社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

顔情報を用いた放送映像中の人物の名寄せ

小笠原　崇†　　高橋　友和†,††　　井手　一郎†,†††　　村瀬　洋†

† 名古屋大学大学院 情報科学研究科　〒 464–8603 愛知県名古屋市千種区不老町
†† 日本学術振興会
††† 国立情報学研究所　〒 101–8430 東京都千代田区一ツ橋 2–1–2
E-mail: †{toga,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ††† ide@nii.ac.jp

あらまし　近年、大量に蓄積された映像データを効率的・効果的に利用するための技術が求められている。放送映像を素材とした時、情報のキーとなる重要な要素の一つとして映像中の登場人物がある。登場人物に注目して知識の抽出や解析をおこなう研究は従来おこなわれているが、それらは専らテキスト情報を用いており、共通した問題として、状況や時流の変化に伴う同一人物の呼称の多様性が挙げられる。これに対処するには、同一人物の異なる呼称を “名寄せ” するための辞書作成が必要となるが、従来はこの辞書作成は人手にておこなわれていた。本研究では、この辞書作成を自動化するために、顔認識の技術を用い映像中に現れる顔の同定（いわば “顔寄せ”）をおこなうことで、それぞれの顔に対応する人物名詞の名寄せをおこなう。
キーワード　放送映像, 名寄せ, 顔認識

Name Identification of People in Broadcasted Video by Face

Takashi OGASAWARA†, Tomokazu TAKAHASHI†,††, Ichiro IDE†,†††, and Hiroshi MURASE†

† Graduate School of Information Science, Nagoya University　Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8603 Japan
†† Japan Society for the Promotion of Science
††† National Institute of Informatics　Hitotsubashi 2–1–2, Chiyoda-ku, Tokyo, 101–8430 Japan
E-mail: †{toga,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ††† ide@nii.ac.jp

Abstract　Recently, the technology to use a large amount of video data efficiently and effectively is requested. When broadcasted videos are used, people who appear in the videos are one of the most important information. So researches that focus on people’s name that appear in videos have been done. However, there is a big problem in these works; the problem that people could have various ways to be called according to the change in the situation and the current of the times. In this work, we will resolve the problem by identifying faces that appear in broadcasted videos.
Key words　Broadcasted Video, Name Identification, Face Recognition

1. ま え が き

1.1 研究の背景

近年、通信技術や記憶装置の発達、また、メディアの多様化に伴い、世の中では大量の映像データが氾濫の一途をたどっている。それに伴い、それら大量の映像データを効率的・効果的に利用するための技術が求められており、検索や閲覧、潜在している知識の抽出などを目指した計算機による内容理解が進められている。

映像の代表的なものとして、放送映像が挙げられる。放送映像は人間社会の貴重な情報源であると考えられるため、そこに現れる “人物” に注目することは自然かつ重要であると考えられる。

これを受け、我々は放送映像を解析し、そこに登場する人物たちが形成する相関関係を抽出することにより、図 1 に示すような人物相関グラフの構築を目指している [1], [2]。

放送映像は、主に画像と音声からなるマルチメディアデータであるが、近年これらに加え、クローズドキャプション (CC) と呼ばれる音声を書き下したテキストデータが付与されはじめている。上記の放送映像からの人物相関抽出に関する研究 [1], [2] においても、CC テキストを処理することで人物間の相関の強さを求めている。

放送映像を対象とした場合に限らず、このような “人物” に注目した研究においては、1 つの大きな問題がある。この問題は、“人物” が人物名詞（一個人を指す固有名詞）によって一意に識別されるものであるにもかかわらず、同一人物であっても状況や時流の変化によって呼称が異なるために生じる。つまり、1 人の人間が複数の人物名詞で言及されうるのである。これにより、処理の上で文字列として人物名詞を扱う場合に、事実上同一のものが、あたかも別のもののように扱われてしまう。

本報告では、この問題に対処すべく、同一人物の異なる呼称を同定（いわゆる “名寄せ”）するための知識を収集し、辞書を作成することを目指す。従来、このような実世界に関する高度な知識に基づく名寄せ辞書の作成は専ら人手にて行われていたが、本研究では自動化を目指している。

