

複合語キーワードの効率的抽出法

林淑隆 獅々堀正幹 伊与田敦 津田和彦 青江順一

徳島大学工学部

文献検索システムなどにおいて、キーワードをいかに効率良く、かつ正確に抽出するかは重要な課題である。本論文では、日本語文書においてキーワードとなることが多い複合語が、キーワード抽出の際に多大なマッチング処理を要することに着目し、複数キーワードのストリングパターンマッチングマシンの手法を応用した複合語キーワードの効率的な抽出法を提案する。本手法は、形態素解析部と複合語キーワード抽出マシンAC部、複合語キーワード候補マシンAC部からなる。14個の複合語文法構造と10個のキーワード評価ルールを定義し、26文書について実験評価を行った結果、形態素解析部を除く平均抽出速度は16.58ミリ秒、文書1KBあたり6.18ミリ秒の結果が得られ、本手法の有効性を確認した。また、抽出キーワードの選別で必要となる重なり語の抽出は、候補マシンACにより効率的に行えるので、利用者はこのマシンACに対する抽出ルールを決定することで、多種多様なキーワードを決定することが可能となる。

An Efficient Extraction Method for Compound Keyword

Yoshitaka HAYASHI, Masami SHISHIBORI, Atsushi IYOTA, Kazuhiko TSUDA and Jun-ichi AOE

The Faculty of Engineering, The University of Tokushima

Extracting keywords efficiently is an important task in text retrieval systems. In Japanese text, there are many compound words consisting some kinds of characters (Katakana, Kanji, etc.) and the text has no delimiter among words. Therefore, extracting keywords from such a text takes a lot of time. This paper presents a technique of detecting keywords from compound keywords by introducing a set of rules, which are conditions for keywords construction. A string pattern matching machine for a finite number of patterns is applied to matching of the rules and storing keyword candidates. From the simulation results for 26 Japanese text files that the algorithm presented has performed 6.2ms/KB.

1. はじめに

キーワード抽出とは、文書から重要語を決定する技術であり、文献検索、テキスト編集など幅広い分野で応用されている基本的かつ重要な技術である。現在、キーワード抽出法は統制語辞書（シソーラス）を使用する統制語方式^[7]と、シソーラスを使用しない自由語方式^[3]の2つの方式がある。統制語方式は、キーワード候補をあらかじめシソーラスに用意しておき、シソーラスのキーワードが対象文書内に存在するか否かによって抽出処理を行う。また自由語方式は、まず対象文書に対して形態素解析などを行いキーワード候補を切り出し、次に候補から頻度情報などの重み付け処理を通してキーワードを抽出する。統制語方式は抽出処理は単純であるが、管理に多大な工数を要するシソーラスが必要となる。自由語方式は抽出処理は複雑であるが、シソーラスの管理が不要である。このため、統制語方式よりも自由語方式の方がより多く利用されている。

この自由語方式は、名詞表記（単語）を基本単位としてキーワードを抽出する方式であるが、日本語文章に関して適切なキーワードを抽出することは、次の理由により複雑な問題となっている。

- (1) 単語と単語の区切りがない「べた書き」であるので、単語の切り出しが複雑である。
- (2) 複合語が無限に造語されるため、単語とそれらを構成要素とする複合語の処理が複雑になる。
- (3) 字種が多いので、(2)の複合語処理をさらに複雑化している。

特に、キーワードとなることが多い複合語に関しては、これらすべてを辞書に登録することは非現実的であり、単語のみを辞書に登録し、複合語の構成情報を別に管理する手法^[4]が一般的である。これまでに提案されている複合語キーワード抽出法では、単語ごとにキーワード性を判定し、その連続部分を複合語として抽出する手法^{[8],[9]}や、キーワードパターンにマッチングする単語の連続部分を複合語として抽出する手法^[3]などがある。しかしながら、いずれの方式も多くのパ

