

文 書 作 成 時 に お け る 論 理 構 造 生 成  
支 援 法 の 検 討

小林 直樹

N T T ヒューマンインタフェース研究所

文書作成の支援について、特に集まった情報の統合、論理化する過程に注目し、その過程における論理構造生成の過程のモデル化ならびに支援方式について検討を行った。その結果、論理構造作成時には、統合化、関係付け、論理展開付け、内容補完の四つの過程を繰り返し行うこと、また各構造を階層的ネットワークで記述可能なことを示した。

さらに、統合化の自動的支援方式として、数量化理論を用いた方式を提案し、特許文書などにおける適用可能性について示した。

A Support System for Constructing Logical Structure in Document Preparation

Naoki KOBAYASHI

N T T Human Interface Laboratories

1-2356 Take Yokosuka Kanagawa , 238-03 Japan

A process model for constructing document logical structure and new idea support system for document preparation is proposed. The model has four different mode that is hierarchical structure making mode, relation making mode , logical structure making mode and making up contents mode. New grouping algorithm using dual scaling for constructing hiarchical structure is also proposed and a simulation get good result.

## 1. はじめに

近年ワードプロセッサやDTP（デスクトップパブリッシング）の発達により、文書作成、編集の分野では、高度な支援機能が実現されるようになってきた。中でも、WISYWIGによるレイアウト処理やアウトラインフォントの採用など、物理的編集、出力機能の発展は目覚ましい。

一方、文書の作成時の支援、例えば、文書の論理展開を決めるとか、論旨の重点を絞るといった行為に対しての支援機能は、現在アウトラインプロセッサ<sup>1)</sup>が実現されているだけで、十分とはいえない状況にある。この文書作成時の支援方法を考えるにあたっては、従来のシステム機能の高度化だけでなく、人間の文書生成の過程に基づいた支援方法の検討が必要となってくる。この点に着目して、文書の処理フローや人間の処理フローを考えたシステム化についての研究が数多くなされている。<sup>2) 3)</sup>本稿では、この中でも文書論理構造を生成する過程に着目し、その支援方法について検討した結果を報告する。

## 2. 文書作成過程の支援方法

報告書、発表原稿など、オフィス等で使用される文書を作成する場合、大きく分けて三つのステップに分けることができる。

- ①書く内容を検討する段階（STEP1）
  - ・必要な情報の収集
  - ・実際に使用する情報の選択
- ②実際の文章の内容を具体化する段階（STEP2）
  - ・情報の整理、位置づけ
  - ・書く内容の論理づけ、論理展開の仕方
  - ・章だて
  - ・内容の記述（文章、図面の作成）
- ③文書の形式を整える段階（STEP3）
  - ・レイアウトの仕方
  - ・テキスト、図面の清書

この文書作成過程とそこで扱われる情報の構造を表したものが図1である。（STEP1）では、扱われる情報はいわゆる文書といった体裁はとっておらず、雑多な情報の集合といえる。

（STEP2）は、この情報に対して、情報間の関係あるいは、重要度などを与え、文書の論理構造を生成する段階といえる。ここで、文書の論理構造とは、図2の下部に示すように文書の持つ内容部（文、段落、図面等）間に、題名、章、節といった章だての関係を与え、階層的構造として記述したものである。

一方（STEP3）は、図2上部のような文書のレイアウト構造を生成する段階であり、ページ、枠、区画といった、階層的な割付位置によって、文書の内容部の構造化を図っている。

