FISH VIEW システム：概念体系に基づく視点情報を活用した文書整理支援

高間康史[†,☆] 石塚満[†]

インターネットに代表される情報環境の急速な成長により，大量の情報が容易に入手可能となりつつある反面，入手可能な情報量が人間の処理能力を超え，かえって効率が低下するという，いわゆる「情報過多（information overflow）」が深刻な問題となってきている．我々は，インターネットなどを通じて大量に収集された文書を熟読し，有効に活用するためには，文書間の関係を図解を用いて整理しつつ，漸進的に読み進めていくことが有効であり，この過程を計算機によって効率的に支援するためには，その時点におけるユーザの興味・視点情報を利用することが必要であると考える．この観点から，我々は Fisheye マッチングと呼ぶ，概念体系を利用した新しい動的ベクトル生成・マッチング機構を提案している．本稿では，Fisheye マッチングを基盤技術として文書整理支援を実現する，ビジュアルインタフェースを備えた FISH VIEW システムを開発したので報告する．FISH VIEW システムは，ユーザが図解として表現した文書間の関係から，Fisheye マッチングを用いてユーザの視点・興味に関する情報を抽出することができ，この情報を基に新規文書を検索したり，ユーザが見落としている文書間の関係を指摘したりするなどの支援を行うことができる．FISH VIEW システムをユーザに実際に使用してもらったところ，文書整理過程において有効な支援が行えていることが確認された．

FISH VIEW System: A Document Ordering Support System Employing Concept-structure-based Viewpoint Extraction

Yasufumi Takama[†,☆] and Mitsuru Ishizuka[†]

In order to deal with the vast collection of electronic documents to be collected, for example, from the Internet space, it becomes important to provide effective support functions for ordering such documents and thus finding some useful ideas. In this paper, we present such a support system called FISH VIEW system with a visual support function. This system interactively allows a user to order the collected documents into a diagram form while his/her reading work, by extracting user's viewpoint/interest from the diagram, finding documents related to his/her current viewpoint, and showing his/her viewpoint in a readable manner. This function is realized on the basis of the Fisheye Matching method, which has been proposed as an extension of the existing Vector Space Model for taking users' viewpoint into account based on the concept structure of an electronic dictionary. The extracted user's viewpoint is used by the system in several ways, such as to retrieve documents in a document database, to indicate relations among documents ordered into plural boxes of the diagram, and to present user's viewpoint as a set of concepts in a readable manner. Furthermore, the resultant or intermediate diagrams of the users' work can be generated as HTML pages, which are structured based on the rectangular regions in the diagrams. Several students have used FISH VIEW system actually, and have given us favorable comments.

1. はじめに

インターネットに代表される情報環境の急速な整備・拡大により，研究や仕事などに必要となる情報を収集する過程はますます容易になりつつある．その反面，入手可能な情報量が人間の情報処理能力を超え，かえって効率が低下するという，いわゆる「情報過多（information overflow）」が問題となってきている．すなわち，今までは情報不足が知的活動の足かせとなっていたのが，今後は収集した情報をいかに生かしきるかが死活問題になるといえよう．

企業においても，**ナレッジマネジメント**というスローガンの下に，社内情報・知識の集約および有効活用が，生産性向上の切札として注目を集めている[3)]．

† 東京大学工学部電子情報工学科
Department of Information & Communication Engineering, School of Engineering, University of Tokyo
☆ 現在，東京工業大学大学院総合理工学研究科
Presently with Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

特に，社内情報は文書の形で蓄積されているものが多く，テキストマイニングや発想法・発想支援[7]，オントロジー・シソーラス[10]などの要素技術が今後ますます重要となるであろう．

ところで，情報環境がもたらすこのような変革は何も企業や組織レベルにおいて起こるだけではない．近年，WWW 上のホームページや CD-ROM といった形態で，個人レベルでも大量の文書情報が容易に入手可能となっている．反面，入手した情報の活用に関しては従来のままであり，せっかく収集した情報を十分に活用できているとはいいがたいだけでなく，無意味な情報に埋もれて重要な情報を見落とすなどの逆効果も考えられる．すなわち個人レベルでの作業プロセスの中心は，関連する情報の収集や文書の編集といったプロセスから，大量のドキュメントをいかに読みこなし，全体を把握し，目的にあった情報，アイデアを取り出すか，に移るべきである．

我々は，個人レベルでの文書情報活用を目的として研究を行っている．これまでに，大量文書情報の熟読につながる文書整理プロセスおよび，このための要素技術として概念体系を利用した新しい動的ベクトル生成・マッチング機構 **Fisheye マッチング**を提案し，適合フィードバック法を用いた通常のベクトル空間モデルと同等の検索性能と，視点の明示化を同時に満たすことを示した[12]．同時に，Fisheye マッチングの持つこれらの特徴を生かし，個人レベルでの文書整理プロセスを有効に支援できる可能性についても例示した．本稿では，これらの成果を基に，図解処理能力や整理結果出力機能，グラフィカルインタフェースなどを整備した，実用的な文書整理支援システム **FISH VIEW システム**を開発したので報告する．

本稿の構成は以下のとおりである．まずはじめに，提案する個人レベルでの文書整理プロセスについて 2 章で説明する．Fisheye マッチングの概要について 3 章で説明し，続く 4 章で Fisheye マッチングの文書整理支援への応用について考察および予備実験を行った後，5 章で開発した FISH VIEW システムを紹介する．評価実験結果について 6 章で紹介した後，7 章で結論を述べる．

2. 個人レベルでの文書整理プロセス

個人レベルでの情報整理・活用における特徴は，「情報整理者＝情報活用者」という点である．たとえば yahoo! などの文書ディレクトリサービスにおいては，不特定多数のユーザによる，多様な利用目的を前提としているため，そのディレクトリ構造は一般的なものにする必要がある．これに対し，個人レベルで自分のために，入手情報を整理する場合に作成される構造は，ユーザのその時点における目的（何のためにこれらの情報を収集したか）や，ユーザの知識状態に依存して決まる．したがって，あらかじめ固定的に用意されている分類構造を利用することは困難であり，ユーザ自身の手で行われる整理，分類結果を尊重し，試行錯誤の過程を支援する必要がある．

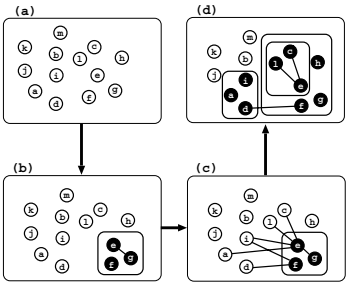


図 1 文書整理プロセス
Fig. 1 Process for document ordering.