ターンマッチング処理を必要とする。また複合語においては、他の多くのキーワード候補を部分文字列として含む場合や、あるいは相互の部分文字列を共有する場合が多く存在している。従って、抽出された複合語キーワード候補の中から、必要とするキーワードを決定する最終処理でも、多くの部分文字列とのマッチング処理を必要とする。このようにキーワード抽出では、形態素解析における辞書引きのコストとともに、マッチングコストも処理時間に大きな影響を与えている。

本論文では、このマッチング処理に対して、複数キーワードのストリングパターンマッチングマシンの手法^[1]を応用することによって、効率的な複合語キーワード抽出を実現する。

本手法では、まず複合語文法構造と複合語キーワード評価ルールを定義する。複合語文法構造には、一般的な複合語文法構造に複合語キーワードの精度向上のための構成要素を追加する。またキーワード評価ルールは、検出された複合語をキーワード候補として残すか否かの判定ルールである。複合語文法構造と複合語キーワード評価ルールから構成された特殊なストリングパターンマッチングマシン（マシンA C）を使用することで、複合語キーワード抽出アルゴリズムにおいて、キーワードの抽出と評価を同時に行うことを可能にした。さらに抽出された複合語キーワードを別のマシンA Cを構築し、部分文字列や共有文字列に対するマッチング処理を効率的に行う。また、これらのマシンA Cはすべて動的に構築できるので、複合語文法構造やキーワード抽出ルールの追加・削除処理が、非常に簡単に行うことが可能である。

以下、2章では本手法で用いる複合語文法構造とキーワード評価ルールについて、3章ではこれらを用いた複合語キーワード抽出アルゴリズムについて述べる。そして、4章では本手法の評価を行い、5章では今後の計画と課題について触れる。

2. 複合語文法構造とキーワード評価ルール

本章では、本システムで使用される複合語文法の構造と、複合語キーワード評価ルールについて説明する。まず、複合語文法構造について説明した後、複合語キーワード評価ルールの特徴について説明する。

2. 1 複合語文法構造

本システムはキーワード抽出法に自由語方式を採用しているため、抽出する複合語の構成情報をもたなければならない。この構成情報を表したものが、複合語文法構造である(表1)。

日本語文章における一般的な複合語に対する自動分割法には宮崎[2]の研究があり、本システムでは宮崎が提案した係り受け規則を基本に、複合語文法構造を定義する。

この文法構造は、複合語キーワード内の品詞の接続関係を表したものである。例えば「約10%以下」は「前置助数詞-数詞-後置助数詞-助数詞承接型接尾辞」と形態素解析され、文法構造1-2-3の順でマッチングする(マッチングアルゴリズムについては次章で詳細に説明する)。

2. 2 複合語キーワード評価ルール

複合語キーワード評価ルールとは、前節で抽出された複合語がキーワードとして扱われるか否かを判別するルールである。この評価ルールは、キーワード抽出ルール(表2)とキーワード削除ルール(表3)の2つに分類されている。

キーワード抽出ルールは、抽出された複合語をキーワードとして扱うルールである。例えば、対象文書内に「形態素解析用辞書を作成した」という文があるとき、「形態素解析用辞書」の後に続く「を」がキーワード抽出ルール2にマッチングするので、複合語「形態素解析用辞書」が、複合語キーワードとして抽出される。抽出ルール1、2、3、5は、ルールの前方で抽出された複合語を対象とし、抽出ルール4は後方で抽出された複合語を、抽出ルール6は前後の複合語をキーワードとする。

またキーワード削除ルールは、抽出された複合

表1 複合語文法構造

	文法構造	例
1	前置助数詞 - 数詞	約10、第八
2	数詞 - 後置助数詞	二本、50パーセント
3	後置助数詞 - 助数詞承接型接尾辞	Kg強、%台
4	数詞 - 助数詞承接型接尾辞	100未満、10以下
5	固有名詞 - 固有名詞	林淑隆
6	固有名詞 - 数詞	エリザベス2世
7	接頭辞 - 名詞	超音波
8	名詞 - 接尾辞	近代化
9	接頭辞 - 固有名詞	準オイラー法
10	固有名詞 - 接尾辞	ニュートン法
11	接尾辞 - 名詞	効率的抽出法
12	名詞 - 名詞	自然言語処理
13	固有名詞 - 名詞	ドイツ人
14	名詞 - 数詞	厚さ4センチ