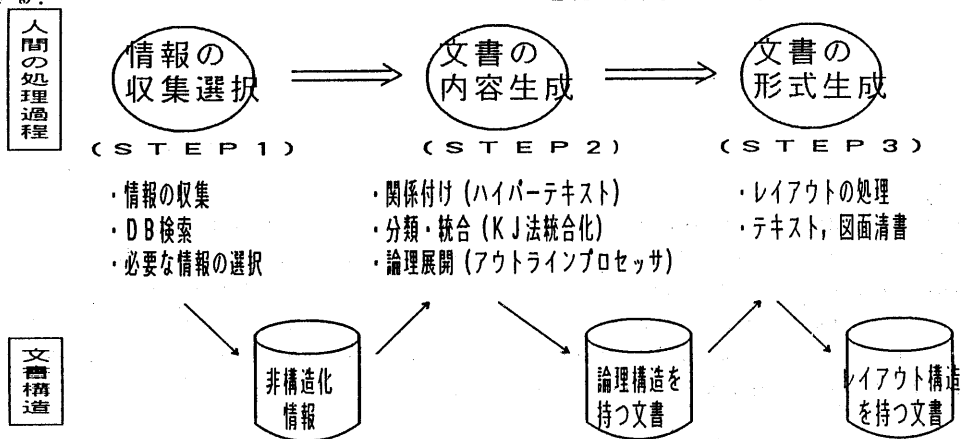


図1 文書作成の処理過程

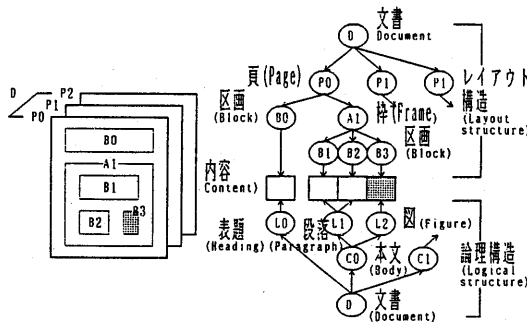


図2. ODAの論理構造とレイアウト構造

従来のワードプロセッサやDTPは(STEP3)の機能,中でも,非常に高度化したレイアウト構造の編集やテキスト,図面を清書する機能をサポートして,(STEP1)(STEP2)の作業をその中で行わせていたと考えることができる。

一方,(STEP2)を支援する機能としてアウトラインプロセッサがあるが,その処理はこの中の「章だて」機能にとどまっているため,(STEP2)全体を十分支援している状態とはなっていない。

本稿では,この論理構造を生成する(STEP2)について,扱われる情報の構造,その処理過程をモデル化するとともに,その支援機能について提案する。

### 3. 文書論理構造の作成

#### 3.1 文書の構造モデル

まず,(STEP2)で扱われる情報の構造について考察する。

完成された文書構造は,ODA(Office Document Architecture)<sup>4)</sup>で検討されているように,物理的な割付からみたレイアウト構造と文書の論理展開を記述している論理構造という階層構造を持っている。(図2)このうち(STEP2)の結果得られる文書は,文書の論理構造のみを持つ文書である。これは,題名,章,節,段落といった論理的階層構造と,これらの情報に対し付与されたシーケンシャルオーダーと呼ばれる順番を構造として持っている。

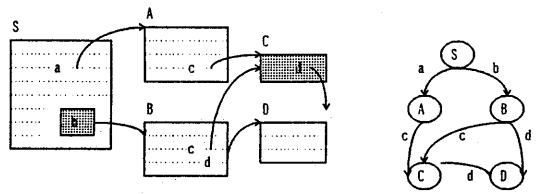


図3. ハイパーテキストの意味ネットワーク構造

また(STEP2)に入力される情報の構造は,全く構造を持たない状態あるいは,ハイパーテキスト<sup>5)</sup>で表せるような意味ネットワーク構造(図3)になっている。

このため(STEP2)では,これらの異なる構造を持った情報を統一的に扱うことが必要となる。ここでは,以前提案した,トータルメディアアーキテクチャ(TMA)

<sup>6)</sup>における情報の構造モデルを適用することとする。TMAでは,情報の構造を図4に示されるような階層的ネットワークで表現している。情報の単位(文,図表,段落など)は基本オブジェクトとなっており,その構造は,基本オブジェクトをもとにしたツリー上の階層構造と,同一階層のオブジェクト間のネットワーク構造を合わせ持っていることが特徴となっている。この構造表現を用いると,前述の文書論理構造は,基本オブジェクト上に構築された階層構造に対し,論理展開を表す順番を与えたものと考えることができ,また,ハイパーテキストのような情報は基本オブジェクトのネットワーク構造が対応していると考えることが可能である。

