

文書画像構造解析のための知識ベースの一構成法

黄瀬 浩 一† 山岡 正輝**
馬場 口 登†† 手塚 慶 一††

文書画像の構造解析とは、文書画像のレイアウト構造を解析し、得られた各部分領域を文書の論理構造に対応付ける処理である。その一手法に、文書のレイアウト構造と論理構造の関連性を知識ベースに蓄え、利用するものがある。知識ベース型構成をもつ構造解析システムでは、対象文書を変更する際、知識ベースの交換により対処するため、システムの適用性は、知識ベースの表現能力、記述性・可読性に大きく左右される。そこで本論文では、これらを高レベルで実現することを目標に、文書モデルと呼ぶ知識ベースを提案する。文書モデルの主な特徴は、①知識を、レイアウト構造、論理構造、および各構造相互の対応規則の3組で表現することにより、レイアウト構造と論理構造の対応関係を宣言的に記述すること、②知識表現の枠組みとしてフレーム表現を用いることにより、レイアウト構造、論理構造の階層性を明示的に表現すること、③センタリング等を表すレイアウト述語を用いて知識を記号的、宣言的に記述すること、④フレーム間の上位下位関係を利用して、レイアウト構造の変動を明示的かつ簡潔に記述することにある。本論文では、レイアウト構造、論理構造の大幅に異なる文書として、名刺と論文タイトルページを取り上げ、記述例を検討する。また、処理結果の検討、および既存手法との比較から、文書モデルの有効性について考察する。

Organizing Knowledge-Base for Structure Analysis of Document Images

KOICHI KISE,[†] MASAKI YAMAOKA,^{††*} NOBORU BABAGUCHI^{††} and YOSHIKAZU TEZUKA^{††}

Systems for structure analysis of document images are often implemented as knowledge based systems. Applicability of such systems, therefore, largely depends on the expressiveness, descriptibility and readability of the knowledge. Focusing on these three points, we propose the knowledge-base called document model. The characteristics of document model are summarized as follows: 1. Document model is composed of the knowledge about layout structure, logical structure and correspondence rules between each other. 2. Frame representation is employed to describe the knowledge hierarchically and declaratively. 3. Layout predicates such as *centering* and *alignment* allow us to represent the knowledge symbolically. 4. IS_A relation between frames enables us to specify variation of layout structure in a simple and clear way. Document model is illustrated with examples of knowledge description for two different kinds of documents: visiting cards and titlepages of papers. From the discussion about analysis results and comparison with other methods, we verified the advantages of document model.

1. ま え が き

オフィス文書の効率的かつ高機能な通信を目指した ODA (Office Document Architecture)¹⁾の規格にも述べられているように、文書には論理構造 (logical structure)、レイアウト構造 (layout structure) と呼

ばれる2種類の構造が存在する。論理構造とは、文書のコンテンツ (content) に関する階層構造であり、“章”、“節”などの論理オブジェクトを要素として持つ。一方、レイアウト構造とは、文書画像の部分領域に関する階層構造であり、ブロック領域などのレイアウトオブジェクトを要素として持つ。

既存の文書に含まれる情報を、コンテンツに応じて柔軟に検索、通信するためには、文書画像から情報を抽出し、論理構造により構造化する必要がある²⁾。その一方策として、論理構造がレイアウト構造と密接な関係を持つことに着目し、文書画像から観測可能なレイアウト構造を解析することにより、論理構造を得る処理が考えられる。ここでは、このような処理を文書

† 大阪府立大学工学部電気工学科

Department of Electrical Engineering, College of Engineering, University of Osaka Prefecture

†† 大阪大学工学部通信工学科

Department of Communication Engineering, Faculty of Engineering, Osaka University

* 現在 NTT データ通信(株)

Presently with NTT Data Communications Systems Corp.

画像の構造解析と呼ぶことにする。具体的には、文書画像からレイアウトオブジェクトを抽出し、それらを論理オブジェクトに対応付ける処理といえる。

文書画像の構造解析システムを構築する場合、論理構造、レイアウト構造が対象文書に強く依存することは大きな問題となる。多種多様な文書に対して高い適用性 (applicability) を持つシステムを構築するためには、それらの構造に関する情報を知識として分離・蓄積し、利用するという知識ベース型の構成が有用となる³⁾。

知識ベース型の構造解析システムでは、対象文書を変更する際に、推論部を変更せず、知識ベースの交換により対処する。したがって、システムの適用性は、知識ベースがどれほど多数の文書に対応できるかに左右されるといっても過言ではない。特に、知識ベースにおける知識の表現能力 (expressiveness)、記述性・可読性 (describability・readability) の2点は重要となる。表現能力とは、多種多様な文書に対して知識を表現する能力である。これが不十分であれば、適用可能な文書の種類が制限される。また、記述性・可読性とは、人間による知識の記述しやすさ、読みやすさを意味し、知識ベースの構築、修正の容易性を決定付ける要因である。表現能力が十分であっても記述性・可読性が低ければ、知識ベースの構築、修正が困難となるため、事実上、システムの適用性は制限される。

システムの適用性の観点から、従来の代表的手法、すなわちルールを用いるもの⁴⁾、座標を用いるもの⁵⁾、領域分割オペレータの AND/OR 木表現を用いるもの⁶⁾等を概観すると、

- 1) 知識表現の主体がレイアウト構造にあり、構造化の基準となる論理構造が明示的に表現されないこと
 - 2) レイアウト構造の種々の変動に対して、表現能力、記述性・可読性が十分とはいえないこと
 - 3) 知識が座標や処理パラメータにより数値的に表現されるため、記述性・可読性に問題があること
 - 4) 知識が推論手続きと明確に分離されていないこと、すなわち知識が推論対象の性質のみを記述したのではなく、特定の推論制御に関する情報も含むものであること
- などの問題が指摘できる。

そこで本論文では、高い適用性を持つ知識ベース型の構造解析システムにおいて主要な役割を果たす知識ベース (文書モデルと呼ぶ) について議論する。文書モデルは、推論対象 (論理オブジェクトおよびレイ

アウトオブジェクト) に着目した知識表現をとり、その主たる特徴は以下のとおりである。

- 1) 知識をレイアウト構造、論理構造、および両者の対応規則の3組により表現すること
- 2) 知識の表現枠組みとしてフレーム表現を用いることにより、レイアウト構造、論理構造の階層性を明示的に表現すること
- 3) センタリングなどを表すレイアウト述語を用いて、知識を記号的、宣言的に記述すること
- 4) フレーム間の上位下位関係を利用して、レイアウト構造の変動を簡潔かつ明示的に記述すること