表2 キーワード抽出ルール

	品詞	例
1	主格	格助詞 は、が
2	目的格	格助詞 を、に、で
3		格助詞相当句 に対して、に対する、をとおして
4	強調表現	指示連体詞の後 この、これらの
5		提題表現の前 についていえば、としては、と言えば、とは
6		提題表現の前後 における

表3 キーワード削除ルール

	品詞	例
1	接頭辞	各、諸、当該、本、両
2	接尾辞	やすい、全体、下、中、以外、等、間、内、時、後半、上、分、割、用
3	並列表現	と、や、あるいは
4	連体修飾語	「AのB」のAを削除する。但し、Bは用言性名詞ではない

語をキーワードとして扱わないルールであり、抽出ルールの逆の評価ルールである。このキーワード削除ルールは、木本[4]が提案した絞り込みアルゴリズムと、吉村ら[5]が提案した接辞辞書を基本に定義する。

削除ルール1、2は、抽出された複合語の先頭または末尾に該当する接頭辞または接尾辞がある場合、その接頭辞ないし接尾辞のみを削除する。例えば「各市長候補が...」という文章がある場合、抽出された複合語は「各市長候補」であるが、「各」を取り除き「市長候補」を複合語キーワードとする。削除ルール3は、並列表現の前後にある複合語はキーワードとして抽出しないルールである。削除ルール4は、連体修飾語を削除す

る。ただし被修飾語は用言性名詞ではない。例えば「明日の天気...」という文に対して抽出された複合語は「明日」であるが、「の」の後に続く「天気」が用言性名詞ではないので「明日」はキーワード候補から削除される。

またこれらのキーワード評価ルールは、次章で説明するストリングパターンマッチングマシンで使用されるため、新たにそれぞれの品詞を登録する。例えば「この」は「指示連体詞語」であり、「あるいは」は「並列表現」という品詞属性として登録する。

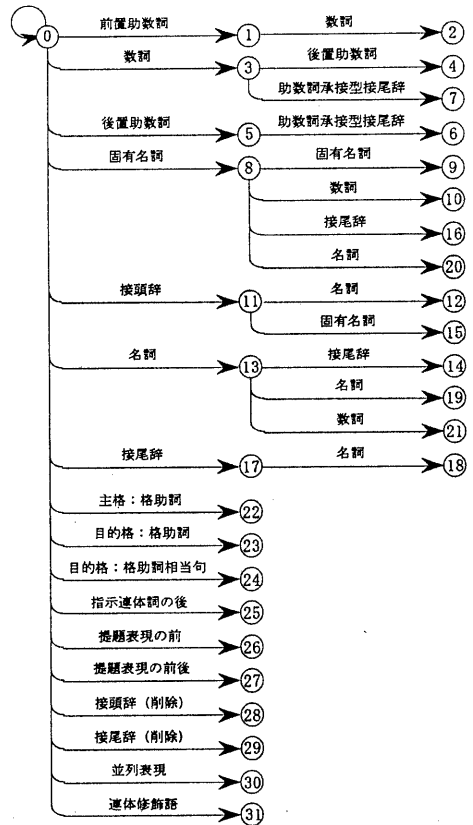
3. 複合語キーワード抽出アルゴリズム

アルゴリズムとしては大きく3つに分割される。第1は形態素解析である。第2は形態素解析の結果を入力とする複合語キーワード抽出部である。第3は抽出された複合語キーワード候補を入力とする複合語キーワード候補マシンACである。本論文での提案の中心は第2、3のマシンACに関するアルゴリズムなので、特に第2、3を詳しく説明する。