このような観点の下に，個人レベルでの情報整理を支援するシステムの開発も行われている[6),11),13)]．また，発想支援，特に収束的思考支援なども同様の視点を共有しているといえる．これらの研究においては，メモのような細切れの情報を対象とし，ユーザの手でそれらを入力していく形態を想定しているものが多い．これは，KJ 法[5]に代表される発想法・発想支援システムにおいても同様である．すなわち，整理対象となるものは文書から作成されたメモであり，そのメモは，ユーザ自身が文書を読み，本質を把握したうえで簡潔に抽出したキーワード，短い文章からなるものである．このような，重要キーワードや概念が明確にされたメモ間の関係，類似性を把握することは，計算機を用いても比較的容易に行えるが，収集文書を読みこなすといった困難な作業に関しては，依然ユーザの手に委ねられたままである．我々が文書整理プロセスで支援したいのは，この各文書の読解・理解に関する作業であり，情報検索作業と図解作成作業を組み合わせた，以下の文書整理プロセスを提案する（**図 1**）．

読書 今までに読んだドキュメントや，頭の中の知識との関連を意識しながらドキュメントを読む．

図解作成 今まで読み進めてきたドキュメント群から**局所図解**を作成し，視点を整理する（図 1(b), (d)）．

検索 得られた視点を基に，次に読むべき文書を決定する（図 1 (c)）．

すなわち，今までの読解結果を基に局所図解を漸進

的に作成する過程を通じて，ユーザは今まで読み進めてきた文脈を整理し，次に読むべき文書を決定する．図解編集は考えをまとめるだけでなく，このような読解作業を助けるうえでも重要な役割を果たすとしているのが，我々の提案する文書整理プロセスの特徴である．

この文書整理プロセスを計算機を用いて効率的に支援するためには，ユーザの視点に関する情報を抽出し，利用する技術が必要となる．これに関して，我々は視点情報を扱うようにベクトル空間モデルを拡張した**Fisheye マッチング**と呼ぶ，新しい動的文書マッチング機構を提案しており[12]，次章でその概要を説明する．

3. Fisheye マッチング

情報検索において，文書の表現形式として用いられる代表的な手法の 1 つにベクトル空間モデルがある[9]．ベクトル空間モデルは，キーワード検索などの Boolean match に比べ，関連度による評価が可能な best match が実現できること，および適合フィードバック法[9]などにより，ユーザの興味を反映した文書を高精度で検索できるといった長所から，多くの研究/システムにおいて使用されているが，以下の問題点を持つブラックボックス的な操作であると見なせる．

- クエリー（質問）ベクトルを人手で生成・修正することは困難である．
- クエリーベクトルが表している（はずの）興味を明示的に把握できない

すなわち，ベクトル空間モデルにおける次元（単語）数は一般に 10^3 以上のオーダであるため，各単語の重みをユーザ自身の手で調整することは非現実的である．一般には，ユーザは現在のクエリーベクトルによる検索結果に適/不適の判断を下し，適合フィードバック法を用いてクエリーベクトルを修正する方法をとらざるをえない．

しかし，適合フィードバック法を用いて判断できるのは興味のあり/なしであり，その興味が何であるか（スポーツに関すること，車に関すること，など）については何も教えてくれない．すなわち，興味の背景にあるユーザの視点に関して，クエリーベクトルから読みとることは困難である．また，ベクトル空間モデルにおいて仮定される，各軸（単語）間の直交性も問題である．すなわち，ある視点から見て共通の特徴と見なせる単語群も，つねに別々に扱われてしまう．

次節で紹介する Fisheye マッチングは，上記問題点を解決するためにベクトル空間を拡張したものである．

3.1 Fisheye マッチングの定義

Fisheye マッチングでは，概念体系を背景知識として用いてベクトル空間を構築することにより，ベクトル空間モデル（および適合フィードバック法）の持つ上記問題点を解消することができる．

Fisheye マッチングでは，概念体系から計算された，ある意味を共有した単語グループ（**意味グループ**と呼ぶ） g_i を単位として，単語の選択（Magnify）/縮退（Shrink）を行うことにより，ユーザの視点を反映した特徴ベクトル空間を構築する．

単語集合を $W = \{w_1, w_2, \ldots, w_n\}$ とすると，これから Shrink，Magnify などの各演算子によって得られる Fisheye ベクトルの特徴集合 $S(g_1, \ldots, g_m|W)$，$M(g_1, \ldots, g_m|W)$ は各々次式のように定義される．

$$S(g_1, \cdots, g_m|W) = \{f_i | f_i = \{w_j | w_j \in g_i \cap W\}, \; i \in [1, m]\}, \quad (1)$$

$$M(g_1, \cdots, g_m|W) = \{f_i | f_i = \{w_i\}, \; w_i \in (g_1 \cup \cdots \cup g_m) \cap W\}, \quad (2)$$

$$\widetilde{S}(g_1, \cdots, g_m|W) = S(g_1, \cdots, g_m|W) \cup \overline{M}(g_1, \cdots, g_m|W), \quad (3)$$

$$\overline{M}(g_1, \cdots, g_m|W) = \{f_i | f_i = \{w_i\}, \; w_i \in W - (g_1 \cup \cdots \cup g_m)\}. \quad (4)$$

意味グループ g_i については，EDR 電子化辞書☆の概念体系辞書中に存在する概念のうち，体系上の下位にある単語数が 2 以上 256 以下であるものを選び，意味グループとして採用する．g_i は単語集合 W とは無関係に，概念体系辞書から求めたものであり，各特徴 f_i は，W に含まれる単語のみを含むように生成する．後述するが，EDR 電子化辞書中の各概念には，それを説明する文章（あるいは単語）が記述されており，これら説明情報をユーザに提示することによって視点情報の外化が行える．たとえば，**図 2** に示す概念体系☆☆（g_i は意味グループを表す）を用いて生成される特徴集合を以下に示す☆☆☆．

$S(g_2, g_3|W)$={{bicycle, car},{apple, lemon}}
$M(g_2, g_3|W)$={{bicycle}, {car}, {apple}, {lemon}}
$\widetilde{S}(g_2, g_3|W)$={{bicycle, car},{apple, lemon},{tomato}}

これより，Shrink は単語を概念に縮退した，粗い特徴空間上において，文書どうしが「同様の話題に関連しているか」をみる場合の操作である．反対に Magnify は，概念体系中の興味ある部分をルーペで拡大し，そこに含まれる単語のみを特徴とすることにより，ある話題に限定した場合の文書間の関連度を

☆ http://www.iijnet.or.jp/edr/
☆☆ EDR 概念体系辞書から実際に抽出したものとは異なる．
☆☆☆ $W = \{apple, lemon, tomato, bicycle, car\}$

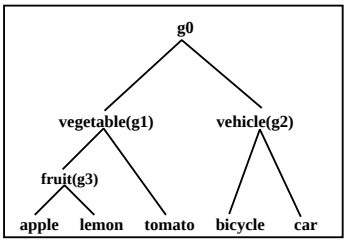


図 2 概念体系の例
Fig. 2 Example of concept structure.