#### 3.2 文書の論理構造作成の過程

次にTMAの構造モデルに従って,実際に文書を作成していく場合の過程と,そこで生成される文書構造の関係について調べることにする。論理構造の作成時における処理モデルを図4に示す。このモデルでは,処理過程を(a)統合化 (b)関係付け (c)論理展開付け (d)内容補完の4つのフェーズに分けている。文書論理構造を作成時にはこの過程を繰り返すことで,より完成した

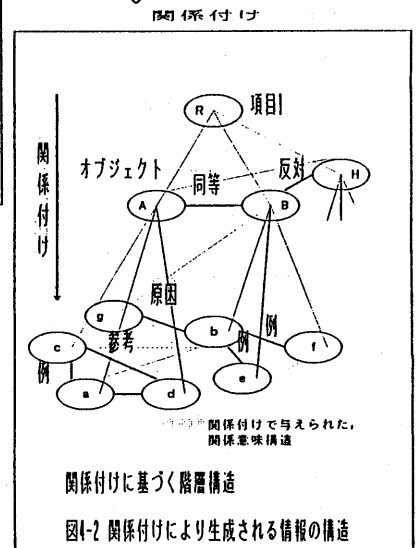
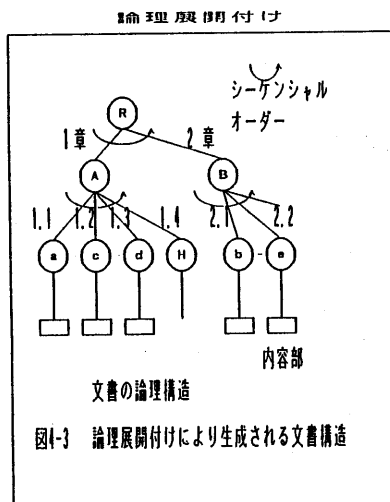
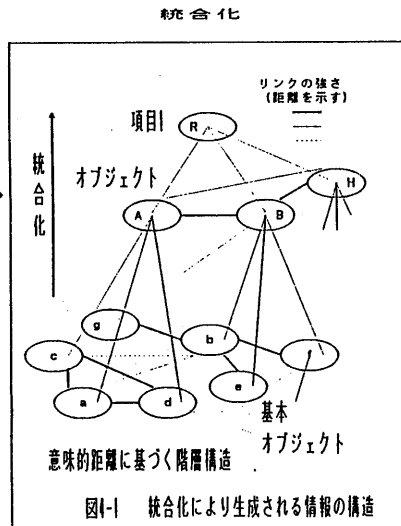
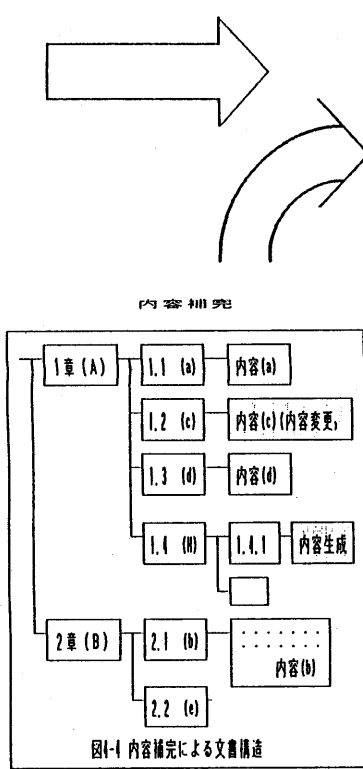
構造を生成していると考えられる。

(a) 統合化

収集した情報（基本オブジェクト）間の関係を基準にして、情報間の階層構造を作成する。この過程は、基本オブジェクト間の意味的距離をもとに、グルーピングし上位の概念（オブジェクト）を生