提案する文書モデルは、筆者らが名刺を対象に開発したもの⁷⁾に対して、表現能力、記述性・可読性の一層の向上を図り、拡張したものである。以下、本論文では、文書モデルの詳細を記述し、レイアウト構造、論理構造の大幅に異なる文書である論文タイトルページと名刺を対象に検討を加えると共に、既存の知識表現法と比較し、本手法の有効性について考察する。

2. 知識ベースに対する要件

一般に、文書画像の構造解析には、対象文書の論理構造、レイアウト構造、および両者の対応関係に関する知識が必要である。これらの知識は、文書の個別例 (インスタンス) に関するものではなく、論文誌など文書のクラスに対して妥当なものでなければならない。したがって知識ベースは、インスタンス間の差から生じる論理構造、レイアウト構造、対応関係の変動に対処する必要がある。ここでは、論理構造、レイアウト構造、対応関係の性質について述べると共に、各々を知識として表現するための要件を整理する。

2.1 論理構造

一般に、論理構造は、“章”、“節”などの文書の論理オブジェクトをノードとする木構造により表される。例えば、論文誌の論理構造は図1のとおりである。ここで、“著者名項目群”とは、著者の氏名、所属などの個人情報を表す論理オブジェクトである。

論理構造に関する知識を表現するとき、考慮すべき変動には、論理オブジェクトの個数変動がある。例えば、論文誌では“著者名”の数が不定である。また、“キーワードリスト”の有無など、論理オブジェクトの有無にかかわる変動も個数変動と見なし得る。

以上から、論理構造の知識表現に対する要件には、木構造が表現可能なこと、論理オブジェクトの個数変動を表現可能なことが挙げられる。

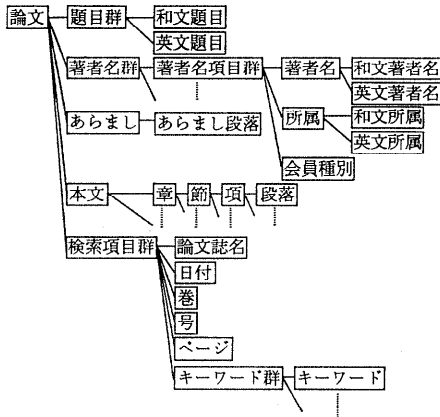


図 1 論文誌の論理構造
Fig. 1 Logical structure of papers.

なお、解析対象に対して知識ベースがカバーする範囲を定めること、換言すれば、どの対象に対して共通の知識ベースを用いるか、また知識ベースを交換するかを定めることは、システム設計上の重要なポイントとなる。我々は、論理オブジェクトの個数変動を許容して表現できる文書のクラスを、一つの知識ベースがカバーする範囲と考えている。上述の例では、作成される知識ベースは各論文誌に個別のものではなく、様々な論文誌に共通のものとなる。ただし、名刺と論文誌など、個数変動により表現不可能な論理構造を持つ文書は別クラスと考え、個別に知識ベースを作成する。

2.2 レイアウト構造

レイアウトの観点から捉えると、文書画像は図2に示す文字領域、文字列領域、ブロック領域の3種類の矩形領域から構成されるといえる。レイアウト構造に関する知識を表現するときに着目すべき領域は、文字領域を除く2種類である。本論文では、これらをレイア

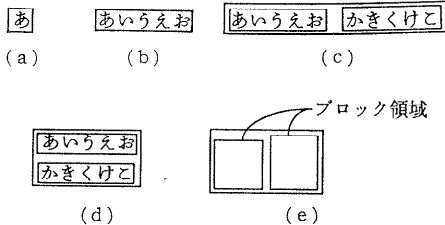


図 2 レイアウトオブジェクト
(a)文字領域, (b), (c)文字列領域,
(d), (e)ブロック領域
Fig. 2 Layout objects.
(a)Character region, (b), (c)Textline region,
(d), (e)Block region.

ウトオブジェクトと呼ぶ。また、図2からもわかるように、縦方向または横方向に並ぶレイアウトオブジェクトを子、それらを包含するレイアウトオブジェクトを親とすることにより木構造が定められ、この木構造をレイアウト構造と呼ぶ。

レイアウト構造に関する知識を表現するとき、考慮すべき変動には、以下の3種類が考えられる。

- 1) 形状変動 文字列領域の大きさなど、個々のレイアウトオブジェクトの形状に影響を及ぼすもの
- 2) 個数変動 段組数、文字列数の変動など、レイアウトオブジェクトの個数に影響を及ぼすもの
- 3) 関係変動 レイアウトオブジェクトの配置に関する変動など、レイアウトオブジェクト間の関係に影響を及ぼすもの

以上から、知識表現への要件としては、論理構造の場合と同様に、木構造を表現可能なことが挙げられる。さらに、知識の記述性・可読性を損なわずに変動を表現するためには、次の2点を考慮する必要がある。第一は、変動のない部分、あるいは変動とは無関係な部分を同一記述とすることである。これにより簡潔な記述が実現できる。第二は、逆に変動部分を明示的に記述することである。このとき、形状変動、関係変動については、変動の程度が画像の絶対的物理量に依存しないように記述すべきである。さらに、関係変動については、レイアウト構造の木において、どの部分木にどのような変動を生じるかを記述する必要がある。

2.3 対応関係

論理オブジェクトとレイアウトオブジェクトの対応関係は、次の三つの場合に分類できる。

- 1) レイアウトオブジェクトと論理オブジェクトが1対1に対応する
- 2) 対応する相手のオブジェクトが存在しない
- 3) 対応する相手のオブジェクトは存在するが、対応関係に変動がある

1) は、最も単純な場合である。例えば、論文誌の論理オブジェクト“あらまし”は、ブロック領域に対応する

2) の具体例としては、章、節、段落などの論理オブジェクト、あるいは段組を表すブロック領域や文字列領域などのレイアウトオブジェクトを挙げることができる。一般に、段組や文字領域には、どのような順序で読まれるかという論理的つながりがある。このことから、章などの論理オブジェクトは、論理的つなが

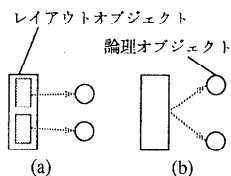


図3 対応関係の変動

Fig. 3 Variations of correspondence relation.