みる場合に用いる．また，Magnify は特徴として使用する単語を概念単位で選択する操作としてみることもでき，$\widetilde{S}(g_1,\cdots,g_m|W)$ では Shrink 操作によって起こる次元数の低下を補うために使用されている．$\widetilde{S}(g_1,\cdots,g_m|W)$ は，Shrink 操作の対象とならなかった W 中の単語についてはそのまま特徴として用いる操作であり，文書検索実験の結果，適合フィードバック法を用いた通常のベクトルモデルと同等の検索性能を維持しつつ，視点情報を可読形式で外化できることを報告している[12]．本稿で記す FISH VIEW システムでも，文書間の関連度計算や新規文書検索に $\widetilde{S}(g_1,\cdots,g_m|W)$ を利用する．

ベクトル空間モデルにおける軸の直交性の問題を解決する方法として LSI（Latent Semantic Indexing）[1] が有名であるが，この手法では，新たな軸を各単語の線形和として求めており，各軸の意味するもの（対応概念）が必ずしも明らかではない．したがって，視点情報を明示的に扱う Fisheye マッチングとは性質が異なるものである．また，シソーラスを用いた検索[2),4)] に関する研究は従来からも行われているが，それらの目的は類義語の扱いであり，ユーザの視点を明示的に扱うことではなかった．また，概念単位で文書をカテゴリ化する研究[8)] も存在するが，1 つの文書は 1 つの概念のみに対応づけられる固定的なものであった．これに対し Fisheye マッチングでは，各概念単位で単語をグループ化し，ベクトル空間を構築することにより，ユーザの視点をベクトル空間を構成する概念の集合という形で，明示的かつ柔軟に表現することが可能である．

以上の操作により求められた特徴に基づいて，ドキュメント d_i に対する Fisheye ベクトル $F_i(v_{i1},\cdots,v_{im})$ は，d_i に対する単語を要素とした通常の特徴ベクトル（**基本特徴ベクトル**） $O_{d_i}(w_{i1},\cdots,w_{in})$ より求める（ここで，O_{d_i} の各要素の値については，TFIDF[9)] により求められているものとする）．すなわち，F_i における j 番目の特徴の値 v_{ij} は，対応する意味グループ f_j に属する単語 $word_k$ の重みの総和より求める（式 (5)）．また，ドキュメント d_i，d_j 間の類似度 $Sim(d_i,d_j)$ については，通常のベクトル空間モデルと同様に，両ドキュメントに対応する特徴ベクトル（ここでは Fisheye ベクトル）$F_i(v_{i1},\cdots,v_{im})$，$F_j(v_{j1},\cdots,v_{jm})$ の内積を基にして求めることができる（式 (6)）．

$$v_{ij} = \sum_{word_k \in f_j} w_{ik} \ , \tag{5}$$

$$Sim(d_i,d_j) = \frac{1}{2}\left(1 + \frac{\sum_{k=1}^{m} v_{ik}\cdot v_{jk}}{Mag(d_i)\cdot Mag(d_j)}\right), \tag{6}$$

$$0 \le Sim(d_i,d_j) \le 1 \ ,$$

$$Mag(d_i) = \sqrt{\sum_{k=1}^{m} v_{ik}^2} \ . \tag{7}$$

3.2 視点意味グループ集合の抽出

Fisheye マッチングにおいて，概念体系中に存在する概念は，単語をグループ化する際のプリミティブ（意味グループ）であり，意味グループ集合の形で特徴ベクトル空間を構築することにより，ユーザの視点を表現する．これは，概念単位で文書を直接グループ化する手法と比較して，辞書中に対応する概念が単独で存在しない場合でも，概念の組合せで視点を表現できる柔軟性や，複数視点を同時に扱えるなどの利点を持っている．しかし，EDR 辞書から求められる意味グループ数は膨大であり，人手で適切な質・量の意味グループを選択，指定することは（単語レベルでの調整ほどではないにしても）困難である．したがって，ユーザの視点を反映した文書の分類結果や図解などから，視点に対応する意味グループ集合を抽出できることが望ましい．また，抽出された意味グループをその説明情報とともにユーザに提示することにより，今まで意識していなかった視点に気づいたり，漠然としていた考えが明確になったりするなどの効果も期待できる．

このような観点から，我々は適合フィードバックを拡張した，意味グループ抽出アルゴリズムを提案している[12]．アルゴリズムついては以下のとおりである．ここで，「ユーザの興味を表す概念（意味グループ）には重要な（重みの大きい）単語が多く属している」との仮定に基づき，重みの大きな単語から順に，グリーディに意味グループを抽出している．

(1) 基本特徴ベクトルの要素として 1 回以上出現する単語の集合を W とする．$word_i \in W$ について，次式により重み w_i を計算し☆，重みが正の単語を $Wlist$ に格納する．ただし，D_P，D_N

☆ α は正負のバランスをとる適当な係数.

