために企業間連携を検討している。企業間の連携には、企業合併、技術提携や共同事業など様々な形態がある。企業合併としては、連携企業会社全てが解散し新規会社を設立する新設合併や連携対象企業を取り込む形での吸収合併がある。また、技術提携は、企業の専門分野の強化と不得意分野の補完のために行う。そして、

*1 総合情報学研究科 06-6396-3915, nakamura@kansai-labo.co.jp

*2 総合情報学部 072-690-2153

共同事業は、1つの目的を達成するために複数の企業が参加する提携である。これらの提携はそれぞれ異なった組織形態を示しているが、適切な企業間の提携を行うためには、以下に示す工程を明確化する必要がある。

- 1) 提携対象企業の模索
- 2) 対象企業についての調査
- 3) 企業の共通認識の確認
- 4) 提携形態の構想模索
- 5) 提携による成果と課題の確認
- 6) 提携に向けた具体的な協議

企業間の提携では、まず、提携対象となる企業の候補を探し出し、その候補企業に関して業績や経営状況、得意分野や関連企業等を詳細に調査する。次に、提携対象企業の決定後、企業間提携における目的や目標などの共通認識を確認し、提携体制の方針を決定する。そして、提携における成果目標を設定し、提携をする上で障壁となる内容を確認する。最後に、具体的な提携の協議を行い、企業間提携を実現する。この中でも、提携対象企業についての調査は、効果の高い企業間提携を実現する上で重要な工程であるが、対象企業についての詳細を把握することは困難である。そこで、企業分析を支援する方法として、既研究では、バランススコアカードを利用して企業の業績を分析する手法⁵⁾や企業活動が円滑に行われているかをコミュニケーションデータから評価する手法⁶⁾などが考案されている。しかし、これらの研究はいずれも、対象企業に関して詳細な情報が必要となる。さらに、複数の企業に同様の調査を行うことは困難であるため、対象となる企業候補を予め取捨選択する必要がある。しかし、取捨選択する際に、主観的な判断が入ってしまうという問題がある。そのため、企業の背景情報を抽出するためにインターネットに公開されている情報に着目し、各 Web ページ間の関係情報を解析し可視化する研究⁷⁾が行われている。しかし、これらの研究は Web 上のリンク構造に着目してコミュニティを形成するため、Web コミュニティ内に不要なサイトが含まれるという問題がある。

そこで、著者らは既研究⁸⁾において、Web 情報か

ら組織間の関係を抽出する手法を提案してきた。既研究は、Web 上のリンク構造に加えて、自然言語処理の技術を利用して関係情報の補完を行うことを目指したものである。この研究により、Web 上のリンク関係にのみ着目した場合の問題点を解消し、より精練された Web コミュニティの生成を実現した。ただし、既研究では Web コミュニティ生成の探索基点となる Web サイト群を取得し、それを初期集団としている。しかし、その初期集団は、検索エンジン Google において、キーワード「co.jp」で検索した結果の上位 100 件の Web サイトとしているため、適用分野に応じたコミュニティ情報の収集が困難なことや精度の問題から有用な初期集団が取得できないという問題がある。そのため、共同企業体の結成に有用となる企業間の関係情報を収集するという本研究の目的に対して既研究の手法を適用することは困難である。

そこで、本研究では、Web 上の情報を任意のカテゴリに自動的に分類することにより、共同企業体の結成に有用となる企業間の関係情報を分野毎に抽出する。そして、リンク構造解析を利用して分野毎に主要な企業の情報を取得する。このことにより、提携対象企業の模索時の問題点を解決し、入札条件の対応や技術の拡充を目的とする特定工事共同企業体結成時の意思決定を支援することを目指す。

2. 研究の概要

Web の構造を解析し関係情報を抽出する研究には、ページ間のリンク関係を中心とした研究⁹⁾や自然言語処理を利用した研究¹⁰⁾がなされている。この中で、リンク構造の解析は、処理負荷が小さいために、Web ページを大まかなカテゴリに分類するといった特徴がある。そのため、大規模な Web コミュニティの解析¹¹⁾に適している。しかし、建設業界は IT 化が遅れており、企業のホームページの約半数が、自社の情報発信を目的として公開¹²⁾しているのみである。そのため、企業のホームページ間のリンク構造のみに着目した従来手法では正確な関係を抽出できないという問題がある。また、自然言語処理は、詳細な解析により、比較的高い精度の情報抽出が可能という点から小規模な Web コミュニティの解析に適して

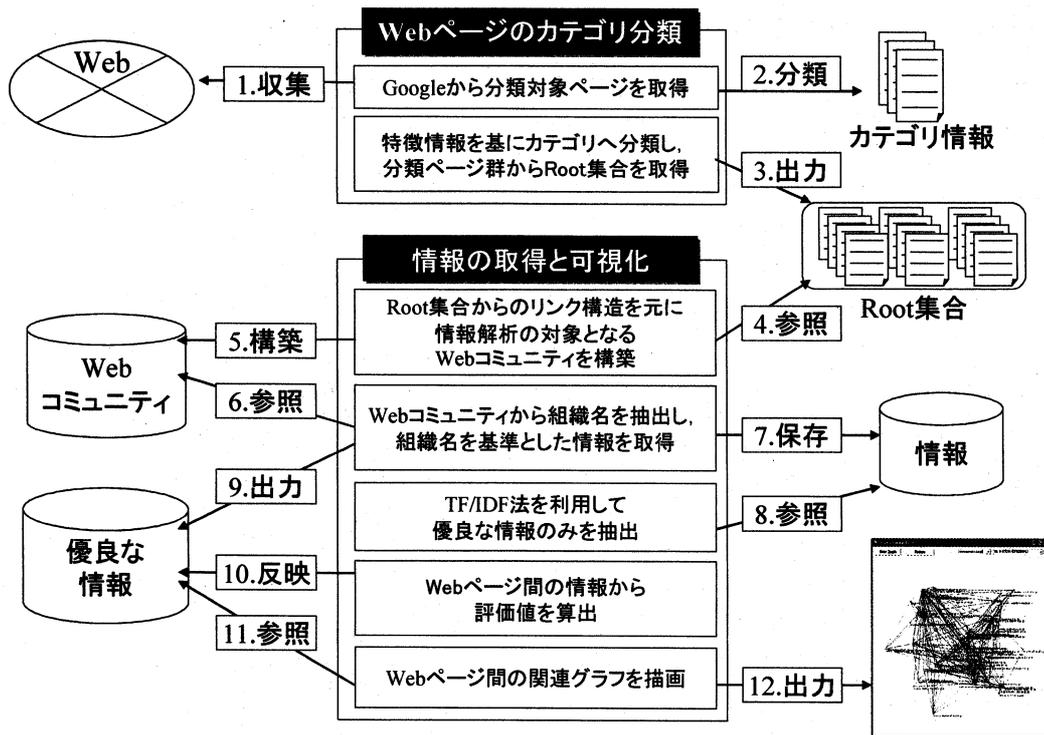


図-1 処理の流れ

いる。しかし、本研究では、関係情報の抽出において、企業間の関係を幅広く取得する必要があるため、本問題にそのまま適用することは困難である。そこで、本研究では、これらの2つの方式を組み合わせることで、互いの問題を補完し、処理負荷が小さく、精度の高い情報解析手法を提案する。