りにより順序付けられたレイアウトオブジェクトから合成されるといえる。したがって、対応関係に関する知識には、論理的つながりに基づいて論理オブジェクトを合成するための規則を記述できなければならない。

3) の場合は、図3に示す2種類に分類できる。図3(a)は、あるレイアウトオブジェクトの領域内に存在する論理オブジェクトは既知であるが、どのようにレイアウトされるかが不定な場合を指す。一方、図3(b)は存在するレイアウトオブジェクトは既知であるが、どの論理オブジェクトに対応するかが不定な場合を指す。したがって、対応関係を知識として表現するには、このような変動を表現する必要がある。

次章では、以上の要件を考慮した知識ベースとして文書モデルを提案し、その詳細について述べる。

3. 文書モデル

3.1 基本構成

文書モデルでは、論理構造、レイアウト構造における木構造を明示的に表現するため、知識表現の枠組みとしてフレーム表現を採用する。各フレームは、木構造のノード、すなわち論理オブジェクト、レイアウトオブジェクトに対応する。以後は、それぞれ論理オブジェクトフレーム (LOOF)、レイアウトオブジェクトフレーム (LAOF) と呼ぶ。

LOOF では、部分全体関係を用いて論理オブジェクト間の木構造を表現する。一方、LAOFでは、図4

に示すように部分全体関係、類似差異関係を通してレイアウト構造に関する知識を表現する。ここで、部分全体関係とは、レイアウトオブジェクトの木構造を規定するものである。また、類似差異関係とは、レイアウトオブジェクトの幾何的な差異を規定するものである。これらの関係は、絶対的な座標に基づいて記述されるのではなく、相対的な座標や位置関係に基づいて記述される。また、類似差異関係については、重複した記述を避けるため、LAOF から独立させ、類似差異フレーム (SIMF) を設けて記述する。

対応関係は、レイアウトオブジェクトから論理オブジェクトを導くルール (対応規則と呼ぶ) として記述する。

変動の記述については、次のような方針をとる。個々のオブジェクトに関する変動、すなわち論理オブジェクトの個数変動、レイアウトオブジェクトの個数変動、形状変動については、フレームごとに記述する。また、レイアウト構造の関係変動については、SIMF を複数設けて記述する。

文書モデルでは、これらの変動を簡潔に表現するため、フレーム表現における上位下位関係を用いる。一般にフレーム表現では、概念の上位下位関係に属性の継承という役割をもたせ、知識記述の重複を大幅に低減している。そこで、変動に依存しない共通特徴を上位クラスのフレームに記述し、個別特徴を下位クラスのフレームに記述することにより重複を回避する。

3.2 論理構造に関する知識

LOOF の構成を表1に示す。ここで、subpart スロットは、論理オブジェクトの部分全体関係を表現するためのものである。number_of スロットでは、論理オブジェクトの個数を区間値 $[m, n]$ により記述する。区間値は、部分全体関係で結ばれた論理オブジェクト1個に対して、子の論理オブジェクトが存在する個数を表す。なお、個数の上限が不定な場合には、 n に ∞ を用いる。

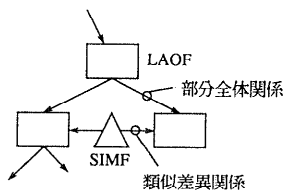


図4 レイアウト構造の表現

Fig. 4 Representation of layout structure.

表1 論理オブジェクトフレーム (LOOF)
Table 1 Logical object frame (LOOF).

スロット	記述される情報
name	論理オブジェクト名
number_of	論理オブジェクトの個数
subpart	子となる論理オブジェクト名のリスト

3.3 レイアウト構造に関する知識

3.3.1 記述子

レイアウトには、揃え、センタリングなど記号的に記述可能な側面が多い。知識が可能な限り記号的に記述されると、記述性・可読性という観点からも有利である。一方、文書画像の構造解析には、例えばレイアウトオブジェクトの面積に関する制約など、記号化が適さない知識も必要な場合がある。

本論文では、これらの相反する要求を満たし、かつ統一的扱いが可能な記述子として、レイアウト記述子を提案する。レイアウト記述子は、矩形領域に関する特徴量を規定する特徴量記述子、センタリング等の記号的なレイアウトを規定するレイアウト述語からなる。レイアウト述語は特徴量記述子を用いて定義されるため、知識を利用する際には、統一的扱いが可能となる。

まず、特徴量記述子を定義する準備として、矩形領域の特徴量（矩形特徴量と呼ぶ）を定義する。ここで用いる矩形特徴量は、個々の矩形領域を規定するもの、および矩形領域間の幾何的關係を規定するものに大別される。以下、それぞれを順に述べる。

一般に矩形領域は、文書画像の左上部を原点とし、横方向に i 軸、縦方向に j 軸をとる座標系を用いて、左上の座標 (i_s, j_s) と右下の座標 (i_e, j_e) により表すことができる。個々の矩形領域を規定する特徴量としては、この座標を用いて定義される縦幅や面積等が考えられる。ところが、これらの特徴量は画像の絶対的物理量に依存するため、知識の記述には適さず、正規化が必要となる。本手法では、部分全体関係に基づいて正規化を施す。すなわち、記述対象となる矩形領域 r の特徴量を、親の矩形領域 p の特徴量により正規化する。表 2 に定義を示す。ここで、 W , H , A , C_i , C_j は、矩形領域の横幅、縦幅、面積、中心の i 座標、 j 座標を表し、添え字 p , r は親子の区別を示す。なお、矩形領域が正方形からどれほど離れているかを表す flatness は割合であるため、正規化は不要である。

一方、矩形領域間の幾何的關係を規定する特徴量には、正規化の要件に加え、幾何的な差異の程度を表現することが求められる。本手法では、そのような特徴量として表 3 に示すものを用いる。これらの特徴量は、 r_2 を基準として r_1 の差異の程度を表す。以後は、 r_1 をターゲット、 r_2 をベースと呼ぶ。なお、 W_{r_m} , H_{r_m} は、 r_1 , r_2 を統合した矩形領域 r_m の横幅、縦幅

表 2 個々の矩形領域を規定する特徴量
Table 2 Characteristic features for a rectangular region.

特徴量名	定	義
n_width	W_r/W_p	
n_height	H_r/H_p	
n_area	A_r/A_p	
n_icenter	$(C_{ir}-i_{sp})/W_p$	
n_jcenter	$(C_{jr}-j_{sp})/H_p$	
n_is	$(i_{sr}-i_{sp})/W_p$	
n_ie	$(i_{er}-i_{sp})/W_p$	
n_js	$(j_{sr}-j_{sp})/H_p$	
n_je	$(j_{er}-j_{sp})/H_p$	
flatness	$1-H_r/W_r$: if $W_r > H_r$, W_r/H_r-1 : otherwise	

表 3 矩形領域間の幾何的關係を規定する特徴量
Table 3 Characteristic features for geometric relationship between rectangular regions.