はそれぞれ，ある視点のもとにユーザが興味あり/なしと判断した文書集合である．$count = 0$ とする．

$$w_i = \alpha \frac{1}{|D_P|} \sum_{d_j \in D_P} w_{ji} - \frac{1}{|D_N|} \sum_{d_k \in D_N} w_{ki}. \quad (8)$$

(2) $Wlist$ 中より，最大の重み w_k を持つ単語 $word_k$ を取り出す．なければ終了．

(3) 意味グループ集合 $G_k = \{g_i | (word_k \in g_i) \wedge (\forall word_j \in g_i \cap W, word_j \in Wlist)\}$ を求める．

(4) (3) で求めた G_k 中の各グループ g_i について，重み W_{g_i} を次式に従って計算する．$G_k = \emptyset$ の場合には (6) へ．

$$W_{g_i} = \frac{1}{|g_i|} \sum_{word_j \in g_i \cap W} w_j \ . \quad (9)$$

(5) G_k 中で，重みが最大のグループ g_l を抽出．$Wlist = Wlist - (g_l \cap W)$ として (7) へ．

(6) $Wlist = Wlist - \{word_k\}$ として (7) へ．

(7) $count = count + 1$ とし，$count = n$ となったら終了．それ以外は (2) へ．

このアルゴリズムを用いて抽出された意味グループを用いて Fisheye ベクトルを生成することにより，ユーザの視点に合致した文書検索が行えることを先に報告している[12]．この実験において抽出された意味グループについて，その説明情報と所属単語について**表 1** に記す．これは，ユーザが医学の記事に関して興味を持っている場合に抽出されたものである．このように，抽出された意味グループに関する説明情報や単語集合をユーザに提示できるため，現在，どのような視点によってドキュメント間の関係を捉えようとしているのかを知る手がかりをユーザに提供できる．また，ユーザの視点（例：野球）よりも上位（例：スポーツ）の概念に対応する意味グループが抽出されてしまったり，形態素解析においてありえない区切りで抽出された単語を含む意味グループが抽出されてしまった場合でも，ユーザが判断して視点から削除したり，視点にふさわしい意味グループを新たに追加したりといった編集作業が容易に行える．Fisheye マッチングの持つこのような性質は，本稿の FISH VIEW システムにおいても積極的に利用している．

また，ステップ (7) において，アルゴリズムの適用対象とする単語数を n に限定している．これは，全単語を対象とした場合に抽出される意味グループ数は 20〜50 程度であり，これらすべてを視点に関する情報としてユーザに提示するのでは多すぎるためである．また，Shrink 操作は，特徴縮退によるマッチング対象の拡大（いわゆるクエリー拡張）効果だけでなく，単語の抽出ミスや意味グループ抽出ミスのため，検索精度が低下するというデメリットも存在する[12]．したがって，視点情報の提供，マッチング精度向上のどちらにおいても，適切な意味グループのみを利用すること，すなわち「視点意味グループの厳選」を行うことが望ましい．前述のように，提案する抽出アルゴリズムでは，重みの大きい単語が集まっているほど，ユーザの興味を表す概念であるとの仮定に基づいているので，単語の重みが上位 n 単語に限定して意味グループの抽出を行うことにより，視点意味グループの厳選が行えるものと考えた．n の値を変えて，検索精度および計算時間を比較した結果より，本稿では $n = 20$ とした．このとき，全単語を対象として意味グループを抽出した場合と比較して，処理速度は 2〜3 倍高速になり☆，また検索文書数が増加しても適合率があまり低下しないという利点が確認された．

表 1 医学関係の記事から抽出された意味グループの例
Table 1 Examples of extracted semantic groups related to medical topics.

説明	所属単語
医薬品	ワクチン 漢方薬 薬剤 目薬 下剤 ...
循環器	心臓 心肺 動脈 毛細血管 大動脈 ...
身体機能の状態捉えた人間	妊婦 患者 入院患者 障害者 痴呆
身体	体 顔 筋肉 口 手足 首 足 皮膚 ...
病気	慢性 うつ病 つわり 痴呆 エイズ ...
天然食品	ニンニク 野菜 果物 なし コメ
雌の生殖器官	卵管 乳 乳房

4. Fisheye マッチングの文書整理支援への適用

2 章で提案した文書整理プロセスを能動的に支援する FISH VIEW システムの開発を行った．本章では，Fisheye マッチングを文書整理プロセスに適用することのメリットについて述べる．具体的には，Fisheye マッチングによって実現可能な支援は以下のとおりである．

(1) 局所図解からユーザの視点を把握・提示する．

(2) 次に読むべき文書を検索し，提示する．

(3) 図解において，ユーザが見落としている文書間の関係を指摘する．

(1) は，(2)，(3) の支援を有効に行ううえでも必

☆ 通常の適合フィードバックと比べると 4〜5 倍の計算時間を要する．

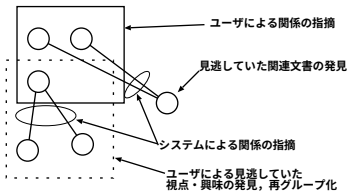


図 3 FISH VIEW システムにおける図解の概要
Fig. 3 Image of diagram employed by FISH VIEW System.

要不可欠である．また上述したように，Fisheye マッチングでは抽出した視点に関する情報を，各意味グループの所属単語や説明により可読な形で提示することができるので，図解に込められた曖昧な視点を明示的な形でユーザに提示する**視点の外化**効果も期待できる．

本研究で採用する図解においては KJ 法[5]と同様に，ユーザがある視点・興味において近い，関連があると思った文書をグループ化することによって表現する（**図 3**）．ユーザは，図解中のグループを指定することによって，現在の視点をシステムに伝えることができる．すなわち，指定されたグループ内部に存在する文書を正例 D_P，外部にある文書を負例 D_N とすることにより，3.2 節で紹介した視点意味グループ集合抽出アルゴリズムを用いて視点意味グループを抽出できる．

(2) については，指定された視点・興味情報を基に，その視点に関連が深いと思われる文書を検索する．すなわち，指定された視点意味グループを引数として $\widetilde{S}(g_1, \cdots, g_m|W)$ 演算を用いて特徴集合を生成し，グループ内部の文書から作成したクエリーベクトルと，各未読文書から求めた Fisheye ベクトルのマッチングを行うことによって，ユーザの視点・興味と関連の深い新規文書の検索を行う．

(3) に関して，大量の文書を対象として整理を行う場合，視点に関連する文書を見逃してしまう可能性が高くなる．本研究では，図解上の文書間の類似度を Fisheye マッチングを用いて計算し，関連度が高い文書間にリンクを張ることによって，このような見落としを防ぐようにする．リンクによりシステムが指摘する情報は，見落としていた関連文書の発見だけでなく，ユーザが想定していなかった文書のまとめ方，すなわち視点に気づかせる効果においても有効であると考えられる（図 3）．

5. 文書整理支援を実現する FISH VIEW システム

図 4 は，我々が開発した文書整理支援を実現する FISH VIEW システムの全体像である．FISH VIEW

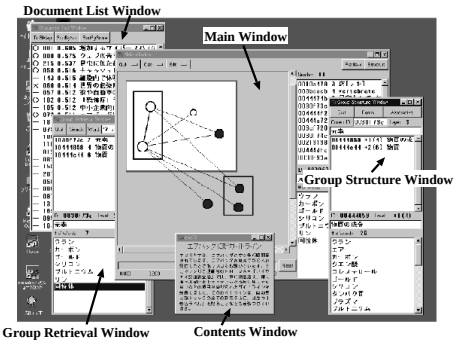


図 4 FISH VIEW システム（クライアント）の画面
Fig. 4 A display image (client) of FISH VIEW System.