Web コミュニティは、一般的に初期集団である Web ページ群からの関係情報に基づいて生成されるため、Web コミュニティの内容は、初期集団に含まれる Web ページに依存するという特徴がある。既研究⁹⁾⁻¹²⁾では、初期集団の構築に、任意のキーワードを検索エンジンで検索した際に得られる Web サイトの URL 群を利用する手法¹³⁾が多く提案されている。しかし、検索エンジンの検索結果には、過剰 SEO (Search Engine Optimization) が施された Web ページが含まれており、対象のキーワードとは無関係の Web ページが取得されるため、Web コミュニティの精度が低下するという問題がある。そこで、この問題を解消するために本研究では、検索エンジンから取得した結果の Web ページをカテゴリに分類することにより root 集合の品質を高める。そして、カテゴリ分類された Web ページ群をリンク構造解析に適用することで、任意のトピックにおける企業間の関係

情報の信頼性を確保する。

以上の問題点を踏まえた上で、Web ページから任意のトピックに基づく企業間の関係情報を抽出するために、本提案手法では図-1に示す3つの処理の流れで関係情報の取得を行う。Web ページのカテゴリ分類では、まず、インターネット上から企業の Web ページ集合を取得する。次に、取得した企業のカテゴリ分類を行う。カテゴリ分類においては、SVM (Support Vector Machine) を用いる手法が提案されているが、SVM への入力値の決定が困難であることや入力値の種類選択によっては良い精度を得られないことなどの問題¹⁴⁾がある。また、SVM では、対象となる問題に合わせて適切なパラメータを指定する必要があり、新たなトピックを入力するたびに学習を行う必要があるため、本研究で採用することは困難である。そのため、本研究では、ベクトル空間モデルを用いて、企業の Web ページ間の類似度からカテゴリへ分類する手法を採用する。

カテゴリの分類体系には、一般的に UDC (Universal Decimal Classification) や NDC (Nippon Decimal Classification) などの学術論文を対象とした分類体系とインターネット上の各ポータルサイトが提供するカテゴリ情報などの Web ページを分類する

ための体系がある。しかし、これらの体系は、全ての Web ページを汎用的に分類できるような形式であるため、特定のトピックに特化したカテゴリ分けではない。そのため、本研究では、各市区町村が提供している格付工種の中で、ポータルサイトのカテゴリ分類と一致するものを分類対象カテゴリとして採用する。

関係情報の取得では、まず、各カテゴリに分類された Web ページ群を基準として、リンク構造解析に必要な Base 集合を形成する。Base 集合とは、リンク構造解析の基点となる Web ページの集合を指す。次に、Base 集合内から組織名を抽出し、抽出した組織名をキーとして Web ページ間の関係情報を確保する。この時、抽出した組織名の重要度を個別に計算することで関係の強度を判断し、弱い関係情報を削除することで有用な関係情報のみを追加の関係情報として抽出する。そして、抽出された有用な関係情報から HITS (Hyperlink-Induced Topic Search) アルゴリズム¹⁵⁾の Authority と Hub の考え方を利用して、各関係の評価値を算出する。

関係評価算出とグラフの可視化では、HITS アルゴリズムによって算出された関係情報の評価値を元にして、関係情報及び Web ページをグラフとして可視化する。

3. Web ページのカテゴリ分類

Web ページのカテゴリ分類は、カテゴリ毎に用意されたカテゴリの特徴を示す文書と新たに追加する Web ページとを比較し、類似性を算出することで行う。カテゴリの特徴を示す文書は、ポータルサイトの採用カテゴリに分類されている Web ページを基点として、その Web ページからリンクする同一ドメインの全ての Web ページから抽出する。類似性の算出では、それぞれの文書に対して TF/IDF (Term Frequency-Inverse Document Frequency) 法¹⁶⁾を利用して特徴ベクトルを生成する。TF/IDF 法とは、1つの文書内における任意の単語の出現頻度を示す TF 値と全文書において任意の単語が含まれる文書数の調整値である IDF 値を利用して、全文書における単語の重要度 TF/IDF 値を算出する手法である。同様に、各カテゴリについても特徴ベクトルを生成する。

そして、生成した特徴ベクトルを多次元空間に展開し、文書とカテゴリとのコサイン相関値を算出する。そして、コサイン相関値をカテゴリ毎に算出し、その値が最大となるカテゴリを分類カテゴリとして一意に判定する。また、カテゴリ分類の既研究¹⁷⁾では、1つの Web ページが複数のカテゴリに属するなどの多重トピック問題があげられる。しかし、企業が複数のカテゴリに分類された場合、その企業から同様の関連情報が取得され、各カテゴリでの関連が同様の傾向を示す可能性がある。そのため、本研究では、Web ページをカテゴリに対して一意に分類することを目的とする。