成し、さらにオブジェクトをグルーピングするといったことを繰り返すことにより、階層構造を生成する過程である。この結果、基本オブジェクト間あるいはオブジェクト間の意味的距離を表すネットワーク構造を持った階層構造が生成される。図4-1の例では、(a, c, d)および(b, e, f, g)を各々統合し、上位概念A, Bを作

(STEP1より)



(STEP3へ)

図4 文書論理構造作成過程のモデル

っている。処理方法としてはK J法のように、カード間の配置で統合する方法などが、実際には行われているが、情報が多い場合などには自動化による支援法が望まれる。

#### (b) 関係付け

(a) で階層化された情報について、論理的な関係、例えば因果関係や並列関係などの意味付けを行う過程で、図4-2に示すように関係付けの結果、各層の間にできるネットワーク構造は、論理関係(因果、例文等)が記述されることとなる。これはK J法における図式化あるいは、階層ネットワークにおけるハイパーテキストのリンク生成に対応している。

#### (c) 論理展開

(a) (b) の結果得られる情報を見つつ、文書の目的、主張に沿った論理展開をもとに、論理構造のみを作成する過程である。ここでは、先の、階層構造を参考に、論理展開に沿った順番をつけることも行われるが、論理展開上必要な新たな概念を導入したり、先の階層関係の一部だけを切り出して使用する、組み換えるなどの作業がなされる。このため、結果は(b)の階層構造と必ずしも一体一に対応するとは限らなく、図4-3に示されるように、結果的に文書の論理構造と同じ構造になるように変換することとなる。すなわち、論理展開上、オブジェクトHはA、Bと同じ階層にあったが、Aの下にその情報群を持ってきているし、内容部が必ずしも存在するとは限らない。

現在、アウトラインプロセッサと呼ばれるものは、この処理を行っているのだが、(a)(b)の処理が支援されていないため十分な威力を発揮していない。

#### (d) 内容補完

(c) の論理展開を行った場合、足りなかった内容部について記述したり、他の情報データを内容として付加す処理である。また、論理展開に応じて行う、内容部の推敲作業も含まれていると考えることができる。これによって、新たに生成された内容部は、再び従来の情報と一

緒にして統合化処理に用いられ、情報の関係の再構築に使われる。

## 4. 論理構造作成時の支援法における課題

### 4.1 情報の統合化の支援

図4の流れに従って、支援方式について検討する。まず、統合化の過程についての支援法を考える。現在、情報を統合化して、構造を明確にしてくれる支援システムは作られていないが、思考支援法で知られるK J法<sup>7)</sup>が、カードに書かれている多くの情報の統合して、階層化を行うというこの過程を実現している。

図4-1にあるような処理を、電子化した文書上で実現するためには、情報間の意味的距離を求めることが重要となる。具体的手段としては、K J法をディスプレイ上で実現して、そのグラフィクス画面上の距離を利用する方法<sup>8)</sup>が試みられているが、ここでは、多くの情報があっても適用でき、しかも電子化されている特徴を生かす方式としてその自動化について検討する必要がある。

### 4.2 関係付けの支援

関係付けの段階では、分類、統合されて、階層関係がつけられた情報間に、意味関係を与える作業となる。

この段階では、作者の意図に応じて関係付けの意味などは変化するため、自動化による支援よりは、情報入力法や結果の提示法といったヒューマンインタフェース設計が重要になってくる。現在、ハイパーテキストでは、このインタフェースの設計を重視しているが、今後さらに改善が必要である。

### 4.3 論理展開の支援

次に、階層化された情報に対し、論理展開を与える場合の支援方法について考える。

アウトラインプロセッサは論理展開を先に作り、論理に足りない情報を生成してゆくという意味で、基本的にはこの過程については支援がされていると考えられる。

ここで、従来のアウトラインプロセッサで問題となる点は、すでに整理された情報をいかに有効に、アウトラインの論理構造上に適用することができるかという点にある。すなわち、情報のシナリオに従って、階層的な論理展開を先ず作成し、適用できる情報をそこに当てはめてゆくため、そのシナリオといかに矛盾なく、既存の情報をあてはめるかという点が問題となる。そこで、重要になるのが先の階層構造に与える情報間の関係であり、これを参考にして、アウトライン上に適切な論理構造とのマッチングをとることにより、全体の論理構造が構築できる。そこで、関係付けされた情報をアウトラインプロセッサに結合する支援環境が重要になる。

