

## 文意解析に基づく有意味キーワード抽出システム\*

高橋昭治<sup>(1)</sup>、高橋真一<sup>(1)</sup>、安田直義<sup>(2)</sup>、高畑信仁<sup>(2)</sup>、石川徹也<sup>(3)</sup>

(1) 株式会社伊國屋書店 (2) 日外アソシエーツ株式会社 (3) 図書館情報大学

### 概要

日本語テキストを対象に、事前設定による大規模キーワード・リストを用いキーワードを抽出するシステムに対し、有意味なキーワードを抽出する為に、文意解析処理を行い、その後大規模キーワード・リストとの照合を計りキーワードを設定するシステム機能について報告する。システムの評価実験の結果、単に大規模キーワード・リストを用いる従来の方式に対し、抽出キーワードの有意味適合率が約8.1%向上することの結果を得た。

### A Meaningful Keywords Extracting System based on A Sentence-Semantic Analysis Method

Shoji Takahashi<sup>(1)</sup>, Shin,ichi Takahashi<sup>(1)</sup>, Naoyoshi Yasuda<sup>(2)</sup>,  
Nobuhito Takahata<sup>(2)</sup>, Tetsuya Ishikawa<sup>(3)</sup>

(1) Kinokuniya Company, Ltd. (2) Nichigai Associates, Inc.  
(3) University of Library and Information Science

### Abstract

This paper describes an indexing system that attempts to extract meaningful keywords from Japanese texts based on a sentence-semantic analysis method before matching against a large-scale keyword-list, as compared with conventional systems that use only pre-determined large-scale keyword-list. Our experiment has shown that the relevance rate of extracted keywords is about 8.1% higher with our method than with the conventional method.

\*当研究は、財団法人データベース振興センターの平成2年度および平成3年度データベースの構築促進および技術開発促進事業の委託事業課題の一つとして実施した。

## 1. はじめに

文書、論文等テキストを検索する場合、書誌データ（例：著者名、標題、出版者名等）を対象に検索する場合と、テキストの内容を対象に検索する場合とがある。テキストの内容を対象に検索することを可能にする為に、現行では、検索性データとして、テキスト内容を紹介する索引語が用いられている。

索引語は、従来から、索引語作成者がテキスト内容を解説し、ソーラス用語（統制索引語という）を用い設定してきている（当方式を、統制索引語索引方式という）。しかし、大量のテキストに対し、索引語作成者による索引語作成には限界があり、情報提供に対するタイム・ラグ、索引語の質の揺れ等の問題点が生じ、検索精度に影響を及ぼしている。当問題回避の為に、テキストの内容を自動解析し、テキスト内に出現する単語を対象にキーワードとして抽出し、索引語とするキーワード自動抽出システムが利用されるようになってきている。

現在、実用に供されているキーワード自動抽出システムを大別すると次の2方式のものがある。

①事前設定による大規模索引語候補語リストを用い、テキスト内に出現する単語との照合によりキーワードを自動抽出するシステム。

②検索語に対し、テキスト内に出現する単語との照合により検索するシステム（フルテキスト検索システムと呼ぶ）<sup>1)</sup>。

しかし、いずれのシステムもテキスト内に出現する単語との照合を前提にしていることから、文意に関係なく抽出する為、再現率は保証されるが、適合率は保証されない欠点がある。適合率を保証するには、テキスト内の文意解析を行い有意なキーワードを抽出する必要がある。

そこで、本稿において、有意なキーワードの自動抽出を目的に、フルテキストに対し文意解析を行い主題表示キーワードもしくはキーセンテ

1) テキスト・データベース（以下、データベースをDBと記す）には、下記の2種類が存在する。書誌データと索引語を蓄積した書誌DBとフルテキストDBとがある。このことに対し、索引語の設定方式に、下記の2方式がある。テキストDBの構築に対し、検索性データ（主に索引語）を事前に設定する方式（事前設定方式と呼ぶ）と検索指示の都度フルテキスト内からキーワードを抽出し、利用する方式（検索時設定方式と呼ぶ）とがある。