(1) 特徴ベクトルの生成

特徴ベクトルは、対象となる Web ページおよびカテゴリ毎に特徴を示す単語を抽出し、その単語を元に生成する。特徴ベクトルの生成は、まず、対象の Web ページに対して形態素解析を行い、Web ページ内に含まれる名詞群を抽出する。本提案手法では、茶筌¹⁸⁾を利用して形態素解析を行う。ここで、Web ページの全体集合を DA、Web ページ x から抽出した名詞を Nx 、名詞群を $Kx=\{Nx_1, Nx_2, Nx_3, \dots, Nx_m\}$ 、Web ページ x における名詞 Nx_k の数を $C(Nx_k)$ とする。また、カテゴリ y に含まれる全ての文書から抽出した名詞を Ny 、名詞群を $Hy=\{Ny_1, Ny_2, Ny_3, \dots, Ny_n\}$ 、カテゴリ y における名詞 Ny_n の数を $C(Ny_n)$ とする。ここで、各カテゴリの Web ページ毎に算出された名詞と Web ページ x に含まれる名詞の総和を $KH(x)$ とする。次に、抽出した名詞に対して Web ページにおける重要度を算出する。本研究では、重要度算出方式として一般的に利用される TF/IDF 法を採用する。TF/IDF 法による、重要度算出は、次式の通りである。

$$TF(Nx_k) = \frac{C(Nx_k)}{\sum_{i=1}^m C(Nx_i)} \quad (1)$$

$$IDF(Nx_k) = \log \frac{|DA|}{df(Nx_k)} \quad (2)$$

$$W(Nx_k) = TF(Nx_k) \times IDF(Nx_k) \quad (3)$$

式(2)における $df(Nx_k)$ は、Web ページ集合 DA において、名詞 Nx_k を含む Web ページの数を指すものとする。最後に、Web ページ x における重要度 $W(Nx_k)$ を特徴ベクトルとして、 $Vk_x=\{W(Nx_1), W(Nx_2), \dots, W(Nx_m)\}$

$W(Nx_3), \dots, W(Nx_m)$ と定義する。また、カテゴリ y の名詞群 H_y に対して同様の重要度算出処理を行い、カテゴリ y における重要度 $W(Ny_n)$ を特徴ベクトルとして、 $V_{h_y} = \{W(Ny_1), W(Ny_2), W(Ny_3), \dots, W(Ny_n)\}$ と定義する。

(2) Web ページとカテゴリとの類似度の判定

Web ページと各カテゴリの類似度を判定するために、Web ページとカテゴリの組み合わせを $D(x,y)$ 次元の仮想空間に展開する。ここで利用する $D(x,y)$ の数は、Web ページ K_x とカテゴリ H_y を比較した場合、式 (4) から算出される。

$$D(x,y) = |K_x \cup H_y| \quad (4)$$

このため、仮想空間に展開された特徴ベクトルは、 $D(x,y)$ 個の要素を持つ必要がある。しかし、特徴ベクトル V_{k_x}, V_{h_y} は、 $D(x,y)$ より少ない次元数しか保持しない。そこで、本研究では次元数を同一にするために、Web ページの特徴ベクトルの値が存在しない特徴要素を 0 として補完する。次に、Web ページとカテゴリの類似性を算出するために特徴ベクトルからコサイン相関値を算出する。コサイン相関値の算出を式 (5) に示す。

$$\text{same}(V_{k_x}, V_{h_y}) = \frac{\sum_{k=1}^n (W(Nx_k) \times W(Ny_k))}{\sqrt{\sum_{k=1}^n W(Nx_k)^2} \times \sqrt{\sum_{k=1}^n W(Ny_k)^2}} \quad (5)$$

ここで、 n は特徴ベクトルの要素の総数、 $W(Nx)$ は Web ページ x に含まれる名詞の TF/IDF 値、 $W(Ny)$ はカテゴリの文書に含まれる名詞の TF/IDF 値を指す。

式 (5) において、分子は Web ページ x とカテゴリ y の特徴ベクトルの内積を表し、分母は各特徴ベクトルの原点との距離の積である。コサイン相関値は、Web ページの特徴ベクトル同士の比較を角度として算出するもので、0 から 1 の範囲で値が得られる。この値が 1 に近いほど Web ページとカテゴリの分類が近いということが言える。そして、全てのカテゴリの組み合わせにおいてコサイン相関値の算出が完了すると、得られたコサイン相関値が最大のも

のを分類カテゴリと判定する。そして、各カテゴリに分類された Web ページの集合をリンク構造の解析の基準となる Root 集合 $R(x_1)$ とする。

4. 関係情報の取得と可視化

関係情報の取得では、リンク構造解析と自然言語処理による関係情報の抽出を組み合わせることにより、品質の高い関係情報を抽出する。まず、リンク構造解析を利用して、Web コミュニティの可能性のある Web ページ群を抽出する。Web コミュニティの収集には、リンク構造解析の既存手法である HITS アルゴリズム¹⁵⁾を利用する。HITS アルゴリズムでは、各 Web ページの有用性を評価するための基準として、Authority Score と Hub Score という 2 種類の値を利用している。Authority Score とは、特定のトピックに関する情報の豊富さを示す値であり、Hub Score とは、Authority Score の高い Web ページへのリンクの豊富さを示す値である。このことから、Authority Score が高い Web ページへリンクをしているページは Hub としての価値が高く、また、Hub Score が高い Web ページからリンクされているページは Authority として価値が高いという仕組みになっている。

次に、抽出された Web コミュニティ内のページ間の関係を取得するために、Web ページ間の関係情報を抽出する。この場合、本研究では Web ページ間のリンク構造を利用せずに組織間の関係情報を抽出する手法を採用する。そして、取得された関係情報に対してリンク構造解析を行い、主要な Web コミュニティを抽出する。

最後に、抽出した Web コミュニティ内での関係情報を表現するために関係グラフの可視化を行う。

(1) リンク構造解析による Web コミュニティの取得

リンク構造解析による Web コミュニティの取得では、関係情報の取得対象となる初期 Base 集合を作成する。本研究で採用する初期 Base 集合 $S(x)$ を式 (6) に示す。

$$S(x) = R(x_1) \cup B(x_2) \quad (6)$$

式 (6) において、Backward 集合 $B(x_2)$ は、Root 集合 $R(x_1)$ に含まれる Web ページに対してリンクを張っ

ているページの集合を指す。Root 集合に関連する 2 つの集合の概念を図-2 に示す。一般的な HITS アルゴリズムの場合は、Root 集合 $R(x_i)$ からリンクしているページの集合である Forward 集合を $S(x)$ に加えたものを利用する。しかし、Root 集合 $R(x_i)$ に無関係なリンク集やポータルサイトなどのページが含まれていた場合に、Forward 集合でのトピックが発散し、本来のキーワードに従った Web コミュニティが抽出できない Topic Drift という問題¹⁹⁾が生じる。そのため、本研究では Root 集合 $R(x_i)$ と Backward 集合 $B(x_j)$ を合わせた Base 集合 $S(x)$ を利用して Web コミュニティを解析対象として採用する。