システムはクライアント・サーバ型として開発されており，クライアントは Tcl/Tk8.1 を用いて実装され，Windows，UNIX の両プラットフォームで実行可能である．サーバは C 言語を用いて UNIX 上に実装され，EDR 電子化辞書から求めた単語，概念（意味グループ）に関するデータベースおよび，文書データベースを持っている．

図 4 の中央に存在するウィンドウ（Main Window）上で，ユーザは図解を作成することにより文書整理を行う．図解中の文書については，図 4 の中央下部にあるウィンドウ（Contents Window）によりその内容（本文）を確認できる．図 4 左側の一番背後にあるウィンドウ（Document List Window）は，整理対象となる文書のリストを表示する☆．リスト中での順番は，ユーザの視点との関連度に基づいてソートすることができる．

左側にあるもう 1 つのウィンドウ（Group Retrieval Window）は，単語をキーとした意味グループの検索に用いる．同様に，右側にあるウィンドウ（Group Structure Window）は，意味グループをキーとして，概念体系上で上下位関係にある意味グループの検索に用いる．これらのウィンドウを用いて，ユーザは視点情報の編集を行うことができる．

限られた紙面で FISH VIEW システムの全機能を紹介することは困難だが，このシステムを用いた文書整理プロセスは，概略すると以下のフェーズを繰り返して文書整理・読解を進めていく．

図解作成 Document List Window からの文書選択，図解への追加．文書内容の確認および図解による

☆ システム起動時は，登録された順序で表示される．また，視点が指定されていない状態で「視点との関連度に基づくソート」が実行された場合は，文書データベース中の全文書に関する基本特徴ベクトルを加算平均して求めたベクトルと，各基本特徴ベクトルとの関連度に基づいて行われる．

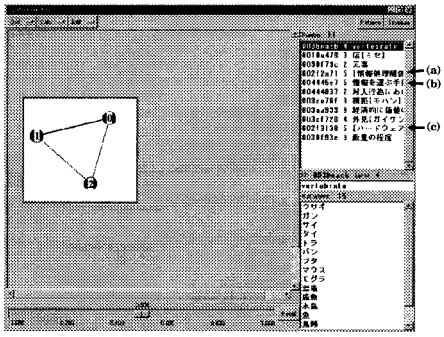


図 5 図解作成および視点抽出の例
Fig. 5 An example of a diagram and feature extraction.

文書整理.

視点情報抽出・編集 グループを指定することによる Fisheye マッチングを用いた視点抽出（3.2 節）．ユーザの手による視点情報の編集.

新規文書検索 Fisheye マッチングによる文書検索．結果は Document List Window に格納.

図解の HTML 化 最終的な図解を HTML 化.

以下では実際の作業例をあげ，上記プロセスの概要を述べる．

5.1 ステップ 1：初期図解の作成・視点の抽出

5.1.1 図 解 作 成

図解作成は，Document List Window から次に読みたい文書を選択する作業から始められる．選択された文書に対応するノードが生成され，図解に追加される．**図 5** は，インターネットに関する 3 文書を選択し，図解を作成した際の Main Window の様子である☆．グループ（矩形），およびノード間のリンクは 4 章で紹介したとおり，それぞれユーザ，システムによる関連性の指摘である☆☆．

リンクは Fisheye マッチングで求めた関連度 $Sim(d_i, d_j)$（式 (6)）の値があるしきい値を超えた文書間に生成される．リンクの太さは $Sim(d_i, d_j)$ の値により 3 種類存在する．また，文書ノードは，その時点におけるクエリーベクトル q との関連度 $Sim(q, d_i)$ に比例し，そのサイズが大きくなる．リンク生成時におけるしきい値や，$Sim(q, d)$ とノードサイズの関係（比例係数）は Main Window 内下部のスケールバーによって調整可能である．

5.1.2 視点意味グループ集合の抽出

ユーザはグループ作成後，Fisheye マッチングにより視点情報を抽出したい場合には FIX 化という操作を行う．FIX 化されたグループは，視点を固定した状態と見なされ，ノードの追加/削除などの変更は行えなくなる．FIX 化されたグループは，図解中で白色で表現される．

抽出された意味グループおよびそれに関する情報（所属単語，説明など）は，Main Window 内右部に表示される．上部の意味グループリストには，対応する概念の ID や，説明などが表示され，所属単語についてはリスト中で選択することにより下部に表示される．

5.2 ステップ 2：視点編集・新規文書検索

5.2.1 視点意味グループ集合の編集

図 5 において抽出された視点情報を見ると，「[情報処理関係の場所]」（図 5 中 (a)），「情報を運ぶ手段やシステム」（図 5 中 (b)），「ハードウェア操作」（図 5 中 (c)）など情報処理に関連する意味グループが抽出されている．また，「経済的に価値のあるもの」や「対人行為における役割で捉えた人間」といった意味グループも抽出されており，これらは「サイバーゴールド（文書 0）」や「インターネット・ガイド（文書 2）」などの話題に関連したものであるといえよう．

Fisheye マッチングにおいてつねに適切な意味グループ集合を抽出することは，単語や意味グループのノイズ[12]の存在により不可能である．この場合，ユーザの手で視点情報に関し以下の修正を行うことができる．

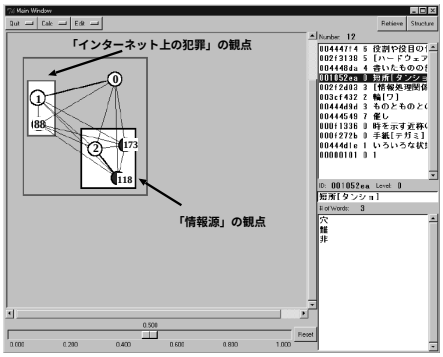
- 不要な意味グループの削除
- ある単語に関する意味グループの，他の意味グループとの入換え
- 新たな意味グループの追加