ンスを事前に認識し、その後上記①の実用システム（日外アソシエーツ<sup>2)</sup>製作のNICEシステム）を基に、キーワードを抽出するシステムについて報告する。

以下、2.でキーワード自動抽出システムについて従来の方式をレビューし、問題点について検討を行い、3.でシステム機能について示し、4.で本システムの評価実験結果を示し、5.で本システムの有効性について考察を行う。

## 2. 従来の方式と問題点

キーワード自動抽出法について従来の方式をレビューし問題点を検討する。

### 2.1. 索引語の機能

索引語機能に以下2点がある<sup>1)</sup>。

- 1) 構造化索引語 (String index term) …事前に設定した索引項目（例：実験材料、実験方式等）に相当する索引語<sup>2)</sup>。ファセット索引語ともいう。
- 2) 主題索引語 (Subject index term) …テキスト内容の主題、即ち、執筆者のテキスト作成主題意図を表示する索引語<sup>3)</sup>。

### 2.2. 従来の方式と問題点<sup>4, 5, 6, 7, 8)</sup>

#### 2.2.1. 事前設定方式

##### i) 出現頻度解析抽出法

##### ii) 単一キーワード抽出法

テキスト内に出現する単語の出現頻度解析を行ない、出現頻度の事前設定値に対し対応する単語をキーワードとする方式。当方式は、Luhnにより提案された<sup>9)</sup>。当方式は、機能語等の単語を意味的に識別することが不可能なことから、機能語等をストップワード・リストとして事前に設定しておき、ストップワードを除く残り全ての単語をキーワードとするストップワード方式により実用化された。しかし、当方式には、例えば英文において複合語の判定が不可能になり欠落が生じる欠点があることから、最近では、下記2)に示すキーワード・リストを事前に設定し利用するキーワード方式が用いられている。日本において、ストップワード方式について二村らによるシステムがある<sup>10)</sup>。

##### iii) キーワード間相関抽出法

複合語判定の為に、テキスト内に出現するキーワード間の相関頻度を基に解析を行い、相関頻度の事前設定値に対し全ての組み合わせを複合語として認知する方法。この方式は、Stilesらにより提案された<sup>11)</sup>。当方式は、テキスト内に記載さ

れていない複合語をも生成してしまうことから、精度に問題があり実用に至っていない。

## 2)大規模キーワード・リストによる抽出法

複合語を含め大規模キーワード・リストを事前に設定しておき、テキスト内に出現する単語を対象に照合し、その単語をキーワードとする方式。当方式は、電算機のファイル容量の発展により、例えば、同義異形語についても維持することが可能になり実用可能になった。当方式は、登録キーワードの網羅性を保証する為に、キーワード・リストのメンテナンスを常に行う必要がある。日本による実用例に、雑誌記事索引作成<sup>12)</sup>、新聞記事DB構築<sup>13)</sup>等がある。

## 3)文構造解析による抽出法

自然言語処理を基にテキスト内の文構造解析を行い抽出する方式。日本語資料を対象に、例えば、形態情報に基づき抽出する吉村ら<sup>14)</sup>、形式動詞「である」に対応し抽出する柴田ら<sup>15)</sup>、字種により抽出する松崎ら<sup>16)</sup>、「格」に対応し抽出する木本<sup>17)</sup>等の研究がある。しかし、いずれも抽出設定条件に合う出現単語を対象に抽出することになり精度面に依然問題点を持つ。このことに対し、石川は有意味なキーワード抽出を目的に動詞の意味を基に抽出する方式を提案している<sup>18)</sup>。以上の方式に対し、報道記事を対象に5W1Hに相当するキーワードに限定し抽出する絹川らの実用システム<sup>19)</sup>およびテキスト内容の構造化表現を目的とするMoritaらの研究<sup>20)</sup>がある。

## 2.2.2.検索時抽出方式（フルテキスト検索方式）

### 1)検索指示内容を直接利用する方式

検索語と並び例えば関連テキストの標題を入力し、その中のキーワード抽出を行い、検索語と共にテキスト内に出現する単語との照合により検索を行う方式。この方式は、SaltonによりSMARTシステムとして提案された<sup>21)</sup>。当方式は、フルテキスト検索方式を可能としたが、照合処理を統計的解析に基づき行うことから依然精度面に問題があり実用に至っていない。しかし、現在は、検索語を基にテキスト内に出現する全ての単語との形態一致照合により検索する方式が実用化されている<sup>22)</sup>。