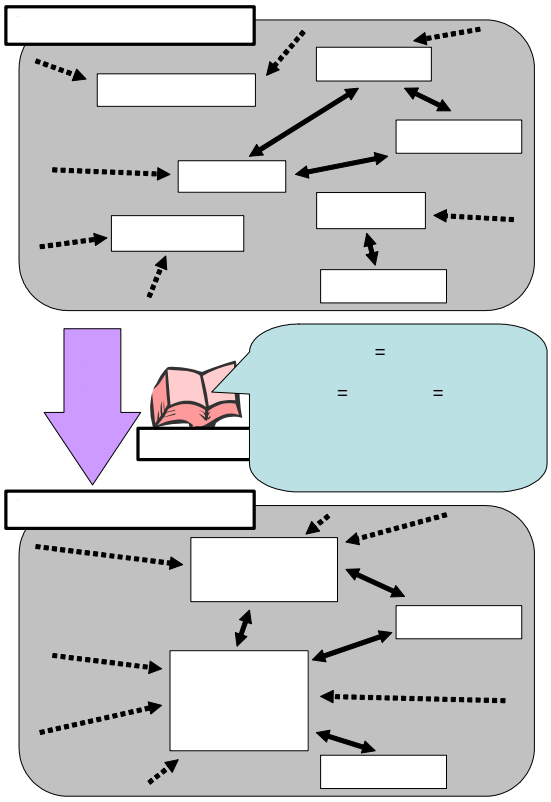

図 2　名寄せのイメージ図

を異なった語で表したために起こった呼称変化である。このような場合は、あらかじめシソーラス (類語辞典) を用意しておき、同義語の言い換えをおこなうことで対応は可能である。

（3） **時間的な変化**

c. から h. までを全て見ると、同義語の置き換えではなく、本来別の概念を指す名詞を付加したことによって、多様な呼称が見られる。図 3 に示した人物は、歴史のある時点では「厚生大臣」という役職についていたが、その後、「自民党総裁」や「総理大臣」となったため、このように呼称が変化している。時間的な変化は、図 3 の例のような役職の変化の他にも、事件を起こしたため逮捕・起訴された場合の「容疑者」や「被告」といった称号の付加や、婚姻等による姓の変化、改名など様々なものがある。長期にわたるデータに対する名寄せにおいては、このような変化はことさら大きな問題となり、社会の変化に対応した辞書更新が必要である。

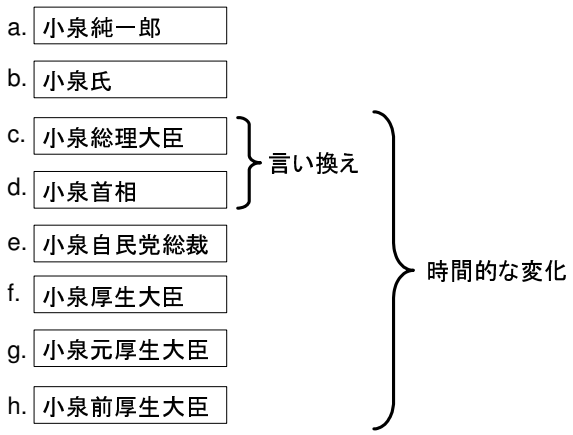

図 3　同一人物に関する呼称の多様性の例

以上のような呼称変化の原因およびその問題点を考えると、名寄せの自動化が大変困難であることがわかる。

1.3　提案手法の概要

以上のような難しさを鑑みると、外部から与えられる社会的な知識なくして、テキスト、すなわち人物名詞を見ただけで名寄せを行うことは困難である。

そこで、本研究では、CC テキストの情報に加え、放送映像中の画像情報を利用することで、名寄せに必要な知識としての辞書作成を自動的に行う。具体的には、顔認識の技術を用い映

像中に現れる顔を同定（いわば “顔寄せ”）することで、それぞれの顔に対応する人物名詞の名寄せを行う。

本報告では、上記の基本的アプローチを踏まえ、以降、2. で処理の詳細を述べる。3. で提案手法の効果を実験によって示し、4. でまとめる。

2. 顔認識を用いた名寄せ

放送映像中の CC から得られた複数の人物名詞に対し、それらに対応する画像中の顔どうしを比較し、その類似度をもとに名寄せを行う。処理の大まかな流れは図 4 のようになる。