(2) 組織関係の抽出

組織関係の抽出では、Web ページに含まれる組織名に着目し関係情報を抽出する。そして、抽出された関係情報に基づいてリンク構造解析を行い、関係評価値を算出する。以下に、抽出された関係評価値に基づいて、関係情報を可視化する手法について説明する。

a) Web ページの関係情報の取得

Web ページの関係情報の取得では、リンク構造に着目して関係情報を抽出すると、Web コミュニティの情報が発散するという Topic Drift 問題¹⁹⁾が発生する。そのため、本研究では、リンク構造を利用せずに Web ページ間の関係を取得する手法として、Root 集合及び Backward 集合から組織名を抽出し、各 Web ページのタイトルとマッチングを行う手法を提案することで新たな関係情報を抽出する。組織間の関係情報の抽出には、組織名の抽出、組織名間の関連付けと優良な関係情報の抽出の 3 つの手順で行う。

組織名の抽出処理では、各 Web ページ間を関連付けるためのキーとして組織名の抽出を行う。組織名の抽出には、まず、Base 集合 $S(x)$ の各 Web ページから HTML タグを削除し、文章、タイトルとリンク文字列を抽出する。次に、抽出された文章に対して形態素解析を適用し文中の名詞を抽出する。そして、抽出された名詞と組織名辞書との比較を行い、一致した名詞を組織名として抽出する。本提案手法の組織名の抽出においては、形態素解析器として、カテゴリ分類と同様に茶筌¹⁸⁾を利用した。組織名の抽出前に形態素解析を行い、辞書とのマッチング対象を名詞のみに制限することにより、名詞以外の単語が

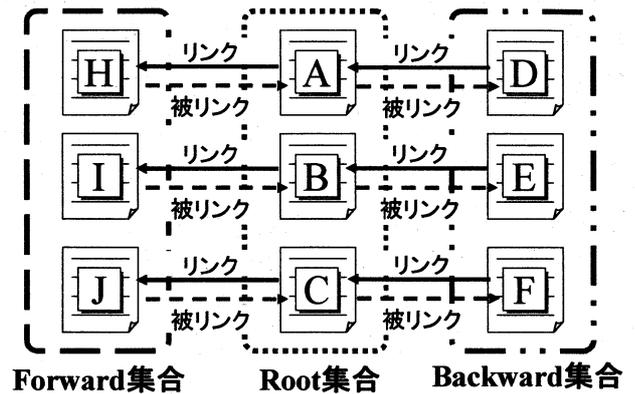


図-2 Root 集合とその他の集合の関係

組織名と誤判定される問題を回避する。また、組織名の抽出で利用する組織名辞書は、茶筌に含まれる辞書を利用した。組織名辞書に収録されている語数は、16,610 件である。そして、抽出した組織名を $TI(xy_n)$ 、その個数を $C(xy_n)$ とする。この時、 y は抽出した組織名を示す。

組織名間の関連付け処理では、組織名の抽出によって取得した組織名をキーとして、Web ページ間の関連付けを行う。まず、組織名と関連付ける Web ページの候補集合として、新たに組織の Web ページ集合を取得し、この集合と $S(x)$ の和を $S_{add}(x)$ とする。次に、 $S_{add}(x)$ の各 Web ページのタイトルに対して形態素解析を行い、抽出した名詞を組織名辞書と比較し、一致した名詞を組織名として抽出する。ここで、抽出した組織名の集合を $T2(xy_m)$ とする。そして、 $TI(xy_n)$ と $T2(xy_m)$ のマッチングを行い、一致したものを関係情報 $RI_{ni}(xy_i)$ とする。

優良な関係情報の抽出処理では、組織名間の関連付け処理において抽出された関係情報 $RI_{ni}(xy_i)$ に含まれる優良な関係情報のみを抽出するために、重要度に基づき関係情報の切捨てを行う。各組織名の重要度算出は、カテゴリ分類時と同様に TF/IDF 法¹⁶⁾を採用した。式 (1) において、Web ページ x における任意の組織名 y_n を Web ページ内部の組織名の総数で割ることで、その Web ページ内部において任意の組織名の重み係数を算出する。そして、式 (2) において対数計算を行い、任意の単語について $S_{add}(x)$ の全体個数と任意の組織名が含まれる個数から、組織名の重みを調整する。ここで、 df は、任意の組織名が含まれる Web ページの合計個数を示す。この処理

により、知名度の高い組織は、組織名 1 件あたりの重みが低下する。最後に、式 (3) において、重み係数の積によって、抽出を行った組織名の Web ページにおける重要度を算出する。

各組織名の重み係数に基づき、閾値による切り捨てを行い、優良な関係情報を抽出する。各組織間の関係情報の切り捨てを行うために閾値を算出する。閾値の算出方法としては、まず、集合 $S_{add}(x)$ から任意の個数の Web ページを抽出し、手作業で組織間の関係を調査し正解データを作成する。そして、正解データと本提案手法で取得した関係情報を比較することで F 値を算出し、これを評価値として最急降下法²⁰⁾により最適な閾値を算出する。 F 値とは、情報検索や特定語句の抽出の精度評価に利用される指標で、抽出データと正解データから算出する。抽出データに含まれる正解データの割合を適合率、全正解データの中から抽出できた正解データの割合を再現率として、これらの調和平均が F 値となる。また、閾値の算出時に局所解を回避するために、初期値をランダム抽出し、複数回の試行を行う。これにより、弱い関係情報を排除し、関係情報の品質を向上させる。

b) 関係評価値の算出

組織間の関係情報の評価値を算出するために、従来のリンク構造解析の評価手法である HITS アルゴリズム¹⁵⁾を採用する。

本研究においては評価値を算出するために、まず、 $S(x)$ のリンク構造から Backward 集合の関係情報を抽出することで、関係情報の集合 $RI_{back}(x)$ を取得する。次に、 $RI_{back}(x)$ と $RI_{in}(xy_i)$ の和集合 $RI_{all}(xy_i)$ を関係情報の全体集合として定義する。そして、 $S_{add}(x)$ と $RI_{all}(xy_i)$ に対して、HITS アルゴリズムを適用し、Authority Score と Hub Score を算出する。最後に、これらの処理から得られた情報を元に、組織の関係グラフ $G(x)$ を作成する。

(3) 関係情報の可視化

本研究で取得した関係情報に基づいて関係グラフの可視化を行う。関係グラフによる可視化に関しては、既研究^{21), 22)}の手法を参照した。本研究では、Authority Score および Hub Score の算出後に、 $S_{add}(x)$ から Authority Score が 0 に収束したものを削除し、残った Web ページの集合を $S_{auth}(x)$ とする。また、削除された Web ページに関連する関係情報も合わせて

