

## 文の意味役割解析に基づく全文検索

三池誠司 住田一男

(株)東芝 研究開発センター

文書構造に基づく検索と抄録生成を行う全文検索システムBREVIDOCの検索機能の評価について述べる。本検索機能は文の意味役割を解析し抽出することにより検索文書の適合性を判定することを特徴とする。技術論文547文書を対象とした6種類の検索命令での実験において、検索語の有無に基づく全文検索方式では適合率が平均31%であるのに対し、本方式では「話題」、「背景」、「従来の問題」、「目的」、「特徴」、「結果」、「結論」、「課題」の意味役割を抽出し利用することにより適合率で平均56%の値を得た。ここで、再現率は前者に対して87%である一方、検索された文書総数の比は前者に対して平均44%であった。本方式が、検索結果の絞り込みや検索文書の適合性の判定に有効であることを確認した。

### FULL-TEXT RETRIEVAL BY ANALYSIS OF SEMANTIC ROLES OF SENTENCES

Seiji Miike Kazuo Sumita

Research and Development Center, Toshiba Corp.

Komukai-Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan

This paper describes evaluation of the retrieval function implemented in a full-text retrieval system BREVIDOC, which retrieves documents and generates their abstracts based on a document structure. The retrieval function extracts semantic roles of sentences in a document to be used for judgment of its document relevancy. In an experiment using 547 technical papers and 6 queries, the precision was 56% on average using following semantic roles, "topic", "purpose", "background", "issues", "feature", "result", "conclusion" and "feature work", while it was 31% on average by traditional retrieval based only on existence of query words. Meanwhile, the recall was 87% on average with respect to the method, and the ratio of the number of documents retrieved by this function to that by the traditional retrieval was 44% on average. We have confirmed that the retrieval function is effective for judgment of document relevancy.

## 1. はじめに

大量の情報が利用可能になるにつれて、情報を様々な側面や特徴から検索し利用する技術の重要性が指摘されている([1]など)。一方、現在の検索システムでは、出力された検索結果から必要な情報を得るための分析を行う必要があり、出力される検索結果が大量である場合にはその負担はさらに大きくなるという問題が指摘されている([2]など)。このような状況において、情報検索理論や技法の向上が求められている[3]。

我々は、効率的な検索を目的とし、文書の構造に基づく検索と抄録生成を行う全文検索システム BREVIDOC(Broadcatching System with an Essence Viewer for Retrieved Documents)を試作した[4]。文献[5]と[6]で、BREVIDOCで用いている、文書中での文の意味役割を解析し抽出することを特徴とする検索機能について報告した。本稿では、この検索機能の評価について述べる。2章と3章で、文書構造に基づく検索とこれを実現した文書構造解析システムについて述べ、4章と5章で検索実験と考察について述べる。

## 2. 文書構造に基づく検索

情報検索において、文書の適合性(relevancy)を識別するための様々な方法が研究開発されてきている。例えば、単語の出現頻度や、段落や文の単位を利用した単語の近接性などを利用した適合性の判定方法などがある。また、passage retrievalの分野では、文書内容や文脈の構造を利用した検索方法が研究されている([7], [8]など)。文献[8]では、sectionに<purpose>, <abstract>, <start>, <summary>, <title>, <supplementary>, <misc>のタグが付けられたTREC conferenceのテキストデータを対象とした検索実験を行い、<purpose>と<summary>の重みを2、<misc>の重みを0.5、その他の重みを1とした場合に最もよい検索結果を得たとの報告がされている。

文献[8]ではタグ付けがされたテキストデータを対象としているのに対し、我々はタグ付けされていないテキストデータを対象とする点が異なる。タグ付けされていないテキストデータを対象としこれを構造解析する理由は、タグ付け、特に文書の内容に基づくタグ付けは労力を要するので、一般のテキストにタグ付けされることが期待できないことと、さらに、検索命令や文書の内容に基づく検索([9], [10]など)を行うには文書構造の解析が欠かせないからである。文書