キーワード (度数値) 0:1 空欄:0

項目	キーワード	キ-1 y <sub>1</sub>	キ-2 y <sub>2</sub>	キ-3 y <sub>3</sub>	キ-4 y <sub>4</sub>	キ-5 y <sub>5</sub>
情報1	x <sub>1</sub>	F <sub>11</sub>	F <sub>12</sub>	F <sub>13</sub>	F <sub>14</sub>	F <sub>15</sub>
情報2	x <sub>2</sub>	F <sub>21</sub>	F <sub>22</sub>	F <sub>23</sub>	F <sub>24</sub>	F <sub>25</sub>
情報3	x <sub>3</sub>	F <sub>31</sub>	F <sub>32</sub>	F <sub>33</sub>	F <sub>34</sub>	F <sub>35</sub>
情報4	x <sub>4</sub>	F <sub>41</sub>	F <sub>42</sub>	F <sub>43</sub>	F <sub>44</sub>	F <sub>45</sub>
情報5	x <sub>5</sub>	F <sub>51</sub>	F <sub>52</sub>	F <sub>53</sub>	F <sub>54</sub>	F <sub>55</sub>
情報6	x <sub>6</sub>	F <sub>61</sub>	F <sub>62</sub>	F <sub>63</sub>	F <sub>64</sub>	F <sub>65</sub>

(a) 情報入力時のキーワード付与形式

キーワード	1	2	3	4	5
情報1		○		○	○
情報2	○	○		○	
情報3				○	
情報4					○
情報5		○	○		
情報6		○	○		○

キーワード	3	5	2	4	1
情報4		○			
情報6	○	○	○		
情報5	○		○		
情報1		○	○	○	
情報2				○	○
情報3					○

(b) 情報入力時のキーワード (c) 数量化理論による項目間の並び替え

図5 数量化理論Ⅲ類による項目間の距離の求め方

#### 4.4 内容部生成の支援

アウトラインプロセッサで、足りない内容部を生成することも、重要な過程となる。このためには、新たな情報検索の支援ならびに関連項目の提示などが、支援環境として重要なものになると考えられる。

#### 5. 情報の統合化方法に対する検討

文書の論理構造を生成における支援方法は、以上のべたように、各フェーズにおいてそれぞれ多くの課題があることを明らかにした。本稿では、この支援法の第一段階として、まず情報の統合化における支援方法を取りあげ、その中で、課題となっている、情報の距離関係を自動的に求める方法について検討する。

##### 5.1 数量化理論による統合化

情報間の距離関係を求める具体的方法としては、KJ法と同様、画面上で配置、編集して求める方法等が考えられる。しかし大量情報を扱えるようにするためには、その自動化が支援方法として重要になる。そこで本章では、数量化理論<sup>9)</sup>を用いた自動的な統合方法を提案し、実験的検討を行う。

本方法では、各情報（基本オブジェクト）に、図5(b)のようなキーワードに対する値を(1, 0)で与える。このキーワードの表において、キーワードの項目を、因子Xとみなし、また各情報の項

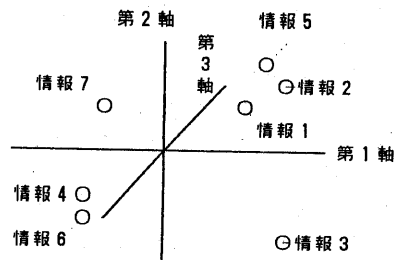


図6 数量化により、多次元空間に配置される情報

目を因子Yとみなすと、図5(b) は因子Xと因子Yの相関を表す度数分布表 $F_{ij}$ とみることができる。(図5(a))