特徴量名	定	義
c_width	$(W_{r_1}-W_{r_2})/\max(W_{r_1}, W_{r_2})$	
c_height	$(H_{r_1}-H_{r_2})/\max(H_{r_1}, H_{r_2})$	
c_area	$(A_{r_1}-A_{r_2})/\max(A_{r_1}, A_{r_2})$	
c_is	$(i_{s1}-i_{s2})/W_{r_m}$	
c_ie	$(i_{e1}-i_{e2})/W_{r_m}$	
c_js	$(j_{s1}-j_{s2})/H_{r_m}$	
c_je	$(j_{e1}-j_{e2})/H_{r_m}$	
h_distance	$(\max(i_{s1}, i_{s2})-\min(i_{e1}, i_{e2})-1)/W_p$	
v_distance	$(\max(j_{s1}, j_{s2})-\min(j_{e1}, j_{e2})-1)/H_p$	

である。

(1) 特徴量記述子 矩形特徴量を記述する記述子として、特徴量記述子を設ける。特徴量記述子は、特徴量名と値の組により次のように記述される。

{〈特徴量名〉〈値〉}

ここで、〈値〉には、レイアウトオブジェクトの形状変動や関係変動を表現するため、区間値を用いる。区間値の上限、下限には、矩形特徴量の値域を適当に量子化した値を用いる。特徴量記述子では、対象となる矩形領域 (r と p , あるいは r_1 と r_2) が省略されているが、これは後述する LAOF の記述形式から、対象が一意に定まるためである。なお、特徴量記述子のうち、表 2 の特徴量を用いたものは LAOF に対し

表 4 レイアウト述語の一覧
Table 4 List of layout predicates.

フレーム	レイアウト述語
LAOF	h_centering, v_centering, r_flush, l_flush, h_flush, t_flush, b_flush, v_flush
SIMF	t_alignment, b_alignment, r_alignment, l_alignment, h_alignment, v_alignment, t_indent, b_indent, r_indent, l_indent

表 5 レイアウト述語の定義例
Table 5 Examples for definitions of layout predicates

レイアウト述語	定義
h_centering(r, p)	{n_icerter [11/27, 5/9]}
l_flush(r, p)	{n_is [0, 1/27]}
h_flush(r, p)	l_flush \wedge r_flush
r_alignment(r_1, r_2)	{c_ie [-5/27, 5/27]}
v_alignment(r_1, r_2)	l_alignment \wedge r_alignment
r_indent(r_1, r_2)	{c_ie [0, 1]} \wedge \neg r_alignment

て、表 3 の特徴量を用いたものは SIMF に対して使用される。実際の記述には、特徴量記述子から適切なものを選択して用いる。

(2) レイアウト述語 表 4 に、LAOF, SIMF の各フレームに対して使用されるレイアウト述語の一覧を示す。ここで、flush, centering, alignment, indent は、それぞれ、寄せ、センタリング、揃え、インデントを表す。また、接頭辞 t, b, l, r, h, v は、上, 下, 左, 右, 横, 縦を示す。レイアウト述語は、表 5 に示すように、特徴量記述子、他のレイアウト述語、あるいはそれらの否定の連言により定義される。ここで、記号 \wedge , \neg はそれぞれ、連言、否定を表す。なお、実際の記述では、特徴量記述子の場合と同様の理由から、レイアウト述語の引数を省略する。

3.3.2 記述形式

LAOF の構成を表 6 に示す。type スロットには、レイアウトオブジェクトの種類を記述する。記述子は、block (ブロック領域), textline (文字列領域) である。

number_of スロットには、レイアウトオブジェクトが連続して出現可能な個数を記述する。記述子には LOOF と同様、区間値を用いる。このスロットは、レ

表 6 レイアウトオブジェクトフレーム (LAOF)
Table 6 Layout object frame (LAOF).

スロット	記述される情報	継承
name	レイアウトオブジェクト名	禁止
type	レイアウトオブジェクトの種類	単純
number_of	レイアウトオブジェクトの個数	単純
order	レイアウトオブジェクトの出現順序	単純
layout_descriptor	レイアウト記述子のリスト	要素
part_of	親のレイアウトオブジェクト名	単純
subpart	子のレイアウトオブジェクト名のリスト	要素
orientation	子のレイアウトオブジェクトが並ぶ方向	単純
similarity	類似差異関係を記述するレイアウトオブジェクト名のリスト	要素
is_a	上位下位関係をなす上位クラスのレイアウトオブジェクト名	禁止
class	レイアウトオブジェクトの属するクラス名	禁止

イアウト構造に関する知識の簡潔かつ効果的な記述に不可欠である。例えば、論文誌には類似した形状の文字列領域が多数連続して現れる。number_of スロットを用いると、それらの特徴をまとめて記述することができる。また、個数が不定な場合にも記述可能となる。

order スロットには、レイアウトオブジェクトが親の内部で出現する順序を記述する。具体的には、レイアウトオブジェクトが親の内部で縦方向に並ぶ場合には上から数えた順序を記述し、横方向に並ぶ場合には左から数えた順序を記述する。

layout_descriptor スロットには、レイアウト記述子のリストを記述する。これは記述子の連言を意味する。

次に、残りのスロットにより記述される部分全体、類似差異、上位下位の関係記述について述べる。

(1) 部分全体関係 LAOF に、part_of, subpart, orientation の 3 スロットを設けて記述する。各々の役割は表 6 に示すとおりである。記述子には次のものを用いる。part_of, subpart スロットには、部分全体関係により結合されるレイアウトオブジェクト名を記述する。orientation スロットの記述子は、vertical (縦方向) または horizontal (横方向) である。

(2) 類似差異関係 SIMF の構成を表 7 に示す。レイアウト記述子のターゲット、ベースとなるレイア

表 7 類似差異フレーム (SIMF)
Table 7 Similarity frame (SIMF).