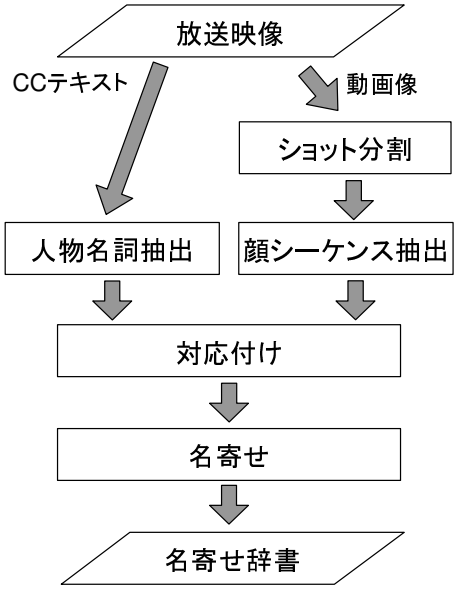

図 4　提案手法の処理の流れ

ここでは、まず 2.1 にて処理を解説する上で必要となる用語を整理し、つづく 2.3 以降で各処理の詳細を述べる。

2.1 用語の整理

以降では、対応付け処理について述べるにあたり必要となる、放送映像の画像的・意味的な構成要素に関する用語を整理しておく。

- フレーム：動画像の最小構成単位である静止画像
- ショット：画像的に連続するフレーム群
- カット：ショット間の不連続点
- シーン：意味的に連続するショット群 (ニュース映像においてはストーリ（話題）に相当)。

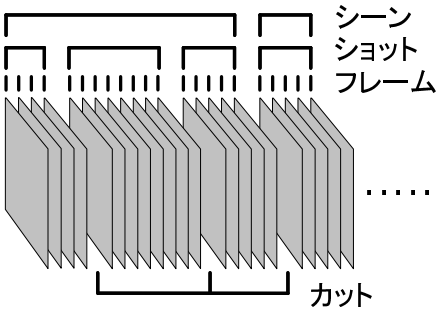

図 5　放送映像の構成

2.2 ショット分割

まず、全ての処理に先立って、ショットの分割をおこなう。以降、2.5 で述べる対応付けおよび 2.6 で述べる名寄せは基本的にショット単位でおこなうため、ここで述べるショット分割はそれらの前処理となる。

放送映像において、ショットは視聴者に対して違和感や誤解のないように配慮されて割り付けられる。そのため、インタビューや記者会見など、ある特定の人物を話題の中心としてショットが作られるとき、何らかの制約がない限り、対象となる人物は画面中央に大きく映るように撮影される。さらに、その次に違う人物を映す場合には、ショットを切り替えるよう配慮される。

この性質を利用し、本研究では人物名詞と顔とを対応付けるための手掛かりとして、ショットを単位として処理を進める。

ショットの分割は、時系列順にフレーム画像の RGB 色ヒストグラムを比較し、前フレームとの類似度が閾値以上となったフレームの直前をカットとする。

2 つのヒストグラム $\mathbf{H_1}, \mathbf{H_2}$ の類似度は、式 (1) で表されるヒストグラムインターセクションを用いる。

$$\mathbf{H_1}\text{と }\mathbf{H_2}\text{の類似度} = \frac{\sum_{i=1}^{I} \min(H_{1,i}, H_{2,i})}{\sum_{i=1}^{I} H_{2,i}} \tag{1}$$

$\mathbf{H_1}, \mathbf{H_2}$: フレーム画像の色ヒストグラム

I : ヒストグラムのビン数

$H_{n,i}$: $\mathbf{H_n}$における i 番目のビン

本報告では、ヒストグラムのビン数 $I = 256 \times 3$ とした。ここで、入力画像の色情報は、RGB それぞれの値が 0-255 の 256 階調である。

2.3 CC テキストからの人物名詞抽出

ここでは、各ショットに対応する CC テキスト中に現れる人物名詞を抽出する。CC テキストは、事前に音声との同期がとれているものを用いた。抽出される人物名詞は、完全一致の重複を除いたすべての人物名詞である。

人物名詞抽出手法は、井手らの手法 [4] を用いており、おおまかに以下に示す手順である。

手順 1　CC テキスト内の各文に形態素解析を施し (JUMAN [3] 3.61 を使用)、名詞列を抽出する。

手順 2　各名詞列に対し、「〜さん」「〜会長」「〜大臣」といった接尾名詞に注目して事前に作成した辞書と照合することで語義属性を解析し、人物名詞を抽出する。

2.4 動画像からの顔シーケンス抽出

ここでは、各ショット中に現れる顔を画像より抽出する。

顔の検出には、照明条件の変動やノイズに対してもロバストであり、かつ解像度に依存せず高速な検出が可能な Joint-Haar like 特徴 [5] を用いる。これは、複数の Haar-like 特徴 [6] の共起に基づく特徴量であり、Haar-like 特徴を組み合わせて顔の構造に基づいた特徴の共起関係を評価することにより高い識別能力を得ている。