数量化理論Ⅲ類では、以上の項目とキーワードから、その因子間の相関が最も大きくなるように、キーワード間の距離、項目間の距離を並びかえて計算し、いくつかのキーワードの結果から得られた新たな概念軸上に、項目を配置してくれる。

実際には、各因子の座標値を $x_i, y_j$ としたとき、相関係数 $\rho_{xy}$ を最大にするように、 $x_i, y_j$ の値を求める。ここで

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$\sigma_{xy} = \sum \sum x_i F_{ij} y_j$$

$$\sigma_x = (\sum F_{ij} x_i^2 - (\sum F_{ij} x_i)^2)^{1/2}$$

$$\sigma_y = (\sum F_{ij} y_j^2 - (\sum F_{ij} y_j)^2)^{1/2}$$

である。

この結果、複数の $(x_i), (y_j)$ のベクトルの対が求まり対応した相関係数が得られる。例えば、図5(c)は(b)に対して、相関が最大となる $(x_i), (y_j)$ のベクトルの対に従って、項目とキーワードをベクトルの座標値順に並びかえたものである。

また、複数の解を情報、またはキーワードごとにベクトルの座標に従って、各々の多次元の空間上へ配置すれば、図6のように示すことができる。以上からキーワードを用いて、比較的容易に多次元空間上の座標から情報の分類統合に必要な距離および階層構造を求めることができる。

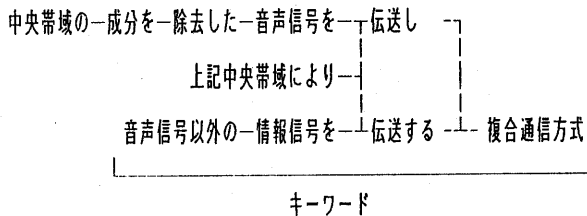


図7 係り受けを用いたキーワード抽出

## 5.2 文書における統合化の実験と考察

数量化による分類、統合化が適用可能性を調べるため、実際の文書に使われている、文を基本単位としてその文書の統合化を行う実験を行った。ここで、使用した文書は特許の明細書を用いることとした。

キーワード抽出法としては、各文の係り受け関係から求める方式を採用することとした<sup>10)</sup>。すなわち係り受けの解析を図7に示すように行って、係り受けの