3.1 複合語キーワード抽出マシンACの構築

まず、複合語キーワードを抽出するためのストリングパターンマッチングマシン(マシンAC)を構築する。このマシンACは複合語文法構造(表1)と複合語キーワード評価ルール(表2、表3)から構成されており、マシンACの構築キーワードとして品詞コードを用いる。また新たに2つの関数を作成する。1つは複合語文法構造と複合語キーワード評価ルールを区別するためのrule_state関数である。もう1つはrule_state関数が評価ルールを示す場合にその複合語キーワード評価ルールを獲得するためのget_rule関数である。get_rule関数の返値はビット列で表現されており、次節のアルゴリズムにおける変数ruleに格納される。図1にこのマシンACの関数の一部を示す。

マシンACの構築に関しては、一般的なマシンACの構築と同様でありrule_state関数と

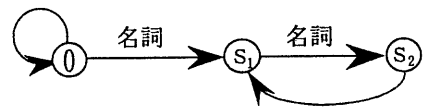


(a) Goto function

1. rule_state(s) = fail for 0 ≤ s ≤ 21,
2. rule_state(s) = true for 22 ≤ s ≤ 31.

(b) rule_state function

図1 複合語キーワード抽出マシンAC



(a) Goto function

s	f(s)	s	output(s)
S ₁	0	S ₁	{名詞}
S ₂	S ₁	S ₂	{名詞-名詞}

(b) Failure function(c) Output function

図2 連続した名詞列への対応

Algorithm 1. Pattern matching machine for compound keywords.

Input. A text string $T = t_1 t_2 t_3 \dots t_n$ where each t_i is an input word, $X = a_1 a_2 a_3 \dots a_n$ where each a_i is a part of speech.

Output. Compound Keywords.

Method.

```

begin
  state ← 0
  keyword ← empty
  queue ← empty
  rule ← no_rule
  for i ← 1 until n do
    begin
      while g( state , ai ) = fail do
        begin
          if state = 0 then
            begin
              if rule = behind then
                begin
                  print keyword
                  keyword ← empty
                  queue ← empty
                  rule ← mask behind
                end
              else keyword ← empty
                  state ← f( state )
            end
          end
        queue ← queue U { ti }
        state ← g( state , ai )
        if rule_state( state ) ≠ fail then
          begin
            rule ← get_rule( state )
            if rule = forward then
              begin
                print keyword
                keyword ← empty
                queue ← empty
                rule ← mask forward
              end
            end
          end
        if output( state ) ≠ empty then
          begin
            keyword ← keyword U queue
            queue ← empty
          end
        end
      end
    end
  end
end

```

Algorithm 1. 複合語キーワード抽出アルゴリズム

get_rule関数が拡張された状態となる。

また特に、連続した名詞列に対処するために、マシンACの構築に際してキー「名詞」と「名詞-名詞」を追加する(図2)。これにより状態番号 S_1 から状態番号 S_2 へのfailure関数が定義されるため、連続した名詞列を処理することが可能となる。この方法は連続した固有名詞列にも使用されている。

3. 2 複合語キーワード抽出アルゴリズム

前節で構築したマシンACを用いて、複合語キーワード抽出処理を行う。この複合語キーワードの抽出処理をするための手順をアルゴリズム1に示す。

このアルゴリズムの入力は、複合語キーワード抽出対象となるテキスト(T)と、テキストの形態素列(X)である。また出力は抽出された複合語キーワード集合(keyword)である。以下、複合語キーワード抽出アルゴリズムについて説明する。

このアルゴリズムでは、品詞列でパターンマッチングを行い、マッチングした品詞列に対応するテキスト文字列が出力される。そのため、従来のマシンACでoutput関数が行っていた出力の代わりに、keywordが出力される。

このkeywordには、抽出された複合語キーワード候補が格納されており、また最長の複合語キーワード候補を抽出するために、queueを使用する。queueはgoto関数によって状態が遷移するたびに、その状態遷移に対応したテキスト文字列を一時的に格納する。また、変数ruleはキーワード評価ルール情報を格納した変数である。