の構造化については文献[11], [12], [13]などの研究があるが、我々は文書構造化の検索への応用を図っている。

文書構造のモデルとして、文書の種類と、その文書に含まれる典型的な情報の種類、あるいは論旨の展開に用いられる情報の種類を想定する。例えば、技術論文では「背景」、「目的」、「結果」、「結論」などが考えられる。また、新聞社説では、「背景」や「意見」などが考えられる。ここでは、これらの文書が伝達する情報の種類を文単位で処理し、この情報の種類、すなわち文書全体に対する文の機能を文の意味役割とよぶ。文の意味役割の抽出においては、検索への応用(検索結果の絞り込みやランキング付け)を目的とするので、すべての文に何らかの意味役割を付与することはせず、抽出できる文のみから抽出する。一つの文から2つ以上の意味役割が抽出されることも許容する。この文の意味役割の抽出と、インデックスファイル作成のために文書構造解析システムを実現した([5],[6])。

## 3. 文書構造解析システム

文書構造解析システムは、図1に示すように、テキストデータを入力とし、文の意味役割を抽出して意味役割ごとに単語インデックスファイルを生成する([5],[6])。日本語解析部には平川らの解析システム([14],[15])を用いており、インデックス作成部には中本らのシステム[16]を用いている。文脈情報抽出部と文脈情報解析部については文献[17]、[18]および[19]を参照されたい。

文の意味役割は、表層的な表現を手掛かりとした抽出規則を用いて、一文単位で抽出する。さらに、文の階層関係を示す文脈構造を参照し、投射規則に応じて意味役割を前後の文に複写する。

## 4. 検索実験

### 4.1 実験の目的

文の意味役割抽出が、検索結果の絞り込みやランキング付けに有効であることを確認する。

### 4.2 実験の方法

技術論文誌である東芝レビューを対象とし、テストデータで文の意味役割抽出規則を記述・評価し、この規則を用いて評価データの検索を行い、検索結果を主題検索[3]の観点から評価す

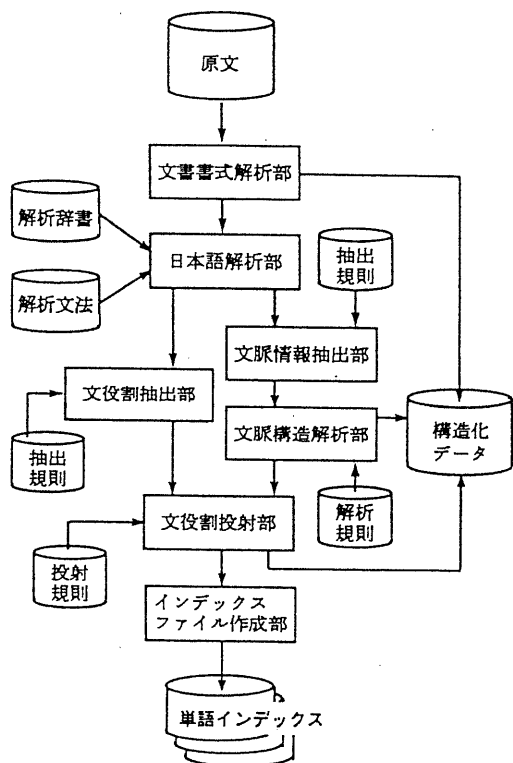


図2 文書構造解析システムの構成

る。論文中の図や表などを除いたテキストを検索対象とする。

#### (1) テストデータを対象とした意味役割抽出規則の記述と評価

テストデータを対象として、文の意味役割抽出規則を記述し、評価する。意味役割として、「話題」、「背景」、「従来の問題」、「目的」、「特徴」、「結果」、「結論」、「課題」の8種類を用いる。ここで、「従来の問題」とは従来の技術や方法の問題を述べた文の種類である。