Haar-like 特徴とは、画像における特徴量として、照明条件

の変動やノイズの影響を受けやすい各画素の明度値をそのまま用いるのではなく、近接する 2 つの矩形領域の明度差を求めることで得られる特徴量である (図 6)。

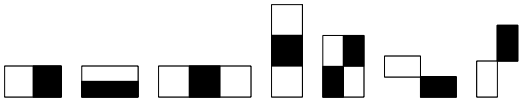

図 6　基本的な Haar-like 特徴のセット

2.2 で述べた放送映像における各ショットの特性から、各ショットから得られると期待される顔は 1 つ以下である。そのため、顔検出処理そのものはフレーム（静止画）1 枚 1 枚に対して行うが、ショット内の全てのフレームから得られた全ての顔は、まとめて 1 つの顔シーケンスと仮定して扱う。

顔を 1 枚の静止画ではなくシーケンスとして抽出することで、2.6 で述べる顔認識の精度を高められると期待される。

2.5　人物名詞と顔シーケンスの対応付け

2.3 にて抽出された人物名詞および、2.4 にて抽出された顔シーケンスを対応付ける。

前述のとおり、各ショットにおいて抽出された人物名詞は複数存在しうるが、顔シーケンスはたかだか 1 つのみである。ここでは、ショット内の顔シーケンスと人物名詞を 1 対多の状態で対応付けておく。

この処理は、佐藤らによる Name-It システム [9] のように、人物名と顔とを 1 対 1 に対応付けることそのものが目的ではなく、あくまでも “名寄せ” の手がかりにするためのものである。そのため、ここでは、一意の対応付けまでは行わず、ある顔に対応しうる人物名の候補を複数挙げるにとどめ、対応付けの誤りによる名寄せ漏れを防ぐ。

2.6　名　寄　せ

対応付けた顔シーケンスと人物名詞のペアを用いて名寄せをおこなう。

顔シーケンスと人物名詞のペアを、全ての組み合わせで比較し、以下の 2 つの条件を満たすとき、両者は同一人物であると判断する。

（1） **顔の類似度が閾値以上**

ここでおこなう顔の類似度計算は、特徴点として抽出した目鼻（瞳、鼻孔）を基準にして顔領域の位置・サイズを正規化した矩形の濃淡パターンを、相互制約部分空間法を用いて認識する手法 [7], [8] により行っている。この手法では、顔の向きや表情変化といった変動を吸収するために、動画像から得られる複数枚の画像を用いた認識を行うため、顔を 1 枚の静止画ではなく、シーケンスとして認識器に与える。

（2） **名詞の類似性が制約を充足**

本来ならば、顔の類似性のみで名寄せするか否かの判断をすべきだが、2.5 での対応付けが 1 対 1 にまで絞り込まれていないため、名詞による制約を設けている。

実際に設ける制約については、3.1 にて具体的に述べる。

3.　放送映像中の人物の名寄せ実験

2. で述べた名寄せ処理の効果を確かめるために、実際の放送映像に適用して実験をおこなった。

3.1　実 験 条 件

素材として使用した放送映像は、2001 年 7 月から 2005 年 8 月までに実際にテレビ放送された「NHK ニュース 7」から、無作為に選んで人手にて切り出した 30 ストーリ（合計約 120 分）である。