最初のwhile文では、従来のAC法と同様にfailure関数による状態遷移を実行する。しかしながら、前節で構築したマシンACは複合語文法構造とキーワード評価ルールから構築されている。状態遷移でstateが0に遷移するとき、文法構造にマッチングしないということを意味する。従って状態番号0に遷移したときには、keywordが空でなければそれまでの遷移で複合語キーワードが抽出されているので、変数ruleの後置抽出フラグが立っていればkeywordを複合語キーワード候補として出力する。そうでなければ複合語キーワード評価ルールにはマッチングしていないのでkeywordをクリアする。

次にgoto関数によって遷移を行い、対応するテキスト文字列をqueueの末尾に追加する。ここでrule_state関数によって遷移した先がキーワード抽出ルールを示す状態番号か否かを調べ、そうならばそのキーワード抽出ルール情報をget_rule関数により変数ruleに設定する。このとき設定した

抽出ルールの前置抽出フラグが設定されていれば、後置抽出のときと同様にkeywordを出力する。

また、現在の状態番号にoutput関数が存在していれば、複合語文法構造に部分的にマッチングしているので、一時的にqueueにためておいたテキスト文字列をkeywordの末尾に追加する。上述の手順を入力テキストの品詞列について最後まで繰り返し、抽出処理を終了する。

次に、入力テキスト「準硬式野球部OB会が明日開かれます。」を例にアルゴリズムを説明する。形態素解析の結果は図3となる。

最初の品詞は接頭辞であり、状態番号11に遷移する。ここでは、queueに「準」が追加され、keywordはまだ空である。次の品詞は名詞であり、状態番号12へ遷移する。queueには「硬式」が追加されて「準硬式」となる。また現在の状態番号にはoutput関数が定義されているので、現在のqueueの値をkeywordへコピーする。この結果queueは空になり、keywordは「準硬式」となる。次の品詞は名詞であるが、現在の状態番号から名詞の遷移が定義されていないため、failure関数によって遷移を行い状態番号13へ遷移する。この遷移によって名詞がマッチするので、queueには「野球」が追加される。次の品詞接尾辞は遷移が定義されているため、状態番号は14となり、queueは「野球部」となる。この状態番号にはoutput関数が定義されているので、先ほどと同様にqueueは空になりkeywordに「野球部」が追加され「準硬式野球部」となる。以下、同様に処理し、keywordは「準硬式野球部OB会」となる。

次の品詞は主格：格助詞であり、状態番号22へ遷移する。この状態番号のrule_state関数の返値はtrueであり、複合語キーワード評価ルールを示す状態番号であることがわかる。そこで、get_rule関数により複合語キーワード評価ルール情報を獲得する。複合語キーワード評価ルール「主格：格助詞」には前置抽出フラグが設定されているので、直ちに「準硬式野球部OB会」を複合語キーワード候補として出力する。そしてqueueとkeywordを空にして、前置抽出が終了したので変数ruleの前置抽出フラグをクリアする。

準--- [接頭辞]
 硬式 --- [名詞]
 野球 --- [名詞]
 部 --- [接尾辞]
 O B --- [名詞]
 会 --- [名詞]
 が --- [格助詞：主格]
 明日 --- [名詞]
 開か --- [カ行五段未然]
 れ --- [助動詞連用]
 ます --- [終止]
 。 --- [句読]

図3 例文の形態素解析結果

表4 抽出アルゴリズムの動作

	状態番号	keyword	queue	rule
1	11		準	
2	12		準硬式	
3	12	準硬式		
4	13	準硬式	野球	
5	14	準硬式	野球部	
6	14	準硬式野球部		
7	18	準硬式野球部	O B	
8	14	準硬式野球部	O B会	
9	14	準硬式野球部O B会		
10	22	準硬式野球部O B会		前置抽出
11	13		明日	
12	13	明日		
13	0			

次の品詞は名詞であるので、queueが「明日」となるが、output関数が定義されているためqueueは空になりkeywordが「明日」となる。続く品詞はカ行五段未然であるが、この状態遷移は定義されていないので、最初のwhile文内においてfailure関数で遷移することになる。このとき状態番号0に遷移するが、先ほど獲得した複合語キーワード評価ルール「格助詞：主格」には後置抽出フラグは設定されていないので、keywordは出力されずにクリアされる。