表-1 F 値によるカテゴリ分類の評価

カテゴリ名	適合率	再現率	F 値
土木	0.9474	0.4222	0.5841
建築	0.6842	0.9286	0.7879
管	0.4737	0.8182	0.6000
塗装	0.4211	0.6667	0.5161
舗装	0.7895	0.6818	0.7317
造園	0.6842	0.8667	0.7647
電気	0.4737	0.6429	0.5455
上下水	0.7895	0.7895	0.7895
平均	0.6579	0.7270	0.6649

$RI_{all}(xy_i)$ から削除する。そして、 $S_{auth}(x)$ を頂点、 $RI_{all}(xy_i)$ を辺として関係グラフ $G(x)$ を作成する。関係グラフ $G(x)$ の描画では、 $S_{auth}(x)$ の頂点を格子状に等間隔で配置し、各頂点において、関係情報にて結び付く全ての頂点との重心計算を行うことで位置の補正を行う。その後、関係情報に従って頂点間に線分を描画する。

5. 評価実験

(1) Web ページのカテゴリ分類の精度評価

Web ページのカテゴリ分類の精度評価を行うために、Web ページ群を収集し、各カテゴリへそれらを分類する。本実験で採用する分類先のカテゴリは、各市区町村が提供している格付工種とインターネットのポータルサイト「Yahoo! Japan」のカテゴリ情報とで重複している土木、建築、管、塗装、舗装、造園、電気と上下水道の 8 つのカテゴリとした。また、それぞれのカテゴリについての学習データとして「Yahoo! Japan」の該当カテゴリ内の企業ホームページの文書情報を採用した。本実験で利用する分類対象の Web ページ集合は、Google から「co.jp」と各カテゴリ名称を組み合わせたキーワードで検索した結果を採用した。その結果、土木カテゴリに 360 件、建築カテゴリに 112 件、管カテゴリに 88 件、塗装カテゴリに 96 件、舗装カテゴリに 176 件、造園カテゴリに 120 件、電気カテゴリに 112 件、上下水カテゴリに 152 件分類された。カテゴリ分類の評価は、各

表-2 F値による関係情報の品質評価

項目	適合率	再現率	F値
閾値切捨てなし	0.3469	0.4275	0.3830
閾値切捨てあり	0.6190	0.5909	0.6046

表-3 Webコミュニティ上位100件の内訳

項目	内訳
組織（土木業界）	91
組織（その他業界）	7
リンク集	0
その他	1

表-4 三井住友建設の主要な関係情報

企業名	TF/IDF値	施工例
日特建設	0.0245	<ul style="list-style-type: none"> ・グランドメゾン玉川上水センタースクエア新築工事（熊谷・三井住友） ・カメラアタワー川口（清水・西武・三井住友・埼玉建興・川口土建） ・東京国際空港第2旅客ターミナル増築工事（清水・大林・東急・三井住友） ・高島トンネル（前田・三井住友・勝村・京急） ・九州横断自動車道日見橋（三井住友・銭高） ・森吉山ダム第二工事（西松・三井住友・銭高） ・新宿七丁目駅二工区土木（大成・三井住友・西武） ・箕面有料道路山岳トンネル築造工事北工区（鹿島・大成・東亜・三井住友・青木あすなろ） ・日本橋1丁目ビルディング（清水・三井住友・東急）
熊谷組	0.0212	
清水建設	0.0193	
清水総合開発	0.0193	
前田建設工業	0.0149	
銭高組	0.0129	
西武建設	0.0127	
日立セメント	0.0120	
東急建設	0.0119	
東亜建設工業	0.0112	

カテゴリへ分類された結果を目視で確認し、それぞれのカテゴリ毎にF値を算出することで評価した。各カテゴリのF値を表-1に示す。提案手法によるカテゴリ分類の評価は、F値の平均が0.6649となっており、カテゴリ間で分類精度に偏りが見られる結果となった。これは、「Yahoo! Japan」のカテゴリ情報から収集した学習情報にカテゴリの特徴を示す単語が含まれていなかったことに起因すると考えられる。また、土木のカテゴリに関しては、カテゴリの特徴を示すキーワードが他のカテゴリのキーワードと多く重複する傾向が見られた。これは、土木カテゴリ内に舗装や配管などのカテゴリが存在し、他のカテゴリの特徴を包括的に含んでいることに起因すると考えられる。このことにより、他のカテゴリと比較して抽出数が3倍になっており、精度が低下していると考えられる。

(2) 関係情報の品質評価

本節では、関係情報を評価するために、各カテゴリの組織のWebページを基点として、Webページ群

を収集し、それぞれのWebページからの関係の品質を評価した。本実験では、本手法で取得した関係情報に対してF値を算出することで精度を評価する。

実験で利用するWebページの集合として、Webページのカテゴリ分類の結果で得られた組織をRoot集合とした。また、本実験では、市区町村の格付工事の中でも有資格者の企業数が多い土木カテゴリに着目して実施した。実験の結果、本提案手法においては、Root集合360件とそのRoot集合を基点として収集したBackward集合1,092件を組み合わせた1,452件をリンク構造解析の候補集合とした。そして、収集したWebページ群に対して、リンク構造解析を実施した結果、835件の関係情報を抽出することができた。

これらの結果で得られた関係情報の品質を評価するために、Root集合から抽出した正解の関係情報1,029件と比較を行った。関係情報の評価結果を表-2に示す。関係情報を評価した結果、F値で0.3830の値を取得した。また、本提案手法では、優良な関