(1.1) テストデータ中の文について人手で意味役割を判定し、その結果を格納する。ここでは、この結果を格納したファイルを正解ファイルとよぶ。

(1.2) 人手で判定した意味役割を抽出するための抽出規則を記述する。この抽出規則を用いて抽出した結果と正解ファイルとを照合して、文の意味役割抽出規則を修正する。

(1.3) 抽出規則を用いて抽出した結果と正解ファイルとを照合して、文の意味役割抽出規則を評

価する。

#### (2) 評価データを用いた検索結果の評価

評価データを用いて検索を行う。検索命令中の検索語の同義語や類似表現などへの展開は行わない。

(2.1) テストデータを対象として記述した抽出規則を用いて評価データ中の文の意味役割を抽出し格納する。

(2.2) 検索命令を設定し、8種類の意味役割と、文の意味役割を区別しない全文検索で検索する。

(2.3) 検索されたすべての文書について検索命令に適合するか否か、すなわち検索命令中の検索語が文書の主題と関係があるか否かを人手で判定する。ここでは、検索命令に適合する文書を適合文書とよぶ。

(2.4) 複数の検索命令について(2.2)と(2.3)を繰り返す、集計する。

### 4.3 実験結果

#### (1) テストデータを対象とした意味役割抽出規則の記述と評価

20文書、1549文(平均77文/文書)からなるテストデータに対し、前節に示した8種類の意味役割について、290の意味役割抽出規則と22の投射規則を記述した。

意味役割の再現率は平均82%であり、適合率は平均72%であった。表1に、文の意味役割ごとの再現率(%), 適合率(%), 当該の意味役割が存在した正解ファイルの数、当該の意味役割が抽出されたファイルの数、および文書全体の文の数に対する図2のa、b、cの範囲の文の数の割合を示す。表1中の意味役割「問題」は「従来の問題」であることを示す。表1から次のことがわかった。

- ・最も多くのファイルから抽出された意味役割は「背景」であり、すべての文書から抽出された。最も少ないファイルから抽出された意味役割は「課題」であった。

- ・図2のa、b、cの範囲の文の数は、平均して文書全体の文の数の1.2%から9.3%であり、数文であった。

文の意味役割を適切に抽出できなかった場合には、次のようなものがあった。なお、例にあげた文書中の文の番号は筆者が付与したものである。

表1 文の意味役割の再現率と適合率

A	B	C	D	E	F	G	H
話題	100	64	16	17	0	3.3	2.3
目的	93	67	6	6	3.8	2.4	1.5
特徴	62	82	10	10	6.8	2.7	1.3
結果	72	81	18	18	5.8	7.8	2.5
結論	95	82	17	17	1.6	5.1	3.7
課題	67	50	3	4	2.2	1.3	1.9
背景	76	83	20	20	3.6	9.3	3.5
問題	90	70	13	14	1.2	1.8	1.9
平均	82	72					

- A: 意味役割  
 B: 再現率(%)  
 C: 適合率(%)  
 D: 当該の意味役割が存在した正解ファイルの数  
 E: 当該の意味役割が抽出されたファイルの数  
 F: 図2のaの領域に含まれる文の数の平均  
 G: 図2のbの領域に含まれる文の数の平均  
 H: 図2のcの領域に含まれる文の数の平均

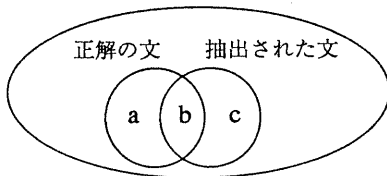


図2 一つの文書の中の文の分布

(a)例1の第2章の第1文から第5文では背景の情報が述べられているが、第4,5文から意味役割が抽出できなかった。文書構造解析システムの文脈構造解析では接続詞などの文書に依存しない言語情報を用いているが、ここでは「高信頼性」、「高品質」の関連性や「InGaAlP」の重複などを解析・利用できる必要がある。