スロット	記述される情報	継承
name	SIMF 名	禁止
layout_descriptor	レイアウト記述子のリスト	要素
is_a	上位クラスの SIMF 名	禁止
class	SIMF の属するクラス名	禁止

ウトオブジェクトを、それぞれ A, B とすると、SIMF の name スロットには、(A, B) を記述する。なお、(A, B) の SIMF を作成する場合には、(B, A) の SIMF は作成しない。一方、LAOF では、similarity スロットに、比較する相手のレイアウトオブジェクト名を記述する。

layout_descriptor スロットには、LAOF と同様にレイアウト記述子のリストを記述する。すべてのレイアウト記述子を LAOF に記述せず、一部を SIMF を設けて記述する理由は、記述子の重複を避けるためである。例えば、二つのレイアウトオブジェクトの左端が揃っているとき、SIMF を用いると l_alignment というレイアウト述語を一つだけ記述すればよい。

なお、SIMF は、すべての LAOF 間に設ける必要はなく、レイアウトオブジェクト間の類似差異関係に記述すべき特徴があるものについてのみ、記述するものとする。

(3) 上位下位関係 上位下位関係では、LAOF, SIMF に一般化のクラス階層を設け、変動を簡潔に記述する。具体的には、is_a スロットを介してフレームを結合することによりクラス階層を表す。

例えば、電子情報通信学会論文誌、情報処理学会論文誌の 2 種類の論文タイトルページを対象とすると、知識記述は図 5 のようになる。ここでは簡単のため、レイアウト述語のみを示している。図 5 (a) は、2 種類の論文誌に共通の知識記述であることから、より一般的な特徴を表現している。また、図 5 (b), (c) には、各々に固有の知識記述であるため、個別の特徴を容易に読み取ることができ。これらは、論文誌という上位クラスに対する下位クラスの知識記述と捉えられる。各クラスの名称は、LAOF の class スロットに記述する。

スロット値の継承方法は、表 6, 7 の継承の欄に示すように、禁止、単純、要素の 3 種類に分類される。継承禁止スロットでは、上位クラスのフレームからスロット値を継承しない。単純継承スロットでは、下位

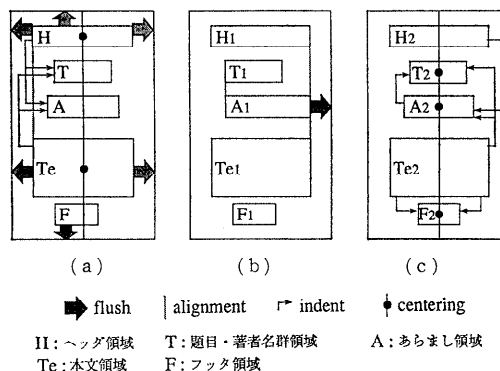


図 5 論文タイトルページのレイアウト構造
(a)論文誌, (b)電子情報通信学会論文誌,
(c)情報処理学会論文誌

Fig. 5 Layout structure of paper titlepages.
(a)Transactions, (b)Transactions of the Institute of Electronics, Information and Communication Engineers, (c)Transactions of Information Processing Society of Japan.

クラスにスロット値が記述されていないとき、上位クラスのスロット値を継承する。また、要素継承スロットでは、下位クラスに記述されたリストの要素について、重複しないものを上位クラスから継承する。

さて、表現能力の観点から上位下位関係の意義を考えると、次のことがいえる。各 LAOF, SIMF に記述されるレイアウト記述子は連言に制限されている。これは、上位下位関係を用いない場合、対象文書のクラスに共通するレイアウト記述子のみが記述可能なことを意味する。一方、上位下位関係を用いると、個別のレイアウト記述子が記述可能となり、表現能力が増す。さらに、情報処理学会論文誌などを表す下位クラスの記述は、各文書画像に対していずれか一つが有効である。したがって、文書画像の構造解析の際に、異なる下位クラスに対する知識を混合して用いることは避けなければならない。文書モデルでは、class スロットに記述される下位クラス名を参照することにより、知識の混合を防ぐ。

3.4 対応規則

一般に対応規則は、次のルール形式を持つ。

$$X_1, \dots, X_m \Rightarrow Y_1, Y_2, \dots, Y_n$$

ここで、 X_i はレイアウトオブジェクト名、 Y_j は論理オブジェクト名を表す。このルールは、論理的つながりを持つレイアウトオブジェクト X_1, \dots, X_m が、論理オブジェクト Y_1, \dots, Y_n を含む可能性があることを示す。ただし、これらのすべてが含まれる必要はない。

実際に知識を記述する際には、2.3節に述べた1)~3)にそれぞれ対応する以下の場合が重要となる。

1) $m=n=1$ のとき：レイアウトオブジェクトと論理オブジェクトが1対1に対応することを示す。

2) $m>1, n=1$ のとき：論理的つながりにより順序付けられた m 個のレイアウトオブジェクトから、論理オブジェクト Y_i が合成されることを示す。単純な例として、ここでは図6(a)に示す文字列領域から段落を合成する規則を取り上げる。LAOFには、段落の先頭にあるインデントされた文字列領域 ST, 段落の途中にある文字列領域 IT, 段落の最後にある文字列領域 ET の3種類が記述される。一般に段落は, ST, IT, ..., IT, ET のように、論理的つながりを持つ文字列領域から合成される。したがって、対応規則は図6(b)のように記述できる。ここで、*は、レイアウトオブジェクトの0回以上の繰り返しを表す。

レイアウトオブジェクトの論理的つながりは、orientation スロットの値から容易に導出できる。値が vertical ならば、子のレイアウトオブジェクトは上から下に論理的つながりを持つ。一方、horizontal の場合には、左から右、右から左の両方の可能性がある。しかし、これは文書のクラスに関して一定であるため、グローバルに規定することができる。

3) $m=1, n>1$ のとき：レイアウトオブジェクト X_i が、 n 個の論理オブジェクトを含む可能性があることを示す。例として、ここでは図7に示す名刺の住所部分における文字列領域 (T4 など) を考える。これらの文字列領域がどの論理オブジェクトに対応するかについては、ほとんど規則性がない。また、住所と郵便番号が同一の文字列領域に記述されることもある。そこで本手法では、

T4=>郵便番号, 住所, 電話番号, ...

のような形式により、可能な論理オブジェクトを列挙する。これらから適切なものを選択し、各々の領域を

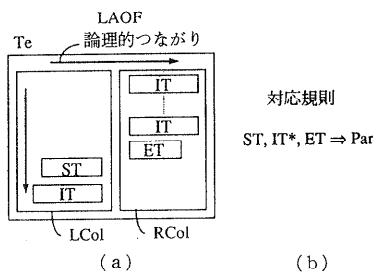


図6 論理オブジェクトを合成するための対応規則
Fig. 6 Correspondence rules to generate logical objects.