残りの品詞はいずれも定義されていないので、状態番号0から遷移することもなく、queueとkeywordは空のままである。

以上で抽出処理を完了し、この結果「準硬式野球部OB会」が複合語キーワード候補として抽出され、次節の複合語キーワード候補マシンACに送られることになる。表4にこの例におけるアル

ゴリズムの動作を示す。

3.3 複合語キーワード候補マシンAC

前節の複合語キーワード抽出アルゴリズムから出力された複合語キーワード候補は、候補マシンACへの追加キーワードとして処理される。候補マシンACのキーワード追加処理自体は一般的なマシンACを同じであるが、キーワードの頻度情報をもたせている。

また複合語キーワード候補の抽出がすべて終了した後に、他の複合語キーワード候補の部分文字列となっている複合語キーワード候補について、そのキーワード候補自体を削除し、包含している複合語キーワード候補の頻度情報をインクリメントする処理を行う。例えば、複合語キーワード候補に「自然言語処理」と「言語」が存在している場合、「言語」を削除し「自然言語処理」の頻度情報をインクリメントする。実際には、output関数が定義されている状態番号について、値が0でないfailure関数が定義されているものをすべて選び出し、output関数が示すキーワードを削除し、そのfailure関数が示す先にあるキーワードの頻度情報をインクリメントする処理を行う。

候補マシンACを利用することで、共有部分文字列をもつキーワードの絞り込みが極めて高速に行えるので、利用者はこの絞り込みのルール（共有部分文字列の併合、分割の基準）を与えることで、短単位中心あるいは長単位中心などの多彩なキーワードを最終的に決定することが可能となる。従来のようにキーワード文字列を列挙する方法では、改めて部分文字列の照合・併合・分割を必要とするので、キーワード数が多くなると多大な時間を要していた。

4. 評価

本システムは、形態素解析、複合語キーワード抽出マシンAC、及びキーワード候補マシンACで構成されており、C言語でインプリメントした。実験に使用した計算機は、PC-9821As2(i486SX 33MHz)である。現在、形態素解析部は既存の形態素解析モジュールを使用しており、その出力結

果の一部を2章で定義した複合語評価ルールに変更して使用している。マシンACは、遷移をリスト構造により構築した。また複合語キーワード候補マシンACも同様にトライ法により構築した。

次に本手法の有効性を確認するため、複合語キーワードの抽出速度とその抽出精度の2点に関する評価を行った。なお、今回の実験では、14個の複合語文法構造と10個の複合語キーワード評価ルールを使用した。また実験対象となる文書には、一般文書、論文、特許に関する比較的短い解説文の内から26文書を選んだ。実験結果を表5に示す。

まず複合語キーワードの抽出速度に関しては、形態素解析の時間が、辞書構造、解析手法により変化するので、この処理時間を除く複合語キーワード抽出マシンACによる抽出処理時間と複合語キーワード候補マシンACの構築時間の和を求めた。この結果、平均で処理時間は16.58ミリ秒、

表5 実験結果

	文書名	文書量 (バイト)	形態素数	処理時間 (ミリ秒)	抽出キーワード数
1	一般1	1,543	468	80	31 (10)
2	一般2	1,367	376	80	33 (14)
3	一般3	2,413	667	150	49 (15)
4	一般4	2,413	746	210	65 (17)
5	一般5	2,387	696	150	53 (13)
6	一般6	2,062	556	110	40 (9)
7	一般7	1,849	515	130	48 (15)
8	一般8	1,783	471	100	32 (16)
9	論文1	1,292	304	60	20 (10)
10	論文2	532	133	30	15 (5)
11	論文3	1,064	305	80	29 (15)
12	論文4	1,125	262	60	27 (10)
13	論文6	1,098	314	80	30 (18)
14	論文8	859	198	40	14 (8)
15	論文9	1,055	289	60	24 (13)
16	論文10	796	241	50	23 (14)
17	論文12	984	246	50	17 (9)
18	論文13	15,835	5,084	1,200	192 (102)
19	論文14	17,505	4,283	1,040	185 (92)
20	論文15	9,330	2,326	340	65 (23)
21	特許3	543	164	30	13 (4)
22	特許7	701	168	30	14 (9)
23	特許9	639	170	30	12 (5)
24	特許11	715	192	40	15 (6)
25	特許12	774	202	40	9 (4)
26	特許23	729	205	40	15 (10)