(b)例1の第3章の第2,3,4,7,8文はこの文書の実験結果に関すると思われるが、意味役割を抽出できなかった。ここでは、それらに先行する第1,5文は参照すべき図を示す文となっている。従って、一文の中に手掛かりとなる表現を含まない場合でも、先行する文の情報を用いて意味役割を抽出するように文書構造解析システムを拡張すれば対応可能であると考えられる。また、図や表を示して述べる内容が文書の主題と密接な関係があるかという問題があるが、例えば当文書の第1文が文書のタイトル中の語を含んでおり、そのような情報の利用も有益である。

例1. "0.6 $\mu$ m 帯赤色半導体レーザーTOLD9200", 本館他, 東芝レビュー Vol.43, No.5, 1988.

## 2. 結晶材料と素子構造

(1)可視光領域で発振する半導体レーザーを実現しようとする場合、結晶材料の選択が重要である。(2)現在、主流として用いられているGaAlAs系材料では、バンドギャップが小さく0.72 $\mu$ mまでが短波長化の限界とされているので、新しい材料を開発することが必要となる。

(3)半導体レーザー用の材料は、結晶欠陥が少なく、高信頼性をもつことが要求される。(4)高品質な基板結晶が得られるGaAsに格子整合がとれ、かつ、短波長化に有利な特長をもつ材料としてInGaAlPがある(1)。

(5)InGaAlP四元混晶は、窒化物を除いたIII-V族化合物半導体混晶の中で、もっとも大きな直接遷移形のバンドギャップをもっており、短波長化にもっとも有利な結晶材料である。(以下省略)

## 3. 素子特性

(1)図3に、TOLD9200の典型的な電流-光出力特性とその温度依存性を示す。(2)周囲温度25℃での連続(CW:Continuous Wave)動作におけるしきい値電流は標準で76mAである。(3)また、スロープ効率は約0.5mW/mAである。(4)周囲温度5℃から40℃の範囲でキックのない特性が得られている。

(5)図4に、光出力2mWでのレーザー光の遠視野像を示す。(6)ガウス分布形状をした基本横モード発振を示している。(7)ビーム広がり角の典型的な値は、接合に垂直方向で34°、平行方向で7°である。(8)周囲温度(5~40℃)および光出力(1~3mW)の値を変えても、広がり角には変化はまったく見られない。(以下省略)

(b)例2では第1,2,3文が背景の情報について述べているが(第2,3文は省略)、意味役割を抽出できなかった。この文書では、第4文に「以上のような標準化経緯からわかるとおり」という表現を含んでいるので、この後方の文の情報から前方の文の意味役割を抽出するように文書構造解析システムを拡張すれば対応可能であると考えられる。また、この文書では、(b)の場合に反してタイトル中にある「ISDN」は本文の主題とは関係が薄い。タイトル中の語と文書の主題との関連の解析も必要である。

例2. "ISDNによる企業情報通信システム、谷野他、東芝レビュー Vol.44, No.4, 1989.

## 2. ISDNの標準化経緯とその特長

(1)ISDNは、電話網のデジタル化をねらいとして1970年代初期から検討が開始され、1980年代初期にはデジタル伝送技術とデジタル交換技術を統合して、一つのデジタル回線上でさまざまな種類・型式の情報を伝送するネットワークとしての姿が明確になった。(途中省略)

(4)以上のような標準化経緯からわかるとおり、ISDNは電話サービス網のデジタル化を基盤としているが、単に電話サービスだけでなく、データ通信や音楽・文書・ファクシミリ・ビデオテックス・ビデオなど非電話系の広範囲の通信サービスを一つのネットワークから提供することを主なねらいとしている。(以下省略)

(d)例3は最終章であり第1文に「完了した」という表現を含むが、当文書の結論ではなく従来技

術についての情報である。このような文書への対応は今後の課題である。

例3. "高品位画像産業用CTスキャナTOSCANER-20000シリーズ", 庄司他, 東芝レビュー Vol.44, No.5, 1989.