### 2)シソーラスを利用する方法

シソーラスを事前に設定しておき、検索語とシソーラス用語との照合を計り、その検索語を含むカテゴリ内の用語全てを基に検索する方式。この方式の例として、英文テキストを対象とするMETAMORPH（Thunderstone/EPI社）がある<sup>23)</sup>。当方

式は、カテゴリ内の用語の範囲内において関連情報の検索が可能である。しかし、原理的には上記の大規模キーワード・リストを用いる方式と同じであり、シソーラスのメンテナンスの課題がある。

## 3. システム機能

### 3.1.システムの構成

有意味な主題表示キーワードを抽出する為のシステムは下記機能からなる（図1）。

①有意味キーワード抽出システム…主題表現動詞を中心、テキストの記述文からキーワードを抽出するシステム機能。我々は、このシステム機能をダイナミックシソーラス・エンジン（以下DTEと呼ぶ）と名付けた。

②キーワード抽出システム…抽出キーワードを対象に、キーワード・リストを基にキーワードを抽出するシステム機能。

尚、検索システムにおいて、質問文に対しても、上記機能を適用し、質問文検索方式を実現する。

本研究では、キーワード抽出システムとして日外アソシエーツ(株)が開発したデータベース用語管理システムNICEを使用した。NICEは、ストップワード方式によるキーワード抽出機能と辞書方式によるキーワード抽出機能を持っている。本研究では、NICEのキーワード抽出機能を利用した。

以下、有意味キーワード抽出機能を中心に紹介する。但し、本システムは、テキストDBとして、東販(株)、日販(株)、(株)紀伊國屋書店、および日外アソシエーツ(株)が制作し、日外アソシエーツ(株)がオンライン検索システムASSISTでサービスしている図書情報データベース「BOOK」の要旨に対し適用することを目的にしているの、その範囲内での機能化である。

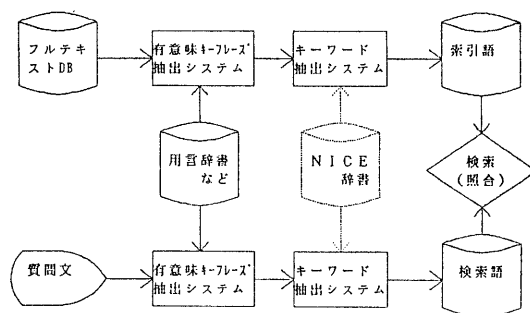


図1：システム機能図

● #3 ISBN4-06-153801-2 【B9110025】

【標題】 言語処理と機械翻訳 野村浩郷編講談社 '91.5.20211p 21cm(A5) ¥3,800(本体¥3,689)

【要旨】 機械翻訳の技術を成熟させるにはまだ多くの課題が残されているが、過去40年にわたり検討してきた技術を統合して機械翻躍システムとして実現する時期にきている。本書は、そのような観点に立ち、現在の機械翻訳の技術と将来の高度な技術を開発するための問題点を述べる。

【内容】 1章 機械翻訳の概要;2章 言語処理の基本技術;3章 機械翻訳システムにおける辞書の構成;4章 概念表現と機械翻訳;5章 意味理解と機械翻訳;6章 文明と意味と機械翻訳

図2：データ例（以下データ例を#No.で表現）

### 3.2. 使用データ

BOOKは、標題・要旨・内容の3部分から構成されている。要旨は、主に本の帯やジャケットなどから採ったもので、内容は本の目次である。本研究では、要旨の部分を対象にした。

BOOKの累積データ数は、1991年度で約20万点である。

本研究では、情報処理分野の要旨付きデータ134件を使用した。

### 3.3. システム機能の概要

#### 3.3.1. キーワード抽出機能

最初にBOOKの要旨に対してDTEを適用し、その出力結果に対してキーワード抽出システムを適用する。DTEは下記3.4に示すキーフレーズを切り出す。キーワード抽出システムは単語切り出し機能を使用してキーフレーズからキーワードを抽出する。以下、データを基にこの流れを説明する。