名寄せ判断のための顔類似度の閾値は、その上下によって、以下に定義する適合率、再現率のトレードオフが起こるが、本実験においては、適合率 100%（False Positive が 0）を保障するような厳しい値に設定した。

$$\text{適合率} = \frac{\text{正しく名寄せできた数}}{\text{名寄せ結果}}$$

$$\text{再現率} = \frac{\text{正しく名寄せできた数}}{\text{正解}}$$

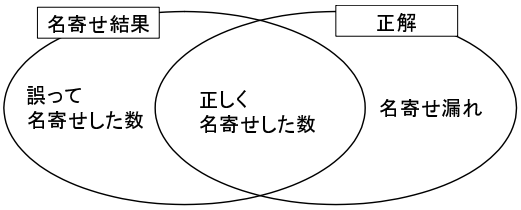

図 7　名寄せ結果の評価

ここで、“名寄せ結果” とは、名寄せ処理によって減らすことのできた異なり数、すなわち「名寄せ前の全人物名詞の異なり数」から「名寄せ後の人物数」を引いた値であり、“正解” とは、「正しくかつ漏れなく名寄せがおこなわれた時に最終的に得られる人物の異なり数」である。ここで定義した適合率の直感的な意味は「本処理によってなされた名寄せのうち、正しく名寄せできたものの割合」であり、再現率の直感的な意味は、「理想的な名寄せのうち、本処理で実現できた名寄せの割合」である。

なお、名詞による制約は「人物名の 1 文字目 (頭文字) が同じ」とした。

3.2　実 験 結 果

実験結果として、表 1 に、2.3 の処理によって抽出された「名寄せ結果」と、人手によって確認した「正解」およびそこから求められる再現率を示す。本報告の実験においては、前述のように、適合率 100%（False Positive が 0）、すなわち「誤って名寄せした数」= 0 を保障するように名寄せ判断のための顔類似度閾値を設定したために、「名寄せ結果」=「正しく名寄せした数」となる。

表 1　実 験 結 果

名寄せ結果（正しく名寄せした数）	正解	再現率
6	45	37%

3.3 考　　察

実験結果を見ると、再現率が 37%と低い値のため、現段階で十分な名寄せができているとは言い難い。しかしながら、テキスト面での条件が「頭文字の一致」という過剰な名寄せをしうるものであるにもかかわらず、顔の認識によって名寄せ誤りを抑制できていることは興味深い。

名寄せに失敗した例を見ると、失敗の原因は、大きく以下の 2 つに分類できる。

（1） **人物名詞に対応する顔が得られなかった**

さらに詳しく分析すると、人物名詞に対し、対応付けるべき顔の現れ方に注目して、以下のように分類できる。

(1-1)　そのショットではなく、前後のショットで現れた

(1-2)　顔向き等の条件の悪さから、検出に失敗した

(1-3)　そもそも画像中に顔が全く現れなかった

(1-1) については、対応付けるべき顔候補を抽出する範囲を、人物名詞と完全に同期をとるショットだけでなく、その前後のショットにも広げることで解決が見込める。(1-2) については、主に横向きの顔等を検出できる検出手法によりある程度は解決できるものの、後に控える顔認識処理に失敗する可能性があるため、困難な問題である。(1-3) については、本手法を適用することが原理的に不可能な問題である。

図 8 の例においても見られたように、このような場合は提案手法が原理的に機能しない。

今後は、顔の類似度計算等、各処理を高性能化することにより再現率の改善を図る。また、人物名詞と顔との対応付けを 1 対 1 に絞り込み、テキストによる制約を廃することで、あだ名や改名などにも対応した高度な名寄せを目指す。

現在は人物名詞どうしが同一人物を指すか否かのみを知識として抽出しているが、将来的な発展としては、時間の前後関係がはっきりしているという放送映像の利点を活かし、人名の時間的変遷を情報として加味することで、広い分野への利用を念頭においた知識抽出も検討している。

謝　　辞

本研究に不可欠である技術（2.4 の顔検出および 2.6 の顔認識）の提供およびご助言をくださった株式会社東芝研究開発センターに深く感謝する。研究に必要な数多くのデータを提供してくださった情報・システム研究機構国立情報学研究所に感謝する。日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に感謝する。本研究の一部は日本学術振興会科学研究費補助金、21 世紀 COE プログラム「社会情報基盤のための音声・映像の知的統合」による。本研究では、画像処理に MIST ライブラリ (http://mist.suenaga.m.is.nagoya-u.ac.jp/) を使用した。

文　　献

[1] 井手一郎, 佐藤真一, “ニュース映像からの人物関係の抽出と索引付けへの利用”, 第 63 回情報処理学会全国大会講演論文集, vol.2, pp.59-60, Sep. 2001.