4. あとがき  
 (1)CT画像のS/Nや空間分解能力の向上は、その要求にしたがって相応のシステムを実現するための技術的基礎固めを完了した。(2)ただし、CTがよりいっそうの飛躍をするためには透過能力、空間分解能、S/Nなど主要な画像性能を満たしたうえで、スキャン速度を向上させる必要がある。  
 (3)人体用のMRI(Magnetic Resonance Imaging)のように、1回のスキャンで数十断面のデータが得られるようなシステムが前記図7の構想では必要である。(4)20000シリーズでは1断面/スキャンで、その所要時間は約3分である。(以下省略)

(2)評価データを用いた意味役割抽出の評価

547文書からなる評価データから文の意味役割を抽出格納した結果について、表2に示す6種類の検索命令で検索を行った。検索命令中のすべての検索語をある意味役割の文に含む文書が当該の意味役割での検索結果の文書となる。各検索命令での検索結果は付録に示す。各検索結果の適合率と再現率をを表3と表4に示す。なお、再現率としては、本来検索対象のすべての文書中の適合文書の数に対する検索結果中の適合文書の数の比を示すべきであるが、ここでは文の意味役割を区別しない全文検索結果の適合文書の数に対する比を示している。表3と表4において、「その他」とはいずれの意味役割も抽出されなかった文にのみ検索語を含む文書の場合である。

「なし」とは文の意味役割を区別しない全文検索の場合である。表3中の「-」は検索された文書がなく、適合率が得られなかったことを示す。

表3と表4から、各々意味役割および「その他」の適合率と再現率について次のことがわかった。

表2 検索命令

番号	検索命令
1	ロボット AND 点検
2	ロボット AND 画像処理
3	画像 AND 診断
4	画像 AND 伝送
5	衛星 AND 通信
6	情報 AND 検索

・「話題」、「目的」、「特徴」、「結果」、「結論」、「課題」の適合率は、65%以上であった。

・文書の主題との関係が薄いと考えられる「背景」と「従来の問題」では、56%と33%の適合率であり、他の意味役割に比べて低かった。

・文の意味役割が抽出されなかった場合である「その他」の適合率は、意味役割の場合に比べて8%と非常に低かった。

・再現率は、意味役割の中では「背景」が最も高く50%であった。

一つの文が複数の意味役割をもつ場合があることと、表8の結果から、意味役割を「課題」、「話題」、「特徴」、「結論」、「目的」、「結果」のAグループと「背景」、「従来の問題」のBグループに分けて次の場合の適合率を調べた。図3に検索された文書の分布を示す。図4から図6の中の円は、図3の中の円に対応する。

- (a)AグループまたはBグループのいずれかの意味役割の文で検索された場合(図4に対応)
- (b)Aグループの意味役割の文で検索された場合(図5に対応)
- (c)Aグループのみの意味役割の文で検索された場合(図6に対応)

表3 意味役割別の適合率

番号	話題	目的	特徴	結果	結論	課題	背景	問題	その他	なし
1	67	100	100	100	50	-	33	-	0	22
2	100	100	100	100	100	-	50	0	0	40
3	100	-	100	100	100	100	92	100	33	57
4	50	-	-	0	67	-	33	-	4	16
5	67	100	100	67	90	-	53	-	8	39
6	100	0	0	25	50	-	75	0	2	10
平均	81	75	80	65	76	100	56	33	8	31

表4 意味役割別の再現率

番号	話題	目的	特徴	結果	結論	課題	背景	問題	その他	なし
1	100	100	50	100	50	0	50	0	0	100
2	50	25	25	50	50	0	25	0	0	100
3	15	0	5	40	5	5	55	15	35	100
4	17	0	0	0	33	0	50	0	17	100
5	31	15	8	31	69	0	69	0	8	100
6	17	0	0	33	17	0	50	0	17	100
平均	38	23	15	42	37	1	50	3	13	100