意味グループの入換え，追加は，Group Retrieval Window や Group Structure Window を用いることにより容易に行える．すなわち，Group Retrieval Window では，単語をキーとし，その所属する意味グループを検索できる．Group Structure Window では，意味グループをキーとし，上下位関係にある意味グループを検索できる．Fisheye マッチングでは，各単語が複数の意味グループに所属することを許していないため，これらを用いて検索された意味グループを視点意味グループ集合に追加した場合，単語を共有する意味グループは視点から削除される．

図 5 では，「“vertebrate”（脊椎動物）」という意味グループも抽出されているが，これは「サイバー」が

☆ 実際は，ノード内にマウスカーソルを入れることによってタイトルがバルーン表示されるが，ここではノード番号を記してある．対応する文書のタイトルは図下部に記した．ここで，タイトルだけでなく本文も含めて単語を抽出し，ベクトル空間を構築している．

☆☆ 視点が指定されていない場合は，基本特徴ベクトルに基づいてリンクが計算される．



88: ウェブの悪質な情報の対策を検討
118: アレルギーに関する情報を満載，ウェブサイト「アレデイズ」
173: コンピュータ・ウィルスの情報を提供

図 6 犯罪に関連する文書グループからの視点抽出直後
Fig. 6 Feature extracton from documents about crime.

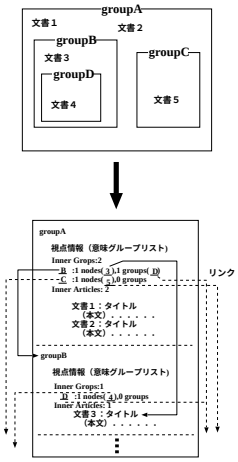


図 7 図解からの HTML 文書の生成
Fig. 7 Generation of HTML pages from diagrams.

形態素解析で “サイ” と誤って認識されたことによる．このような視点として不適切な意味グループについては，ユーザの手によって視点情報リストから削除することができる．

5.2.2 新規文書検索

図解中の（FIX 化された）グループを選択し，新規文書検索を指示すると，指定されたグループに関する視点意味グループ集合（Main Window 内右部に表示）を基に $\widetilde{S}(g_1, \cdots, g_m|W)$ 演算を行って特徴空間を構築する．このとき，クエリーベクトルの各要素の値は，意味グループに対応する特徴の値は式 (9)，単語に対応する特徴の値は式 (8) で計算した値を用いる．D_P，D_N については 4 章で述べたとおりである．生成されたクエリーベクトルと各文書 d_i の間の関連度 $Sim(q, d_i)$ を計算し，その値に基づいてソートした結果を Document List Window に表示する．このとき，図解中のリンクやノードサイズについても再計算が行われる．

図 6 は，“vertebrate” に関する意味グループを視点情報から削除した後，新規文書検索を行い，Document List Window 中で上位の 3 文書を図解に追加し，再編集した図解を示している．ここで，文書 2 と 118，文書 0 と 1 の間にそれぞれ張られている太いリンクの意味を吟味すると，文書 2 と 118 はともに，百科辞典，データベースといったインターネット上の情報源に関する話であり，文書 0 と 1 はともに，インターネット上の経済，価値あるものに関する話であるといった視点を見出すことができる．

図 6 では，システムによるこれらの関連性の指摘のうち，文書 2，118 間の関係を生かし，「情報源」の観点からグループ化している（図解中右下）．また，図解中左上のグループは「インターネット上の犯罪」の観点から作成したものであり，図 6 右側にその視点情報が表示されている．これを見ると，情報処理関係の意味グループに加え，「短所」（図 6 中白抜きで表示）という否定的な意味合いの意味グループが抽出されている．これは，犯罪に関する記事は，一般に否定的，非難する立場から記述されることが多く，文中に「難」や「非」といった単語（漢字）が多く使用されるためである．これらの単語を「短所」という意味でまとめ，犯罪に関する話題の持つ，否定的意味合いを抽出できたことは，Fisheye マッチングによって扱うことのできる視点情報の一例として興味深い．

5.3 ステップ 3：文書整理結果の出力

作成した図解は，文書を読み進めた結果得られたユーザの思考空間，概念構造を表したものと考えることができる．本研究では，ユーザが最終的に完成させた図解を HTML ページに変換して出力可能である．これは，KJ 法 B 型プロセス（図解の文章化）[5] を自動化したものと考えることができる．

図 7 に示すように，図解から HTML への変換はグループ単位で，階層構造を生かして行われる．文書番号やグループ番号からは，対応する文書，グループに関する実際の情報が記述されている部分へのリンクを張り，ハイパーテキスト化する．これにより内容確認が容易となり，サーベイや簡単なレポート代わりに活用することができる．

6. 評価・考察

本章では，FISH VIEW を実際に複数のユーザに使ってもらった結果について報告する．実験用のデータとしては，専門知識を必要とせず，ほとんどの被験

表 2 アンケート回答結果
Table 2 Results of questionnaire.

	有効回答数	平均	○	×
HTML 出力は便利か	7	6.29	7	0
視点の自動抽出は便利か	7	4.43	3	1
視点情報は分かりやすいか	7	3.29	2	4
文書検索結果は適切か	6	4.83	3	1
リンクは役立ったか	7	4.43	4	2
リンクは適切か	5	4	2	3
今後も使いたいか	7	5.29	5	2

者に興味を持ってもらえそうであること，およびいろいろな話題（視点）を含んでいること，などを考慮して，WWW から映画の批評に関する文書を 101 用意し，これを用いて文書整理を行ってもらった．文書整理作業という短期集中的な作業において，すべてに目を通すには多すぎる程度の文書が存在する場合に，システムによる支援が有効となると考えているが，これは数十〜百文書程度であると想定し，収集文書数を設定した．また，1 文書あたりは平均して 200〜400 文字である．工学系大学院生 10 人程度に FISH VIEW システムのクライアントプログラム，および操作方法や文書整理プロセスの例を記したオンラインマニュアルを配布して実験を依頼したところ，7 人から有効な回答が得られた．ここで，タスクは特に設定せず，各自の興味に任せ，自由にシステムを利用してもらった．