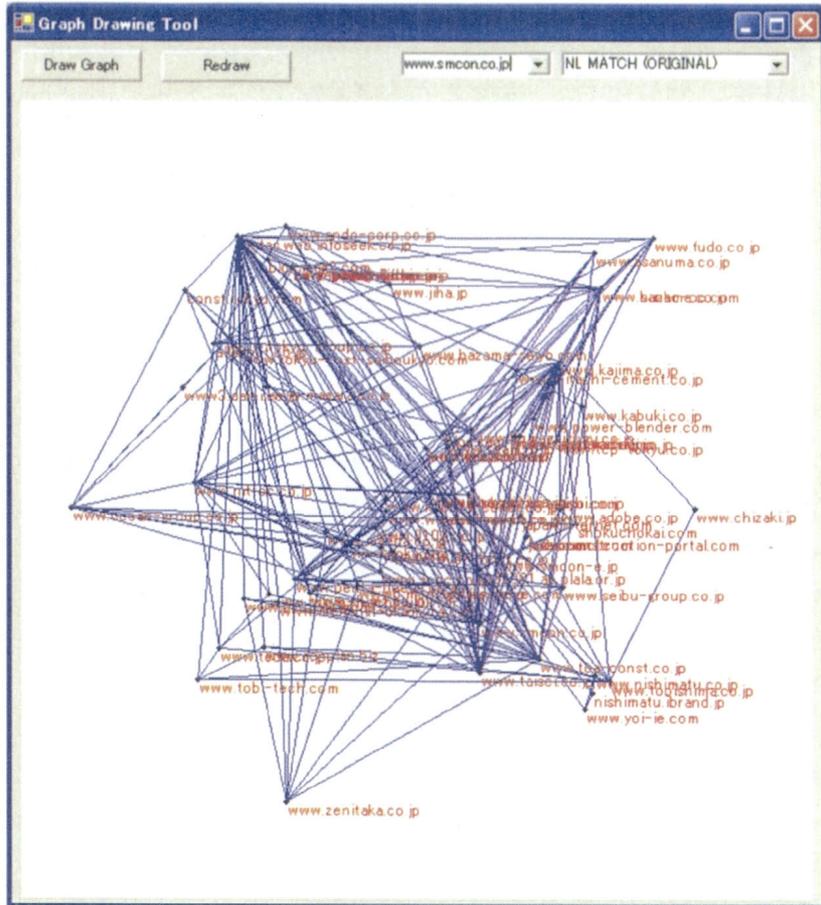


図-3 自然言語処理による関係情報の可視化

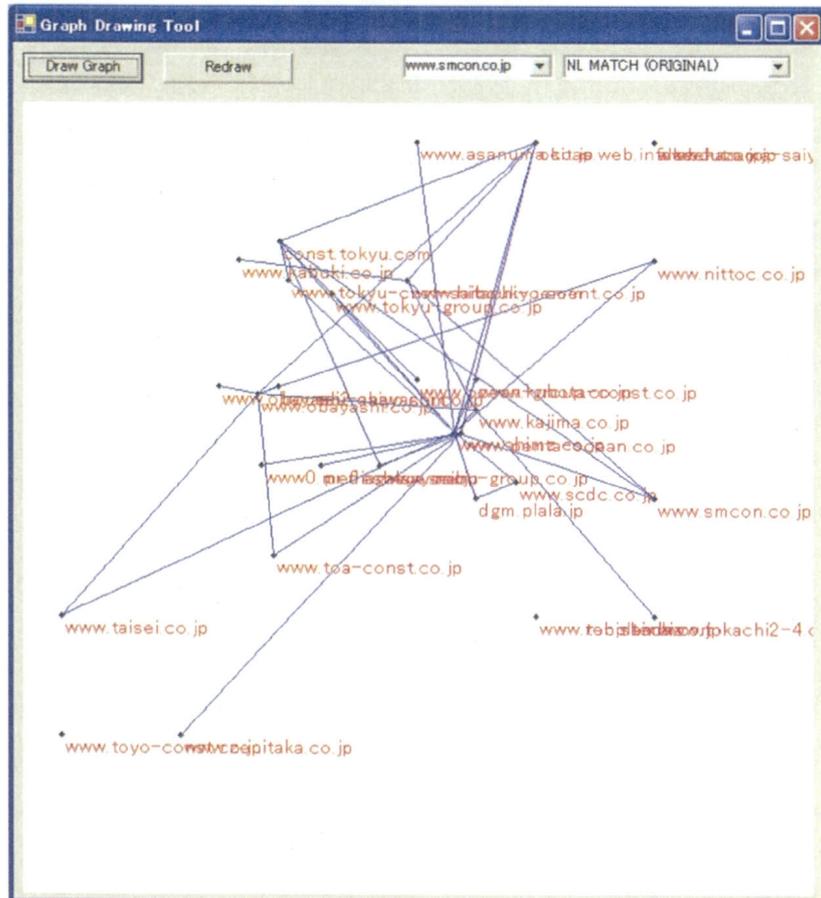


図-4 www.smcon.co.jp の主要な関係情報の可視化

係のみを抽出するために 0.0314 を閾値とした。この閾値は、 F 値を基準として初期値のランダム取得と最急降下法²⁰⁾により算出した。閾値にて切り捨てた結果、 F 値において 0.6046 の値を取得することができた。 F 値が 0.6046 にとどまった原因としては、建設業界で Web ページを公開している企業の数が他の業界と比べて少ないことが原因と考えられる。

(3) Web コミュニティの評価

本実験では、本提案手法により取得した関係情報により構成される Web コミュニティが、土木のトピックに基づいて抽出されているかを確認する。しかし、任意のトピックに最も関連性の高い Web コミュニティを抽出する場合には、各 Web ページ間の関連性を加味して主要な Web ページのみを抽出する必要がある。そのため、本実験では、HITS アルゴリズムにより算出される Authority Score を用いて、抽出された Web コミュニティの評価を行う。本提案手法により抽出された Web コミュニティの上位 100 件の結果を表-3に示す。

本提案手法で抽出された Web コミュニティは、土木業界に関するトピックを多く含む良質な Web コミュニティであることがわかった。さらに、Web コミュニティ内の上位の企業を確認したところ、過去に共同企業体を結成した経験のある企業が多く含まれており、土木業界における主要なコミュニティを抽出できたと考えられる。

(4) 組織関係グラフの可視化

本実験では、本提案手法により取得した関係情報から構成する Web コミュニティを確認するために、グラフによる可視化を行う。ここでは、www.smcon.co.jp (三井住友建設) についての組織関係をグラフ化することを試みる。具体的な可視化方式については、基点となる組織の Web ページから、その組織を中心とした Web コミュニティを抽出する。また、Web コミュニティの形成では、Authority Score が 0 に収束した Web ページを排除し、主要な関連のみから構築する。そして、Web ページを点としてランダムに配置し、これらの Web ページ集合に関する関係情報のみを線分として描画した。また、各点において重心計算による点の再配置を行った。本提案手法の結果について、グラフの可視化を行った結果