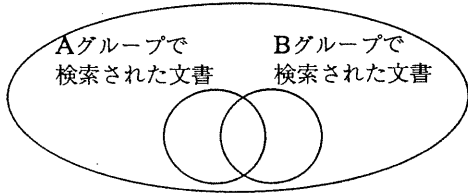


図3 検索された文書の分布

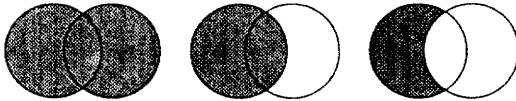


図4 AグループのみまたはBグループのみ  
図5 Aグループのみ  
図6 AグループのみまたはBグループのみ

表5に全文検索の場合の検索文書、適合文書および適合率を示す。表6、表7、表8に、(a)、(b)、(c)の場合の検索された文書と適合文書の数、適合率、全文検索の適合率との比、検索文書総数の比、および再現率を示す。表10、表11、表12の中のDからIがこれらを表している。

- D: 検索された文書の数
- E: 検索された文書の中の適合文書の数
- F: 適合率(E/Dの値、%)
- G: 適合率の比(F/Cの値、%)
- H: 検索文書総数比(D/Aの値、%)
- I: 再現率(E/Bの値、%)

表6、表7、表8から次のことがわかった。

- ・(a)、(b)、(c)の順で適合率が上昇し、検索文書総数比が減少した。ただし、(b)と(c)の適合率の差は大きくなかった。
- ・(a)、(b)、(c)の順で再現率が低下した。(b)と(c)の再現率の差は大きかった。

以上から、意味役割を用いることにより、検索文書総数の減少と、適合率の向上を得られることを確認した。

## 5. 考察

### (1) 意味役割抽出規則の記述について

情報抽出規則の記述にあたっては、意味役割の再現率を高くするような規則は不適切な文にもマッチする危険性が高く、逆に個別な規則では抽出の網羅性を期待できない。表層的な表現を手掛かりとする本方法では、その傾向は著し

表5 全文検索における適合率

番号	A	B	C
1	9	2	22
2	10	4	40
3	35	20	57
4	38	6	16
5	33	13	39
6	63	6	10
平均			31

- A: 検索された文書の数
- B: 検索された文書の中の適合文書の数
- C: 適合率(B/Aの値、%)

表6 (a)の場合の適合率と再現率

番号	D	E	F	G	H	I
1	4	2	50	2.3	44	100
2	6	4	67	1.7	60	100
3	14	13	93	1.7	39	65
4	13	5	38	2.4	34	83
5	21	12	57	1.5	64	92
6	15	5	33	3.7	23	83
平均			56	2.2	44	87

表7 (b)の場合の適合率と再現率

番号	D	E	F	G	H	I
1	3	2	67	3.0	33	100
2	4	4	100	2.5	40	100
3	9	9	100	1.8	25	45
4	7	3	43	2.7	18	50
5	13	10	77	2.1	37	77
6	13	4	31	3.4	20	67
平均			70	2.6	29	73

表8 (c)の場合の適合率と再現率

番号	D	E	F	G	H	I
1	1	1	100	4.5	11	50
2	3	3	100	2.5	30	75
3	2	2	100	1.8	5.6	10
4	4	2	50	3.1	11	33
5	4	3	75	2.0	11	23
6	10	2	20	2.2	15	33
平均			74	2.7	14	37

い。意味役割の抽出精度の向上のためには、深層的な情報や前後の文の情報を解析し利用することが必要である。

### (2) 意味役割抽出の検索における効果について

本方法では従来の全文検索方式に比べて高い

適合率を得られることから、特に適合率を重視する検索の場合において、検索作業の効率化に寄与できると考えられる。

一方、本方法により得られる適合文書の数は、全文検索で得られる適合文書の数の平均87%以下であるという問題がある。また、本方式では、文からの意味役割の抽出において、その精度が形態素解析および構文・意味解析の精度に依存する。検索システムとしては、形態素解析を用いない文字レベルの検索、形態素解析を用いた単語レベルの検索などと、本方式とを組合わせて後処理・絞り込み機能として用いることにより、システム全体では漏れがないようにすることが必要であるが、これは実現可能である。また、検索された文書を検索命令に最も適合する文書から提示するランキング表示や、relevance feedbackにおいて、検索文書の適合度の尺度の一つとして用いることが有効であると考えられる。