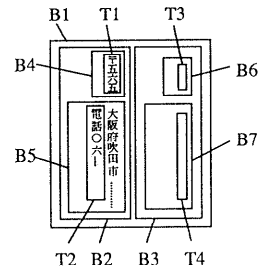


図7 名刺の住所部分のレイアウトオブジェクト
Fig. 7 Layout objects in address parts of visiting cards.

抽出するためには、文字認識の後、別種の知識を適用する必要がある⁹⁾。

4. 記述例と検討

まず、論文タイトルページ、名刺に対する記述例を示し、次に他手法との比較検討を行う。

4.1 記述例

4.1.1 論文タイトルページ

図5に対する記述例を図8に示す。ここでは、変動の表現と属性の継承に焦点をあて、検討する。

あらましに対応するレイアウトオブジェクト A1, A2 に共通の特徴は、上位クラスのAに記述される。A1, A2 に記述されていないスロット、すなわち type, number_of, orientation, order, subpart の値はAから継承する。また、A1, A2 の子となるレイアウトオブジェクトは同一構造を持つため、それぞれの subpart スロットに記述されず、Aの subpart スロットに記述される。

layout_descriptor スロットは要素継承のスロットであるため、各レイアウト記述子について値を継承する。例えば、A1 の場合について継承を行うと、

{n_height [0, 2/9], r_flush)

となる。SIMF についても、同様にスロット値を継承する。なお、v_distance など下位クラスに固有の特徴量は、それぞれのフレームに記述される。

以上のように、文書モデルでは、変動に影響を受けるフレームのスロットを上位クラスから抜き出し、下位クラスフレームに差分のみを記述する。このことにより、変動の簡潔かつ明示的な記述が可能となる。

4.1.2 名刺

図7に示す領域を例に取り上げる。ここで、B1 は名刺の住所表記に対応するブロック領域である。B2 は、住所表記が2種類存在する場合に必要となる。

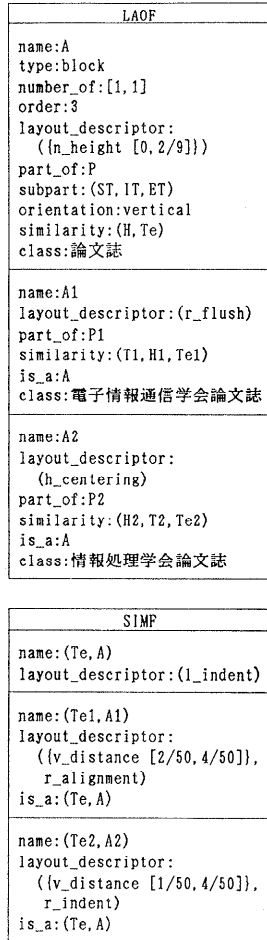


図 8 論文フロントページに対する記述例
 Fig. 8 Knowledge description for paper titlepages.

LAOF の構造を図 9 に、太線により示されるフレームの具体的な記述を図 10 に示す。ここでは、特に上位下位関係の異なる利用法、および LOOF の役割について述べる。

上位下位関係を用いると、変動の場合に限らず、知識記述の冗長性を削減することができる。これは、文書のクラスとは別に、個々のレイアウトオブジェクトにクラス階層を設けることにより実現される。例えば、B4 と B6 の場合、number_of スロットの値は異なるものの、内部に存在する文字列領域 (T1, T3) のレイアウトは共通 (HT として記述) である。したがって、上位クラスに HB を設け、下位クラスには個別特徴のみを記述する。

LOOF の役割としては次のことがいえる。対応規

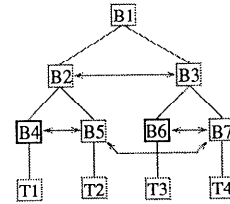


図 9 名刺の住所部分に対する LAOF の構造
 Fig. 9 Structure of LAOFs for address parts of visiting cards.

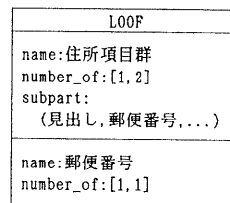
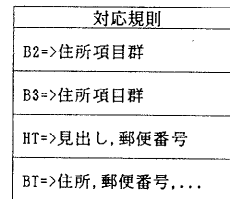
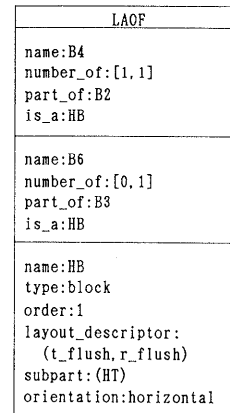


図 10 名刺に対する記述例
 Fig. 10 Knowledge description for visiting cards.

則には、HT, BT の二つの文字列領域に“郵便番号”が存在可能なことが記述されている。しかし、“郵便番号”は一つの住所表記 (“住所項目群”という論理オブジェクトとして記述) に唯一である。LOOF を用いると、このような知識を表現できる。この例の場合、“郵便番号”に対する LOOF の number_of スロットに [1,1] を記述すると、“郵便番号”の個数

が親（“住所項目群”）に対して1個であることを表している。

以上の記述例から、文書モデルは、論文タイトルページ、名刺という全く異質な文書に対しても、知識を簡潔に記述可能であるといえる。

4.2 上下位関係の表現能力

上位下位関係の表現能力を検証するため、電子情報通信学会論文誌、および情報処理学会論文誌のタイトルページ（標題の頁）を対象に実験を行った。実験対象は30サンプル（各15サンプル）である。このうち、20サンプル（各10）をS1、残りの10サンプル（各5）をS2と呼ぶ。実験の概略を以下に示す。

(1) 文書モデル作成 サンプル集合S1を参照し、M1、M2の2種類の文書モデルを作成した。M1は上位下位関係を用いた文書モデル、M2は上位下位関係を用いない文書モデルである。3.3.2項の(3)に述べたように、M2ではスロットの記述子が連言に制限される。なお、特徴量記述子には、 n_width , n_height , $h_distance$, $v_distance$ の4種類の特徴量を用いた。値域の量子化幅は、 n_width , n_height に対しては1/9または1/27、他に対しては1/10または1/50である。

(2) 構造解析⁹⁾ 実験結果について述べる前に、まず解析手法の概略を示す。本手法は、レイアウト構造の根ノードから葉ノードへと、段階的にレイアウトオブジェクトの領域を抽出するものである。各段階では、先に抽出された親のレイアウトオブジェクトの領域を、文書モデルの記述を満たすように分割し、子のレイアウトオブジェクトの領域を抽出する。