を図-3に示す。また、主要な関係情報のみを抽出した結果を図-4と表-4に示す。

本結果を見ると、www.smcon.co.jp から直接の関係として、www.shimz.co.jp (清水建設)、const.tokyu.com (東急建設) へ繋がっており、1つの組織を介して、www.obayashi.co.jp (大林組) へ繋がっている。これらの企業は、清水・大林・東急・三井住友特定建設工事共同企業体として実際に活動しており、正確な企業間の関係が抽出できていることがわかる。さらに、表-4の結果からwww.smcon.co.jp の Web ページを基点として、幅広い他企業の Web サイトへ関係しており、既に共同企業体結成実績のある企業に加えて、新たな共同企業体への参加候補企業として www.nittoc.co.jp (日特建設) や www.hitachi-cement.co.jp (日立セメント) が抽出できていることがわかる。このことから、本提案手法で抽出した Web コミュニティは、対象のカテゴリにおいて有用な関係を抽出しており、共同企業体結成時の1つの判断材料を提供できたと考えられる。

(5) 関係情報の個別評価

本提案手法により取得した関係情報が正しい関係を表しているかどうかを確認するために、企業に対する関係情報を個別に検討する。個別評価では、大手建設会社であるスーパーゼネコン5社 (大成建設、大林組、鹿島建設、清水建設と竹中工務店) を対象とする。本実験で取得したスーパーゼネコンと関連する企業の一覧を表-5に示す。

本実験の結果から、各スーパーゼネコンにおいて、主要な関係として、実際の施工例に基づいて関連企業が抽出されていることがわかる。各スーパーゼネコンの関係企業を分析すると、実際の施工実績以外の次に示すような様々な関係が獲得できていることがわかった。

- 大成建設と関連する TC プロパティーズは、東急建設の会社分割により設立された企業であり、東急建設との施工例の関連から抽出されたと考えられる。
- 大林組と日特建設の関係は、CFRP ラミネート工法の共同開発などの情報から抽出されていると考えられる。

表-5 スーパーゼネコンとの関連情報（上位5件）

企業名	関連企業名	TF/IDF 値	施工例
大成建設	鹿島建設	0.2084	・中部国際空港旅客ターミナルビル（大成・鹿島・大林・東急・戸田・ベクテル・佐藤・矢作） ・JR セントラルタワーズ（大成・鹿島・大林・清水・熊谷・竹中・鉄建・名工・ジェイアール東海建設・新生テクノス・東海交通機械）
	竹中工務店	0.2365	
	TC プロパティーズ	0.0921	
	清水建設	0.0639	
	大林組	0.0625	
大林組	竹中工務店	0.8610	・シーサイドホテル舞子ビラ神戸（大林・竹中・長谷工・三井・日本国土・新井・舞子ビラ） ・名古屋第二赤十字病院（大林・鹿島・銭高） ・ホテル日航ベイサイド大阪（大林・大成） ・横浜ベイブリッジ下部工事（鹿島・大林・東亜）
	鹿島建設	0.0769	
	大成建設	0.0625	
	東亜建設工業	0.0581	
	日特建設	0.0531	
鹿島建設	鹿島組	2.8419	・松坂屋新南館（竹中・鹿島） ・名駅四丁目7番地区再開発ビル新築工事（竹中・大林・鹿島・清水） ・箕面有料道路山岳トンネル築造工事北工区（鹿島・大成・東亜・三井住友・青木あすなる） ・名古屋第二赤十字病院（大林・鹿島・銭高）
	竹中工務店	1.1428	
	清水建設	0.2365	
	大成建設	0.2084	
	大林組	0.0769	
清水建設	清水総合開発	2.8005	・千葉東テクノグリーンパーク（清水・東急・鹿島） ・山王パークタワー（清水・鹿島・大林・東急） ・渋谷マークシティ（東急・鹿島・大成・戸田・清水・京王） ・赤坂2丁目共同ビル（清水・竹中）
	東急建設	2.2005	
	鹿島建設	0.2611	
	竹中工務店	0.2365	
	クボタ建設	0.1319	
竹中工務店	清水建設	2.8005	・神戸国際会館（竹中・清水・新井・イチケン・岡） ・ラ・フォルテ西神中央2番館（竹中・戸田・三井・前田） ・学校法人関西学院神戸三田キャンパス（竹中・大林・熊谷・浅沼） ・愛媛県武道館（竹中・鹿島・二神）
	りんかい日産建設	1.2105	
	戸田建設	1.1810	
	大林組	1.1428	
	鹿島建設	0.8610	

- 清水建設とクボタ建設の関係は、スーパーピン工法の共同開発や適用実績の情報から抽出されていると考えられる。また、清水建設と清水総合開発の関係は、グループ会社の関係が抽出されたと考えられる。
- 竹中工務店とりんかい日産建設の関係は、竹中工務店のグループ会社である竹中土木とりんかい日産建設間の施工実績が多くあることから抽出されたと考えられる。
- 大成建設、大林組、鹿島建設、清水建設と竹中工務店のスーパーゼネコンは、それぞれの企業間で相互に関係していることがわかる。これらの相互

関係は、共同企業体における技術の補完の関係のみではなく、再生建材販売会社であるグリーン建材社や建設資材のマーケットプレイス事業を行うコンストラクション・イーシー・ドットコム社の設立などの様々な共同事業化活動から抽出できたと考えられる。

これらの結果から、本提案手法で抽出した関係情報は、実際の企業間関係を反映しており、企業の背景情報を判断するのに役立つと考えられる。

また、鹿島建設と鹿島組の関係は、関係情報を分析すると実際の施工例および企業間の繋がりが存在しないことがわかった。この関係は、自然言語処理

を用いて関係情報を獲得しているため、組織名が同一であると認識されたことで高い TF/IDF 値が出ていると考えられる。この場合においても、関係情報の獲得時に各 Web ページにおける組織名の共起関係やそれぞれの企業の Web ページへリンクしている Web ページの一致率を考慮することで、より精練された関係が獲得できると考えられる。

6. おわりに

本研究では、Web から共同企業体結成時に重要となる、対象企業についての背景情報を抽出することに成功した。評価実験の結果から、本提案で取得できた背景情報は、過去に共同企業体を結成したことのある関係、提携関係と取引関係であることがわかった。これらの情報は、企業の実績や評価情報と組み合わせることで、対象企業の信頼度や共同事業時の成果を判別することができる。例えば、任意の企業についての過去の共同企業体の結成実績とその共同企業体を実施した事業の工事内容および評価情報を蓄積することにより、新たな共同企業体結成時において、その企業が参画した際の効果を推測することが可能となる。また、本研究では、共同企業体結成時の意思決定を支援するための情報として、建設業界に特化した組織間の関係情報を生成する実験を行ったが、同様に、組織名を専門用語へ置き換えることにより、分野毎に特化した専門用語辞書および専門用語間の関連情報の抽出にも有用であると予想される。