## 6. おわりに

文の意味役割解析に基づく全文検索の実験と評価を行った。技術論文547文書を対象とした6種類の検索命令での検索実験において、検索語の有無のみを判定基準とする全文検索方式では適合率が平均31%であるのに対し、「話題」、「背景」、「従来の問題」、「目的」、「特徴」、「結果」、「結論」、「課題」の8種類の意味役割を利用することにより適合率平均56%を得た(再現率は前者に対して平均87%)。このとき、検索文書の総数は前者に対し44%であった。さらに、「背景」、「従来の問題」を除く6種類の意味役割では、再現率が73%に下がるものの適合率は70%に上昇し、検索文書総数の比は29%に減少できた。本方式が、検索結果の絞り込みに有効であることを確認した。また、本方式と従来の語の出現頻度などの統計情報に基づく検索方式(20)などとの予備実験での比較評価を別途行い、本方式が後者の検索方式と比べて適合率を約1割向上できることを確認している[21]。

今後は、技術論文以外の文書についての適応・拡張を図る予定である。

[謝辞] 本稿で述べた文書検索システムの研究開発に際して、ご指導や討論をして頂いたマルチメディア研究所の岩井勇氏、中本幸夫氏、研究開発センターの田中克己氏、小野顕司氏、平川秀樹氏に感謝する。

## [参考文献]

- [1]長尾真、情報社会の生態学、情処学会人文科学とコンピュータ研究会 CH-11-6, 1991.
- [2]三輪真木子、データベースサーチャーの視点、情報処理 Vol.33 No.10, pp.1162-1170, 1992.
- [3]細野公男、情報検索理論・技法の問題点とその解決の方向、情処学会情報学基礎研究会 FI-24-5, 1991.
- [4]Miike, S., et al., A Full-Text Retrieval System with a Dynamic Abstract Generation Function, Proc. 17th Annual Int. Conf. Research and Development in Information Retrieval (SIGIR'94), pp.152-161, 1994.
- [5]三池誠司他、文書の構造解析に基づく文書情報検索、情処学会情報学基礎研究会 FI-31-6, pp.39-46, 1993.
- [6]Sumita, K., et al., Document Structure Extraction for Interactive Document Retrieval Systems, Proc. SIGDOC'93, pp.301-310, 1993.
- [7]Callan, J.P., et al, Passage-Level Evidence in Document Retrieval, Proc. 17th Annual Int. Conf. Research and Development in Information Retrieval (SIGIR'94), pp.302-310, 1994.
- [8]Wilkinson R., Effective Retrieval of Structured Documents, Proc. 17th Annual Int. Conf. Research and Development in Information Retrieval (SIGIR'94), pp.311-317, 1994.
- [9]新開正史他、主題の意味構造に基づく論文検索法の提案、情処学会第46回全国大会 6G-6, pp.4-207-208, 1993.
- [10]岸本行生他、テキストの構造化に基づく検索システム、情処学会論文誌、Vol.35 No.5, pp.908-916, 1994.
- [11]神門典子、構成要素カテゴリを用いた原著論文の内部構造分析、情処学会情報学基礎研究会 FI-25-7, pp.39-46, 1992.
- [12]神門典子、文章特性が異なる複数領域の原著論文の構造、情処学会情報学基礎研究会 FI-33-4, pp.17-22, 1994.
- [13]西村健士他、特定表現の重点的解析による科学技術論文構造化手法、情処学会情報学基礎研究会 FI-29-5, pp.35-42, 1993.
- [14]平川秀樹他、構文/意味優先規則による日本語解析、人工知能学会第3回全国大会 8-4, 1989.
- [15]平川秀樹他、日本語解析における最適探索、情処学会自然言語処理研究会 NL-74, 1989.
- [16]中本幸夫他、日本語解析を用いたフルテキストサーチの実験、情処学会第46回全国大会 4B-4, pp.3-125-126, 1993.
- [17]小野顕司他、文脈構造の分析、情処学会自然言語処理研究会 NL-70-2, 1989.
- [18]Sumita, K., et al., A Discourse Structure Analyzer for Japanese Text, Proc. Int. Conf.