親から子を抽出する手順を図11に示す。親は文書モデルに記述されたLAOFとすでに対応しているため、子の並ぶ方向はLAOFのorientationスロットから求められる。次に、子の並ぶ方向とは異なる方向に親の領域を投影し、いくつかの部分領域に分割する。図11では、子が縦方向に並ぶため、横方向に投影処理を施している。得られた部分領域は、子の領域の一部、あるいは全部であるため、関連するレイアウト記述子をすべて満たすように統合すれば、子を抽出することができる。抽出された子は、対応規則の適用により論理オブジェクトに対応付けられる。

レイアウト記述子は、特徴量の区間値を用いて定義されるため、子が一意に定まらない場合がある。このような場合、本手法では、得られた子を仮説として扱い、仮説を親としてさらに子を抽出していく。このよ

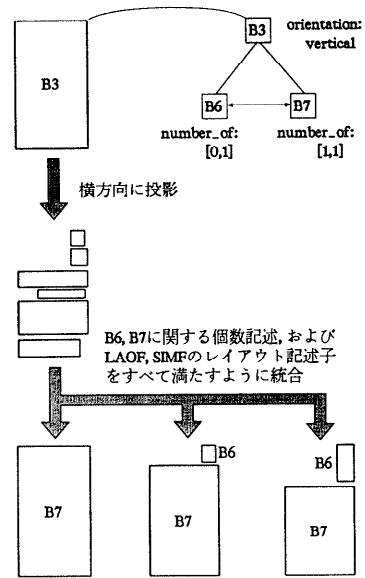


図 11 構造解析法

Fig. 11 The method of structure analysis.

うな処理形態をとると、仮説の組合せ的爆発が問題となる。本手法では、仮説間の依存関係、矛盾に基づく仮説管理機構¹⁰⁾を用いて、不適当な仮説を早期に除去することにより、組合せ的爆発を回避している。以上の処理により、文書モデルの記述を満たすレイアウトオブジェクトをすべて抽出する。

さて実験としては、サンプル集合S2を対象に構造解析を施した。結果を表8に示す。ここで、平均領域数とは文書モデルに記述されたレイアウトオブジェクト1個あたりに抽出された領域数、信頼度とは生成された領域に正解が含まれている割合、処理時間比とはM1による処理時間を1としたときの比を表す。なお、抽出対象となったレイアウトオブジェクトは合計779個である。

M1、M2のいずれを用いた場合にも、正解の領域はすべて抽出されているが、M1を用いたほうが、生成された領域は少数である。このことから、M1はM2に比べて、より詳細かつ妥当なレイアウト構造に

表 8 処理結果
Table 8 Processing results.

	M1	M2
平均領域数	1.02	1.09
信頼度	100%	100%
処理時間比	1	2.20

表 9 他手法との比較
Table 9 Comparison with other methods.

手 法	知識表現の枠組み				記 述 子			変 動	
	表現方法	表現単位	表現される知識 特徴・関係	複雑さ	表現方法	特 徴	了解性	表現方法	明示される変動
Niyogi ⁴⁾ (手法 a)	ルール	レイアウトオブジェクト	表現単位ごとの 特徴・関係	単純	述語	記号的	低い	確信度	なし
東野 ⁵⁾ (手法 b)	書式定義言語	レイアウトオブジェクト	領域分割手続き	中程度	座標の関数	数值的	低い	座標変動 if 文	形状変動, 関係 変動
駱 ⁶⁾ (手法 c)	AND/OR 木	領域分割オペレータ	領域分割手続き	単純	領域分割オペレータ	記号的	低い	OR ノード	関係変動
本手法	フレーム	レイアウトオブジェクト	木構造	複雑	レイアウト記述子	記号的・数值的	高い	number_of, is_a 表現	個数変動, 形状 変動, 関係変動

関する知識を表現していると考えられる。M2を用いた場合の処理時間が、M1を用いた場合に比べて2倍以上であることを考え合わせると、上位下位関係は、処理精度、処理効率の両面に寄与するものといえる。

4.3 他手法との比較

既存の知識ベース型手法として、Niyogiらの手法⁴⁾(手法 a)、東野らの手法⁵⁾(手法 b)、駱らの手法⁶⁾(手法 c)を取り上げ、知識の表現能力、および記述性・可読性の観点から本手法との比較を試みる。これらと比較対象とする理由は、各々が異なる知識表現法をもち、かつ代表的なことである。手法 a はルール表現、手法 b は書式定義言語、手法 c は領域分割オペレータの AND/OR 木表現に基づくものである。以下では、論理構造、レイアウト構造、対応関係に関する知識の表現について考察していく。

(1) 論理構造 論理構造については、手法 a～c のすべてが明示的な表現手段を持たず、レイアウトオブジェクトと論理オブジェクトの対応関係の記述に留まっている。この理由は、これらの手法が、文書画像の部分領域を論理オブジェクト名によりラベル付けすることを目標とするためである。しかし、文書データベース²⁾への自動入力などの種々の応用を考えると、文書画像の構造解析は、部分領域のラベル付けのみならず、各部分領域が論理構造においてどのような親子関係をもつかも明らかにすること、すなわち論理構造による構造化を図ることが望ましい。本手法では、このような観点から、論理構造をレイアウト構造とは独立な表現としている。また、論理構造を独立に表現することにより、知説の表現能力についても、4.1.2 項に述べた利点を得ることができる。

(2) レイアウト構造 手法 a～c は、すべてレイアウト構造を中心とした知識表現をとっている。したがって、レイアウト構造に関する知識表現の比較は特に

重要となる。ここでは、知識表現の枠組み、記述子、変動の3点から、各手法を詳細に比較する。まず、これまでの議論に基づき、知識の表現能力、記述性・可読性から、以上の3点に求められる具体的な条件を整理しておく。

1) 知識表現の枠組み レイアウトオブジェクトを単位とした知識表現であり、かつレイアウトの本質である木構造を明示的に表現できること

2) 記述子 可能な限り記号的であり、かつ人間にとって了解性*が高いこと

3) 変動 知識表現の枠組み、あるいは記述子が、文書のインスタンス間の変動(個数・形状・関係)を明示的に表現できること

比較の結果を表9にまとめる。本手法は、1)、3)の条件をすべて満たすように設計されているため、他手法に比べて知識表現の枠組みが複雑となっている。しかし、これは本手法の問題点を表すのではなく、むしろ記述される知識の量と質の差、すなわち知識ベースにおける表現能力、記述性・可読性の差を反映していると捉えるべきである。具体的にいうと、例えば個数変動を明示的に表現できなければ、レイアウトオブジェクトの個数が不定な場合に、知識を表現することが困難となる。また、変動を明示的に表現できなければ、知識の記述性・可読性が損なわれる。2)については、本手法では形状変動を特徴量記述子により数值的に表すため、記号的という点からは手法 a、c が優れている。しかし、了解性を考慮すると本手法の記述子はより適切であると考えられる。