また文書 1 K B あたりにすると 6.18 ミリ秒という速度が得られた。

また抽出精度に関しては、実際に抽出されたキーワードと対象文書のキーワードとして定義されているものを比較した。本手法で抽出されたキーワードの総数は、表 5 の抽出キーワード数に記述している。また括弧内の数字は、抽出キーワードのうち 2 単語以上からなるキーワードの総数である。実験結果より本手法は、ほぼ対象文書のキーワードを約 20% 包含している状態となっている。これは、さらにキーワードとしての絞り込みを強化しなければならないことを意味している。表 5 の実験結果の抽出キーワード数は、複合語キーワード候補マシン A C がもっているキーワードの総数を記述しており、実際にキーワードとして有効なものは、マシン A C が持つ頻度や重み情報を参照することでさらに絞り込むことが可能である。その一例が括弧内の数字である。

また本手法は動的なストリングパターンマッチングマシンを使用していることで、複合語文法構造やキーワード評価ルールを、簡単かつ高速に追加変更することが可能である。このことは、さらに有効な文法構造や抽出ルールを使用することで、抽出精度の結果も向上し得ると思われる。また形態素解析部をよりキーワード抽出に向けたものに変更することも、抽出精度の向上に関与すると思われる。

5. まとめ

本論文では、複合語キーワード抽出においてマッチング処理が多大な工数を必要とすることに着目し、複数キーワードのストリングパターンマッチングマシンの手法を応用した、より効率的な複合語キーワード抽出法を提案し、およびこれに用いる複合語文法構造と複合語キーワード評価ルールを定義した。また 19,581 語 26 文書のデータに対する実験結果から、本手法による複合語キーワードの抽出速度の有効性を実証した。

先に述べたように、本手法は従来法ほどの抽出精度を得るに至っていない。より有効的な文法構造や評価ルール、形態素解析部を今後の課題とし

て研究を進める計画である。

文 献

- [1] 津田和彦、入口浩一、青江順一：「ストリングパターンマッチングマシンの動的構成法」, 信学論(D), J77-D-INo.4, 1994
- [2] 宮崎正弘：「係り受け解析を用いた複合語の自動分割法」, 情処論, Vol.25, No.6, pp.970-979, 1984
- [3] 小川泰嗣、望主雅子、別所礼子：「複合語キーワードの自動抽出法」, 自然言語処理, 97-15, 1993-9-17
- [4] 木本晴夫：「日本語新聞記事からのキーワード自動抽出と重要度評価」, 信学論, D-I, Vol.J74-D-I, No.9, pp.556-566, 1991
- [5] 吉村賢治、日高達、吉田将：「日本語科学技術文における専門用語の自動抽出システム」, 情処論, Vol.27, No.1, pp.33-40, 1986
- [6] 伊藤哲、丹羽寿男、萱島一弘、丸野進、木泰治：「利用目的に応じて最適化可能なキーワード抽出手法」, 電気情報通信学会, 信学技報, NLC93-53, 1993-12)
- [7] 中園薫、白井諭：「日本語牽引自動生成システム」, 情報処理学会, 自然言語技術シンポジウム(1984-11)
- [8] 会森清、依田透、嵩原哲：「日本語キーワード抽出システムの開発および今後の課題」, ドキュメンテーション・シンポジウム予稿集, pp.15-19, 1988
- [9] 山口義一、杉山時之：「自然言語による索引語自動抽出システムの概要とその索引語の分析」, 科学技術文献サービス, No.85, pp.31-40, 1988
- [10] 細野公男、後藤智範、諸橋正幸：「パターン・マッチングによる重要語の自動抽出」, 情処学会 N L 研会, Vol.39, No.1, pp.1-8, 1983