[2] 小笠原崇, 高橋友和, 井手一郎, 村瀬洋, “放送映像からの人物相関グラフの構築”, 第 19 回人工知能学会全国大会, 1F4-02, pp.1-4,

June 2005.
[3] 京都大学長尾研究室，東京大学黒橋研究室：“日本語形態素解析システム JUMAN”, http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman.html
[4] 井手一郎, 浜田玲子, 坂井修一, 田中英彦, “テレビニュース字幕の語義属性解析のための辞書作成”, 電子情報通信学会論文誌 (D-II), vol.J85-D-II, no.7, pp.1201-1210, July 2002.
[5] 三田雄志, 金子敏充, 堀修, “顔検出に適した Joint Haar-like 特徴の提案”, 画像の認識・理解シンポジウム (MIRU2005) 論文集, pp.104-111, July 2005.
[6] C. P. Papageorgiou, M. Oren and T. Poggio, “A general framework for object detection”, Proc. of ICCV, pp.555-562, Jan. 1998.
[7] 山口修, 福井和広, “顔向き表情変化にロバストな顔認識システム “Smartface””, 電子情報通信学会論文誌 (D-II), vol.J84-D-II, no.6, pp.1045-1052, June 2001.
[8] 福井和広, 山口修, 鈴木薫, 前田賢一, “制約相互部分空間法を用いた環境変動にロバストな顔画像認識－照明変動を抑える制約部分空間の学習－”, 電子情報通信学会論文誌 (D-II), vol.J82-D-II, no.4, pp.613-620, April 1999.
[9] Shin’ichi Satoh, Yuichi Nakamura and Takeo Kanade, “Name-It: Naming and detecting faces in news videos”, IEEE MultiMedia, vol.6, no.1, pp.22-35, Jan.-Mar. 1999.
[10] Ichiro Ide, Hiroshi Mo, Norio Katayama and Shin’ichi Satoh, “Topic threading for structuring a large-scale news video archive”, Image and Video Retrieval -Third Intl. Conf. CIVR2004, Dublin, Ireland, Proceedings- P. Enser, Y. Kompatsiaris, N.E. O’Connor, A.F. Smeaton, A.W.M. Smeulders eds., Lecture Notes in Computer Science, vol.3115, pp.123-131, July 2004.
[11] 井手一郎, 佐藤真一, “人物関係に基づくニュース映像の検索と閲覧”, 電子情報通信学会 (パターン認識とメディア理解研究会) 技術報告 PRMU2001-48, July 2001
[12] “FOAF, the ‘friend of a friend’ vocabulary”, http://xmlns.com/foaf/0.1/
[13] 安田雪, “社会ネットワーク分析 -何が行為を決定するか-”, 新曜社, 1997.
[14] 原田昌紀, 佐藤進也, 風間一洋, “Web 上のキーパーソンの発見と関係の可視化”, 情報処理学会研究報告 DBS-130-3/FI71-3, May 2003.
[15] F. Yoshikane and K. Kageura, “Comparative analysis of coauthorship networks of different domains the growth and change of networks”, Scientometrics, vol.60, no.3, pp.435-446, Mar. 2004.
[16] E. Garfield, I. Sher and R. Torpie, “The use of citation data in writing the history of science”, Technical report, Philadelphia Institute of Scientific Information, 1964.
[17] 濱崎雅弘, 武田英明, 松塚健, 谷口雄一郎, 河野恭之, 木戸出正継, “Bookmark からの共通話題ネットワークの発見手法の提案とその評価”, 人工知能学会論文誌, vol.17, pp.276-284, Mar. 2002.
[18] J. Tyler, D. Wikinson and B. Huberman, “Email as spectroscopy: Automated discovery of community structure within organizations”, pp.81-96, Kluwer, B.V., 2003.
[19] 村田剛志, “参照の共起性に基づく Web コミュニティの発見”, 人工知能学会誌, vol.16, no.3, pp.316-323, May 2001.
[20] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment”, Proc. ACM-SIAM Symposium on Discrete Algorithms, pp.668-677, 1998.
[21] S. R. Kumar, P. Raghavan, S. Rajagopalan and A. Tokins, “Trawling the web for emerging cyber communities”, Proc. 8th WWW Conf., 1999.
[22] H. Kautz, B. Selman and M. Shah, “The Hidden Web”, AI Magazine, vol.18, no.2, pp.27-35, 1997.
[23] 松尾豊, 友部博教, 橋田浩一, 中島秀之, 石塚満, “Web 上の情報からの人間関係ネットワークの抽出”, 人工知能学会論文誌, vol.20, no.1, E, 2005.