#126【標題】 LISPによる自然言語処理

【要旨】 本書はAIのコンピュータ言語ともいえるべきLISPを用いた自然言語処理を出来るだけやさしく解説することを目的としている。したがって、LISPの関数は出来るだけ基本的なものを少数用いるようにした。

DTEは、要旨の中から、「本書は」と波線部の表現動作動詞「目的としている」に基づき、実線部をキーフレーズとして抽出する。次に、キーワード抽出システムは、DTEが抽出したキーフレーズの中から以下のキーワードを抽出する。

AI、コンピュータ言語、LISP、自然言語処理

これに対し、DTEを使用しない場合、単独のキーワード抽出システムでは以下の用語を抽出する。

本書、AI、コンピュータ言語、LISP、自然言語処理、関数、少数

この要旨に対して、人間が真のキーワードとして判定できるのは「LISP」と「自然言語処理」である。したがって、当システム機能は、単独のキーワード抽出システムよりも有意義なキーワードを抽出する確率が高いと言える。

#### 3.3.2 検索機能

検索者が自然文で入力した質問文に対しても、要旨の場合と同様の処理を行ない、質問文からキーワードを抽出し、BOOK内抽出キーワードとの照合を行い、検索する。

### 3.4 キーフレーズ抽出ルール

DTEのキーフレーズ抽出ルールは、大きく3グループに分かれている。先に記したグループを優先し、また各グループ内でも先に記したルールを優先する。各センテンスに対して1つのルールが適用されたら、それ以後のルールは適用しない。

注1：各ルールの中で、下線部( )をキーフレーズとする。

注2：表現動作動詞とは、動詞辞書でACTADEコードを持つものである。

注3：以下の例で、( )で囲まれた部分がキーフレーズとして抽出される。

#### 第1グループ(本の要旨に特有なルール)

ルール11 \*\*\* +        + である/です/にある +

文末

\*\*\*のリスト例：本書は/本書では/本著作は/本特集は/本講座は/本号は/目的は/内容は/本書のねらいは/この辞典で

#24【標題】 図解 人工知能入門

【要旨】 本書は、{人工知能に関する初心者向けの解説書}です。特に、図を多くし、文章をやさしくして、初心者にもわかりやすいように心がけています。

ルール12 \*\*\* + 〜 + 表現動作動詞 + 文末  
\*\*\*のリスト例：本書は／本書では／この本では  
／本書の理論は／このマニュアルには

#3 【標題】言語処理と機械翻訳

【要旨】機械翻訳の技術を成熟させるにはまだ多くの課題が残されているが、過去40年にわたり検討してきた技術を統合して機械翻訳システムとして実現する時期にきている。本書は、{そのような観点に立ち、現在の機械翻訳の技術と将来の高度な技術を開発するための問題点を}述べる。

ルール13 〜 (前全体) + \*\*\* + 文末

\*\*\*のリスト例：論文集／論／エッセイ集／書  
／ハンドブック／講座／意欲作／特集／入門  
／一冊／読本／作品／入門書／著述／名著  
／招待／最前線／提唱／焦点／対話／鳥瞰図

#51 【標題】アクティブ・マインドー人間は動きのなかで考える

【要旨】{身体と思考のみごとな相互関連を解きあかし、認知科学の新しいパースペクティブを示した意欲作}。

ルール14 〜 (前全体) + は/とは + なにか/  
何か + 文末  
〜 (前全体) + とは + 文末

#64 【標題】岩波講座 転換期における人間 6  
科学とは

【要旨】{人間にとって、社会にとって、科学}とはなにか。生き方と価値観の指針を提示する。

ルール15 〜 (前全体) + 動詞 (連体形) +  
か/のか/のだろうか + 文末

#167 【標題】応用言語学

【要旨】{言語学の成果はどのように}応用されるのか。子供の言語発達のしくみから説き明かし、2言語併用、翻訳理論、自動翻訳の問題を探究、国語・外国語の教育方法にも言語学の応用を説く。