Fifth Generation Computer Systems 1992 (FGCS'92), pp.1133-1140, 1992.

[19]小野頼司他、日本語論説文の自動抄録のための文脈構造解析、情処学会第46回全国大会 7B-10, pp.3-187-188, 1993.

[20]Turtle, H., et al., Evaluation of an Inference

Network-Based Retrieval Model, ACM Trans. Information Systems, Vol.9, No.3, pp.187-222, 1991.

[21]住田一男他、文の意味解析に基づく全文検索、人工知能学会第8回全国大会 24-2, 1994.

[付録] 検索結果

各検索命令について、意味役割ごとに検索された文書の数と、適合文書の数、適合率(%), 再現率(%)を示す。付表中の「その他」、「なし」、「-」の意味は本文表3と同じである。

一つの文が複数の意味役割をもつ場合があるので、各意味役割の検索文書の数と「その他」の検索文書の数の合計は、「なし」の検索文書の数と一致しない。適合文書についても同様である。

付表1 「ロボット AND 点検」の場合

意味役割	検索文書	適合文書	適合率	再現率
話題	3	2	67	100
目的	2	2	100	100
特徴	1	1	100	50
結果	2	2	100	100
結論	2	1	50	50
課題	0	0	-	0
背景	3	1	33	50
従来の問題	0	0	-	0
その他	5	0	0	0
なし	9	2	22	100

付表4 「画像 AND 伝送」の場合

意味役割	検索文書	適合文書	適合率	再現率
話題	2	1	50	17
目的	0	0	-	0
特徴	0	0	-	0
結果	2	0	0	0
結論	3	2	67	33
課題	0	0	-	0
背景	9	3	33	50
従来の問題	0	0	-	0
その他	25	1	4	17
なし	38	6	16	100

付表2 「ロボット AND 画像処理」の場合

意味役割	検索文書	適合文書	適合率	再現率
話題	2	2	100	50
目的	1	1	100	25
特徴	1	1	100	25
結果	2	2	100	50
結論	2	2	100	50
課題	0	0	-	0
背景	2	1	50	25
従来の問題	1	0	0	0
その他	4	0	0	0
なし	10	4	40	100

付表5 「衛星 AND 通信」の場合

意味役割	検索文書	適合文書	適合率	再現率
話題	6	4	67	31
目的	2	2	100	15
特徴	1	1	100	8
結果	6	4	67	31
結論	10	9	90	69
課題	0	0	-	0
背景	17	9	53	69
従来の問題	0	0	-	0
その他	12	1	8	8
なし	33	13	39	100

付表3 「画像 AND 診断」の場合

意味役割	検索文書	適合文書	適合率	再現率
話題	3	3	100	15
目的	0	0	-	0
特徴	1	1	100	5
結果	8	8	100	40
結論	1	1	100	5
課題	1	1	100	5
背景	12	11	92	55
従来の問題	3	3	100	15
その他	21	7	33	35
なし	35	20	57	100

付表6 「情報 AND 検索」の場合

意味役割	検索文書	適合文書	適合率	再現率
話題	1	1	100	17
目的	2	0	0	0
特徴	1	0	0	0
結果	8	2	25	33
結論	2	1	50	17
課題	0	0	-	0
背景	4	3	75	50
従来の問題	1	0	0	0
その他	51	1	2	17
なし	63	6	10	100