実験終了後，HTML 出力とアンケートの回答を提出してもらい，通信ログとあわせて結果の分析を行った．アンケート結果の一部について**表 2** に記す．各項目は 7 段階で評価してもらい（7 が最高），1〜3 を×，5〜7 を○とした．被験者が少数のため，これら 7 項目のうちで統計的に有意な水準で肯定的な結果といえるのは,「HTML 出力は便利か」，および「今後も使いたいか」の 2 項目だけであった．そこで以下では，図解作成作業，および視点情報を利用した支援機能（検索性能，リンク生成能力，可読性）に分け，通信ログの分析結果やユーザの感想などもふまえて考察する．

6.1 図解作成に関する考察

システムの使い勝手については，不満や改良案なども多少あったが,「FISH VIEW システムを今後も使ってみたいか」という質問に肯定的な回答が多かったように，図解を作成しながら文書を読み進め，整理していくというプロセスは認められたと考えている．ちなみにシステムの用途については，論文の分類・検索や新聞スクラップやビデオなどの整理，変わったところでは芸能人に関する記事，コメントを整理し，印象やアピールポイントを分析する，などの提案があった．また，図解を HTML 化して出力する機能は評価が高いことも確認された．

図解作成にかかった時間は 1 人あたり 30〜90 分程度であり，図解中に含まれるノード（文書）数は平均 17.14（最大 29，最少 5）であった．上述のように，文書整理作業として想定していた文書数は数十〜百程度であるとしたが，実験後の感想より，このような情報整理に関するツールの使用経験を持つユーザは皆無であったこと，および 1 人のユーザは 1，2 回しかシステムを使用しなかったことを考慮すれば，十分な量の文書を読み，整理を行うことができたと考えられる．

6.2 視点情報に関する考察

FISH VIEW システムが視点情報を用いて行う支援機能については，4章で示した以下の 3 機能について考察する．

(1) 局所図解から視点抽出・可読形式での提示
(2) 視点を反映した文書検索
(3) 文書間の関連性計算，リンク生成

6.2.1 局所図解から視点抽出・可読形式での提示

視点情報の提示に関する表 2 中のアンケート項目は,「視点情報は分かりやすいか」であるが，この項目に対する評価は他の項目と比較して低い傾向にあることが確認できる．本項目に関する最低評価は 2 であり，3 人のユーザがこの評価を下したが，彼らが作成したグループの中には，視点として抽出された意味グループ数が極端に少ない（0，1）ものや，その反対に多すぎる（10 以上）ものが存在することが確認された．これより，視点あたりの意味グループ数が，視点情報の可読性に影響を与えていることが推測でき，抽出意味グループ数を制御可能となるように，3.2 節のアルゴリズムを改良することが，今後の課題としてあげられる．

通信ログの分析より，ユーザによる図解作成の傾向として，映画のジャンルや，出演俳優，監督に基づいてグループ化を行うことが確認された．ユーザの意図した視点意味グループが正しく抽出できたと見なせるジャンルは次のとおりである．

- アクション（セイント，アベンジャーズ，007/トゥモロー・ネバー・ダイ，ホワイトハウスの陰謀など）
- サスペンス（スリーパーズ，評決の時，陪審員，密告）
- SF 映画（コンタクト，スターシップ・トゥルーパーズ，フラッド，インディペンデンス・デイ）
- 怪獣映画（モスラ，ガメラ 2，ゴジラ）

このうち，最初の 2 つは，文書中に「アクション」や「サスペンス」という単語が含まれていたため，対応する意味グループが正しく抽出，提示できていた．こ

表 3 怪獣映画に関して抽出された視点意味グループ
Table 3 Semantic groups extracted as viewpoint of a monster film.

説明	所属単語
軍事組織	自衛隊，海軍，軍団，人民解放軍，部隊
空中戦	空中戦
神話や伝説や人々の心の中でのみ存在する疑似人間や疑似生物	キャラクター，ヒーロー，ヒロイン，主人公，主役，竜，ドラゴン，怪獣，人魚，悪魔，宇宙人，魔女，神，天使

れに対し SF 映画に関しては，ジャンルそのものを表す単語は存在しなかったが，内容に深くかかわっている単語「宇宙，地球，日，彗星，隕石，溶岩，サンド，パール，一石」が 1 つの概念「自然物」としてまとめられ，抽出されていた．最も興味深く，かつ Fisheye マッチングの能力を示したと考えられるのが，怪獣映画に関する視点意味グループである（**表 3**）．後者 2 つのジャンルに関しては，従来のベクトル空間モデルで扱うことは困難であり，Fisheye マッチングの利点が生かされたものと考えている．反対に，視点意味グループの抽出がうまくいかなかった例としては，「コメディ」に関する映画のグループや，出演俳優，監督に基づくグループがあげられる．俳優や監督の名前といった固有名詞は，静的な電子辞書を用いる場合に対応が困難なものであり，統計的手法などを組み合わせ，辞書を補強，整備する必要があると考えている．

今回の実験で，視点情報の編集を行ったユーザは少なかった（7 人中 4 人）が，なかには我々の想定以上に視点編集機能を活用したユーザも存在した．このユーザは，サスペンス映画に関するグループに，「暴力，犯罪，ミス，奇行，死」などの単語を含む意味グループ「所為」を視点として追加したり，表 3 にも示されている意味グループ「神話や伝説や ...」の上位/下位にあたる意味グループを検索し，入換えを試みるといった，我々の想定していた編集作業のほかに，「意味グループ検索中に，当初の視点とは関係ない意味グループに興味が移り，図解を構築し直す」といった，「文書空間のブラウジング」ならぬ「視点情報空間のブラウジング」とでもいうべき活動も行っていた．すなわち，タランティーノ監督の作品を含むグループを作成中に，誤って検出されていた単語「ティー」のために抽出された，視点とは関係のない意味グループ「嗜好品である飲物」に対し，これを単に削除するのではなく，新たな視点構成要素を思いつき，「カフェ」という単語をキーとして，関連意味グループの検索を行い，視点の修正・変更を試みたのである．このようなユーザは，被験者中ただ 1 人のみであったが，概念体系を背景知識として用い，可読形式で視点情報を提供する我々の提案手法の可能性を感じさせるものであり，システムの利用経験を重ねていけば，他のユーザも同様な活動を行うようになることが期待できる．