今後は、カテゴリの分類において多重トピック自動分類問題へ対応することにより異業種間での共同企業体の結成を支援するような展開研究を行う予定である。また、本研究では、HTML のテキストデータのみを対象としたが、それらの情報に加えて PDF データや画像データに含まれるテキストデータも加えた解析を行うことにより、従来のリンク構造解析では取得できない関係情報を収集する予定である。そして、本研究で取得した関係情報の有効活用するために、企業の評価と企業の保有する Web ページの評価を結び付け、Web 上の情報の信頼度を算出する研究に発展させる予定である。

本研究の一部は、平成 19～21 年度関西大学先端科

学技術推進機構 e-Business 研究グループ（研究課題「電子国土のための e-Business に関する研究」）から助成を受け、その成果を公表するものである。

【参考文献】

- 1) 国土交通省：平成 16 年度特定建設工事共同企業体結成を義務付けた工事とその理由，2005 年 11 月。
- 2) Chung, J. C. : The effects of knowledge attribute, alliance characteristics, and absorptive capacity on knowledge transfer performance , *R&D Management*, SSCI, Vol.34, No.3, pp.311-321, 2004.
- 3) Nakamura, A. and Nakamura, M. : Firm performance, knowledge transfer and international joint ventures, *International Journal of Technology Management*, Inderscience Publishers, Vol.27, No.8, pp.731-746, 2004.
- 4) 宮崎県：建設業者のための企業間連携マニュアル，2005 年 12 月。
- 5) 芝野治郎：バランススコアカード導入のメリットと考慮すべき留意点 ～導入研究会の作業をベースとして～，情報システムと社会環境研究会研究報告，情報処理学会，Vol.2004, No.53, pp.9-12, 2004 年 5 月。
- 6) 中村英史，水田秀行：企業をコミュニケーションから評価する，情報処理学会論文誌，情報処理学会，Vol.45, No.9, pp.950-955, 2004 年 9 月。
- 7) 福地健太郎，豊田正史，喜連川優：Web Community Browser における探索機構の実装と評価，データベースシステム研究会技術研究報告，電子情報通信学会，Vol.102, No.209, pp.79-84, 2002 年 7 月。
- 8) 池辺正典，田中成典，古田均，中村健二：Web リンク構造解析と自然言語処理による組織関係の抽出についての研究，情報処理学会論文誌，情報処理学会，Vol.47, No.6, pp.1687-1695, 2006 年 6 月。
- 9) 加藤一民，松尾啓志：Markov Cluster Algorithm を用いた Web コミュニティ群の発見手法，自然言語処理研究会研究報告，情報処理学会，

- Vol.2005, No.22, pp.87-93, 2005年3月.
- 10) 山本仁志, 太田敏澄, 石田和成, 岡田勇: リンク構造と共起関係を用いた Web 空間の視覚化, デジタルドキュメント研究会研究報告, 情報処理学会, Vol.2004, No.36, pp.95-101, 2004年3月.
 - 11) 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, 電子情報通信学会, Vol.J87-D-1, No.2, pp.256-265, 2004年2月
 - 12) 建設経済研究所: 2005年建設企業 IT 活用状況に関するアンケート調査, 2005年8月.
 - 13) 野村早恵子, 小山聡, 早水哲雄, 石田亨: WEBコミュニティ発見のための HITS アルゴリズムの分析と改善, 電子情報通信学会論文誌, 電子情報通信学会, Vol.J85-D-1, No.8, pp.741-750, 2002年8月.
 - 14) 宮部泰成, 高村大也, 奥村学: 異なる文書中の文間関係の特定, 自然言語処理研究会研究報告, 情報処理学会, Vol.2005-NL-168, No.73, pp.35-42, 2005年7月.
 - 15) Kleinberg, J. M. : Authoritative Sources in A Hyperlinked Environment, *Journal of the ACM*, ACM, Vol.46, No.5, pp.604-632, 1999.
 - 16) Salton, G. and Yang, C. : On the specification of term values in automatic indexing, *Journal of Documentation*, Emerald Group Publishing Limited, Vol.29, No.4, pp.351-372, 1973.
 - 17) 賀沢秀人, 泉谷知範, 平博順, 前田英作: 最大マージン原理にもとづく多重トピック文書の自動分類, 自然言語処理研究会研究報告, 情報処理学会, Vol.2004-NL-163, pp.53-60, 2004年9月.
 - 18) 茶 筌 : ChaSen's Wiki - FrontPage , <<http://chasen.naist.jp/hiki/ChaSen/>>
 - 19) Bharat, K. and Henzinger, M. R. : Improved algorithms for topic distillation in a hyperlinked environment, *Proceedings of the 21th annual international ACM SIGIR conference on research and development in information retrieval*, ACM Press, pp.104-111, 1998.
 - 20) 中川徹, 小柳義夫: 最小二乗法による実験データ解析, 東京大学出版会, 1982年5月.
 - 21) 田地晶, 宮寺庸造, 樋山淳雄, 横山節雄: ユーザ思考に基づく学術論文関係図の可視化手法の提案, 教育工学研究会技術研究報告, 電子情報通信学会, Vol.100, No.420, pp.37-44, 2000年11月.
 - 22) 土橋喜, 山内平行, 立花隆輝: キータームの関連性の視覚化による知識連鎖の発見支援-TermLinker システムの可視化機能-, 知能と複雑系研究会研究報告, 情報処理学会, Vol.103, No.304, pp.41-46, 2003年9月.

Research on Decision Support for Organizing Joint Venture

By Kenji NAKAMURA, Shigenori TANAKA, Hitoshi FURUTA

Recently, several companies are organizing JV (Joint Venture) to join large-scale building and construction works. Then, in order to form a JV, it is inevitable to collect various information and backbone data about those companies. However it is difficult to gain them in details. In this study, we aim to extract useful information of relationship among some companies from the Web information, using such technologies as link structure analysis and natural language processing, and to make a decision for organizing JV using this information. To validate the applicability of the proposed method, numerical simulations are performed to compare the results obtained by the proposal method with existing cases of JV. As a result, it is confirmed that the proposed method can efficiently provide available information for practical cases. At results, the method and information were available effectively.