(3) 対応関係 1対1対応の場合は全手法が表現可能であり、同等の記述性・可読性を持つ。論理オブジ

* 了解性が高いとは、センタリングやインデントなど、人間が文書を作成する際に意識する書式用語として直接記述できることを意味する。

エクトの合成を必要とする対応関係については、本手法、手法 b において表現可能である。ただし、手法 b では座標変換により知識を記述するのに対して、本手法では記号的なルールにより記述するという相違点がある。したがって、記述性・可読性は本手法が優れているといえる。変動については、本手法、手法 d において、同等の記述性・可読性のもとに表現可能である。

以上の議論を総合すると、本手法は従来手法に比べて知識の表現能力、記述性・可読性が高いといえる。

5. おわりに

文書画像の構造解析システムにおいて、適用性を高めるためには、知識ベースの表現能力、記述性・可読性を高いレベルで実現する必要があると観点から、本論文では、オブジェクトに着目した知識ベースとして文書モデルを提案した。文書モデルは、レイアウト構造と論理構造を別種のフレームに記述し、両者を対応規則を用いて結合するものであり、以下のような特徴を持つものである。

- 1) フレーム表現を用いることにより、オブジェクトに関する知識を階層的に記述可能である。
- 2) レイアウト述語を用いることにより、知識を記号的に記述可能である。
- 3) 上位下位関係、オブジェクトの個数記述、ルール形式の対応規則により、レイアウト構造、論理構造、対応関係の変動を簡潔に記述可能である。

今後の課題には、画像例、処理誤りから知識ベースを自動的に構築、修正すること¹¹⁾がある。なお、文書モデルを用いる構造解析システムの詳細については、稿を改めて報告する。

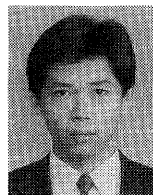
謝辞 筆者の一人黄瀬が日頃ご激励頂く、大阪府立大学の日下浩次助教、高松忍講師に感謝する。

参 考 文 献

- 1) 野口健一郎, 大谷 真: OSI の実現とその課題, 情報処理, Vol. 31, No. 9, pp. 1235-1244 (1990).
- 2) 田口和男, 坂下善彦: OA システムと文書データベース, 情報処理, Vol. 28, No. 6, pp. 721-729 (1987).
- 3) Kise, K., Yamada, K., Tanaka, N., Babaguchi, N. and Tezuka, Y.: Visiting Card Understanding System, *Proc. of 9th ICPR*, pp. 425-429 (1988).
- 4) Niyogi, D. and Srihari, S. N.: A Rule-based System for Document Understanding, *Proc. AAAI-86*, pp. 789-793 (1986).
- 5) 東野純一, 藤澤浩道, 中野康明, 江尻正員: 矩形領域の集合表現に基づく知識表現言語 FDL と文書画像理解への応用, 電子通信学会技術研究報告, PRU 86-31 (1986).
- 6) 駱 琴, 渡邊豊英, 吉田雄二, 稲垣康善, 斉藤隆夫: 知識ベースに基づいた図書目録カードの理解, 情報処理学会論文誌, Vol. 31, No. 12, pp. 1755-1767 (1990).
- 7) 山岡正輝, 黄瀬浩一, 馬場口登, 手塚慶一: 文書画像理解におけるドメイン知識記述の一手法, 電子情報通信学会技術研究報告, PRU 89-75 (1989).
- 8) Kise, K., Momota, K., Yamaoka, M., Sugiyama, J., Babaguchi, N. and Tezuka, Y.: Model Based Understanding of Document Images, *Proc. of IARP Workshop on Machine Vision Applications, MVA '90*, pp. 471-473 (1990).
- 9) 山岡正輝, 黄瀬浩一, 馬場口登, 手塚慶一: 知識ベース型文書構造解析システムの汎用性に関する一考察, 電子情報通信学会技術研究報告, PRU 90-119 (1991).
- 10) 黄瀬浩一, 馬場口登, 手塚慶一: 文書画像理解における推論と制御の一提案, 電子情報通信学会技術研究報告, PRU 89-76 (1989).
- 11) 黄瀬浩一, 馬場口登, 手塚慶一: 文書構造解析システムにおける知識ベースの構築・修正について, 平成 3 年電気関係学会関西支部連合大会, S 57 (1991).

(平成 4 年 2 月 6 日受付)
(平成 4 年 10 月 8 日採録)

黄瀬 浩一 (正会員)



昭和 38 年生。昭和 61 年大阪大学工学部通信工学科卒業。昭和 63 年同大学院前期課程修了。平成 2 年より大阪府立大学工学部助手。パターン認識、画像理解の研究に従事。工學博士。電子情報通信学会、人工知能学会各会員。

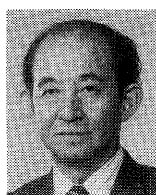
山岡 正輝 (正会員)



昭和 42 年生。平成元年大阪大学工学部通信工学科卒業。平成 3 年同大学院前期課程修了。現在 NTT データ通信(株)勤務。在学中、文書画像理解に関する研究に従事。電子情報通信学会、人工知能学会各会員。

**馬場口 登 (正会員)**

昭和32年生。昭和54年大阪大学工学部通信工学科卒業。昭和56年同大学院前期課程修了。昭和57年愛媛大学助手。昭和62年大阪大学助手。平成2年同大学工学部講師。現在に至る。工学博士。人工知能、パターン認識、画像処理の研究に従事。IEEE, 電子情報通信学会, 人工知能学会各会員。

**手塚 慶一 (正会員)**

昭和3年生。昭和26年大阪大学工学部通信工学科卒業。同大学院特別研究生。愛媛大学, 山口大学, 大阪大学助教授を経て, 大阪大学工学部教授。平成4年大阪大学名誉教授。現在関西大学工学部教授。データ通信, 通信ネットワーク, オートマトン, パターン処理, データベース, 知識処理などの研究に従事。工学博士。著書「電子計算機基礎論」, 「電子計算機システム論」, 「デジタル画像処理工学」など。IEEE, 電子情報通信学会, 電気学会, 人工知能学会各会員。