6.2.2 視点を反映した文書検索

表 2 の項目「文書検索結果は適切か」は，ほかに比べて評価が高い傾向にあるものの，統計的に有意な水準には達していない．しかし，実験後の感想として，便利な機能の 1 つに「視点に基づく検索機能」をあげるユーザも多く，非常に有効な支援が行えたことが確認できた．具体的には，上述の視点意味グループが正しく抽出できた場合，関連文書の検索も有効に行われていた．視点意味グループの抽出に失敗した場合であっても，たとえば「コメディ」に関する文書の検索が有効に行えた，との報告もユーザより得られている．

また，視点を用いた支援全般に関する評価項目「視点の自動抽出は便利か」に対して 4 以上の評価をしたユーザ（5 人）と，「文書検索結果は適切か」で 4 以上の評価をしたユーザは一致していた．これより，視点情報に基づく 3 つの支援機能のうち，文書の検索作業は，文書整理作業全体の主要部分を占めている，と推測できる．

6.2.3 文書間の関連性計算，リンク生成

これについては，関係を指摘する手段としてリンクを用いることに対する評価，およびリンクによって指摘された文書間の関係がユーザにとって納得のいくものであったか，に分けて評価を試みた．前者に対応するアンケート項目は「リンクは役立ったか」であり，後者は「リンクは適切か」である．両項目に対する評価差として，有効な回答を行ったユーザ全員が「リンクは役立ったか」を，「リンクが適切か」より高く評価していたため，統計的に有意な差があったということができる．通信ログの分析より，リンクは大量に生成される傾向があり，これにより図解が繁雑になることが，リンクが不適切であるという評価につながったと考えている．FISH VIEW システムでは，Main Window 下部にあるスクロールバーを操作し，リンク生成に関する閾値を調整することによってリンク数を手動で適切に調整することを想定していたが，この機能を用いたユーザは存在しなかった．この点は反省すべきであり，ある範囲内に収まる数のリンクを自動生成するなどの改良が必要と考えている．

7. おわりに

本稿では，大量情報が容易に入手可能となった現代の我々にとって必要不可欠な，個人レベルでの情報活

用を目的として，文書整理プロセスの提案および，この過程を支援する FISH VIEW システムについて述べた．Fisheye マッチングは高い文書検索能力だけでなく，視点抽出・外化能力も備えた枠組みであり，個人レベルでの視点を反映した文書整理プロセスを支援するための基盤技術として適していると考えられる．視点情報の分かりやすさなど，改善すべき点は多く存在するが，FISH VIEW システムの有効性，将来性については実際のユーザによる評価実験により認められたと考えている．

評価実験の結果からも分かるとおり，本研究で提案したような，個人レベルでの情報活用を支援するシステムの必要性，将来性については多くのユーザの認めるところである．しかしその反面，現状における認知度，普及度は低く，インターネットに代表される膨大な情報空間の活用は，低機能な WWW ブラウザを用いたアクセスなどによる，情報収集段階にとどまっているのが現実である．現行の WWW ブラウザ並みの使いやすさ，親しみやすさと支援能力とを両立させることが，知的活動全般におけるシームレスな支援を実現するうえで重要であり，今後の課題であろう．

謝辞 本研究の一部は，日本学術振興会未来開拓学術研究推進事業研究プロジェクト「生物的適応システム」の支援のもとに行われました．記して，感謝いたします．

参 考 文 献

1) Dumais, S.T., Furnas, G.W. and Landauer, T.K.: Using Latent Semantic Analysis to Improve Access to Textual Information, *CHI'88, Conf. on Human Factors in Computing*, pp.281–285 (1988).
2) 藤崎博也ほか：キー概念に基づく情報検索システム方式の高度化（2)—キーワードの同表記異義の処理，第 57 回情報処理学会全国大会論文集，pp.3-239–240 (1998).
3) 幡鎌 博，津田 宏，益岡竜介：ナレッジマネジメントへむけて—知識検索・整理および基盤技術，人工知能学会誌，Vol.13, No.6, pp.912–919 (1998).
4) 岩爪道昭，武田英明，西田豊明：弱構造化オントロジーを用いたインターネットからの情報獲得，信学技報，AI95-32, pp.79–86 (1995).
5) 川喜田二郎：発想法，中公新書 (1967).
6) 糀谷和人，前田晴美，西田豊明：弱構造知識メディアを用いた情報ベース構築支援，信学技報，AI95-30, pp.63–70 (1995).
7) 國藤 進：発想支援システムの研究開発動向とその課題，人工知能学会誌，Vol.8, No.5, pp.552–558 (1993).
8) Mock, K.J.: Hybrid Hill-Climbing and Knowledge-Based Methods for Intelligent News Filtering, *13th Nat'l Conf. on Artificial Intelligence* (*AAAI-96*), Vol.1, pp.48–53 (1996).
9) 長尾 真：自然言語処理，chapter 11，岩波書店 (1996).
10) Nishida, T.: The Knowledgeable Community: Towards Knowledge Level Communication, *Int'l Forum on Frontier of Telecommunication Tech.* (1995).
11) 大見嘉弘ほか：インターネット上の情報を利用できるカード操作ツール PAN-WWW，情報処理学会論文誌，Vol.37, No.1, pp.154–162 (1996).
12) 高間康史，石塚 満：Fish Eye マッチング：概念体系を利用した視点抽出に基づく文書整理支援機能，人工知能学会誌，Vol.14, No.1, pp.93–101 (1999).
13) Torrance, M.C.: Active Notebook: A Personal and Group Productivity Tool for Managing Information, *AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval* (1995).

(平成 11 年 4 月 20 日受付)
(平成 12 年 5 月 11 日採録)

高間 康史（正会員）

1994 年東京大学工学部電子工学科卒業．1999 年同大学院博士課程修了．同年，東京工業大学大学院総合理工学研究科助手．現在に至る．博士（工学）．情報検索や感性情報処理，知的インタフェースの研究に従事．主要著書は「図解 人工生命を見る」（同文書院）．人工知能学会，日本ファジィ学会各会員．

石塚 満（正会員）

1971 年東京大学工学部電子工学科卒業．1976 年同大学院博士課程修了．工学博士．同年 NTT 入社，横須賀研究所．1978 年東京大学生産技術研究所助教授，教授を経て，1992 年より工学部電子情報工学科教授．研究分野は人工知能，知識処理，マルチモーダル擬人化インタフェース，知能的ネットワーク化情報環境．IEEE，AAAI，電子情報通信学会，人工知能学会，映像情報メディア学会，画像電子学会各会員．