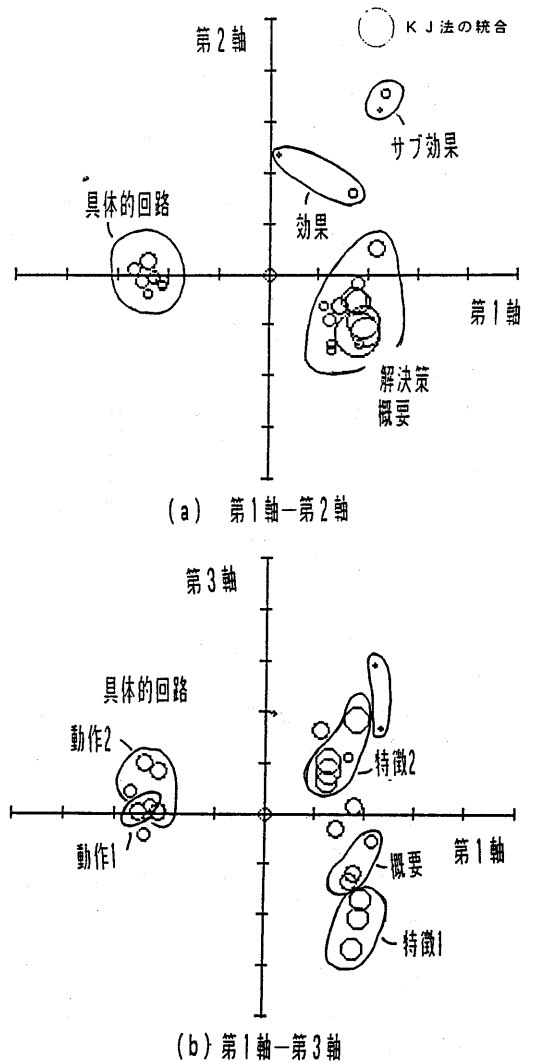


図8 統合化支援の実験結果

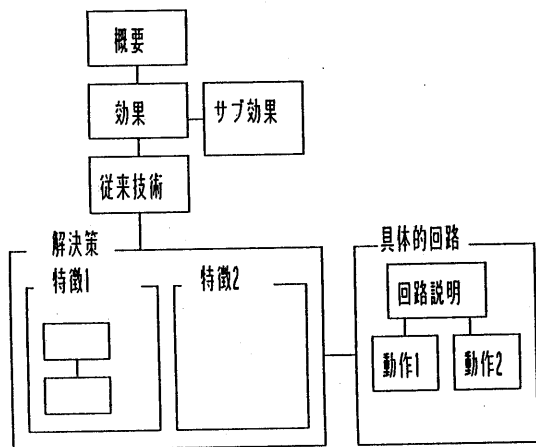


図9 KJ法による、特許文の構成

深い方より5層分選び、この結果を各文の重要語と仮定してキーワードとした。本文書の文の数は27文で各文に付与されたキーワード数は文の長さによって異なるが、10から20程度であり、全キーワード数は58であった。

図8は本方式で得られた、各文の間の距離関係を示したものである。(a)は第1軸と第2軸の関係をXY軸に第3軸を点の大きさで示したものであり、(b)は第1軸と第3軸の関係をXY軸に第2軸を点の大きさで示したものである。図中、実線で囲んだものは、別途行ったKJ法の結果(図9)に従って分類、統合された文書の構造を示したものである。

この結果、本方式による分類ではKJ法の結果と同様、方式の説明部と実施例の部分をはっきり区別でき、また、内容の分類に関してもかなりの一致点が見られる。

この理由としては、特許文書の場合、各文の論理的役割がはっきりしている点、使用される言葉が、その文の特徴を表している点、同じ概念に対して、キーワードが1対1に対応可能である点などがうまく作用するためと考えられる。

以上の結果より、比較的論理性の高い文書を作成する場合には、本方式による統合化の支援の適用可能性があることが示された。

## 6. むすび

文書作成の過程の中でも、特に文書の論理構造を作成する過程での、情報の構造に注目し、その過程のモデル化ならびに支援方式について検討を行った。その結果、論理構造作成時には、統合化、関係付け、論理展開付け、内容補完の四つの過程を繰り返し行うこと、また各構造を階層的ネットワークで記述可能なことを示した。

さらに、統合化の自動的支援方式として、数量化理論を用いた方式を提案し、特許文書などにおける適用可能性について示した。

### (参考文献)

- 1) W. Hershey "Idea Processor", Byte Jun. 1985 337-350
- 2) 平野, 日吉, 永井 "オフィス文書作成支援システム -基本概念と機能-", 情報処理学会 日本語文書処理研究会 10-2 Jan. 1987
- 3) 土井, 岩井 "知的文書処理システムにおける文書処理モデル" 信学会 オフィスシステム研究会 OS86-20 Mar. 1986
- 4) ISO/TC97/SC18 IS 8613
- 5) J. Conklin "Hypertext An Introduction and Survey", IEEE Computer, Sep. 1987 17-41
- 6) 鈴木他 "トータルメディアアーキテクチャの検討" TV学会研究会 IPA '88-16
- 7) 発想法 川喜田二郎 (1967) 中央公論社
- 8) 河合, 小山, 大岩 "ワークステーションを用いたKJ法: 概念ネットワーク形成ツール" 信学会 オフィスシステム研究会 OS86-50 Mar. 1987
- 9) 小林 数量化理論入門 日科技連
- 10) 稲垣 他 "特許請求範囲文の係り受け解析について" 情報第35回全大予稿133 (1987)