第2グループ (表現動作動詞に関する一般ルール)

注1：動詞の終止形と連用中止形は、全て有効とするが、連体形と連用形は、特定の句接続詞につながるもののみを有効とする。

連体形：が／ために／ための／とともに／よう

連用形：つつ／つつも／ながら

ルール21 〜 (前の動詞まで) + 助詞 (関連格／主格) + 表現動作動詞 (受動態)

#111 【標題】学習の理論 下

【要旨】本書のねらいは現代の学習理論を理解させ、その歴史的脈とその背景とを提供することにある。下巻では理性主義による認知一体系化の考え方に影響された{ゲシュタルト派やトールマンの学習理論や情報処理理論とその発展}について述べられ、さらにこれらの理論と関連する神経科学や教育の応用にも及んでいる。

ルール22 〜 (前の動詞まで) + 助詞 (疑問格) + 表現動作動詞 (「説明」に近い概念を表わすもの)

#124 【標題】メンタルモデルー言語・推論・意識の認知科学

【要旨】著者は、本書で認知科学の研究のやり方の一つを示している。{実験心理学、論理学、言語学、人工知能、そして哲学からのアイデアと方法とをいかに統合する}かを教えてくれる。とりわけ、計算理論は、心の複雑さを理解するのに必要な決定的な方法とメタファーのいくつかを提供する。本書は、広く科学に興味を抱く読者を念頭に書かれたものである。

ルール23 〜 (前全体) + 動詞 (連体形) + 抽象名詞 + 助詞 (関連格／目的格) + 表現動作動詞

#188 【標題】コアとプロトタイプー基本動詞の意味論

【要旨】takeやgetなどの基本動詞の意味は多岐に亘り、相互の関連性が截然と把えにくい。{本書では、これをコアとプロトタイプの概念を組み合わせ}整理する理論を提案し、検証を経てかなり成功している。意欲あふれる研究書である。

ルール24 〜 (前の動詞まで) + 助詞 (関連格／目的格) + 形式名詞 + 抽象名詞 + 助詞 (形式格) + 形式動詞 + 抽象名詞 + 助詞 (始点格) + 表現動作動詞

#85 【標題】神経回路モデルとコネクショニズム  
【要旨】{今、認知科学でもっともホットな脳の並列分散処理の可能性と限界を}その基礎となる原理からていねいに解説。

ルール25 ～ (前全体) + 表現動作動詞 (終止形、  
前方に 関連格/目的格の助詞なし)

#134【標題】岩波講座 ソフトウェア科学 8 記号  
処理プログラミング

【要旨】人工知能や知識処理のソフトウェア  
を作成するうえで必要となる記号処理のプ  
ログラム技法を系統的に紹介する。{例題はプ  
ログダクションシステム、自然言語の構文解析、  
定理の自動証明などの実際の人工知能システ  
ムからとり、プログラム作成の手順を追って  
詳しく} 解説した。また、LISPやPROLOGは、  
おのずとマスターできるように工夫した。

ルール26 ～ (前の動詞まで) + 助詞 (関連格/  
目的格) + 表現動作動詞

#159【標題】比喩と理解

【要旨】生きた実例を豊富に駆使して、{比  
喩を通じた認知のメカニズム}を詳細に分析。  
本格的比喩研究の書。

### 第3グループ (括弧に関するルール)

注1: 以下の括弧の中をキーフレーズとする。

< > 《 》 「 」 『 』 “ ”

#147【標題】意味と情報

【要旨】情報なるものが単に大量に氾濫して  
いるだけでは、人々間のコミュニケーションも、  
社会的な統合も促進されるとは限らない。  
情報は適切な意味を担って適切な場で伝  
えられるのでなければ役に立たない。つまり、  
情報にとっては、単に形式的な量だけでなく、  
その『{意味}』と『{文脈}』が重要である。  
そこでそういう観点から、情報というものを  
見ようというのが、この一連の共通のテーマ  
であり、全体をまとめて『{意味と情報}』と  
名づけた次第である。

#### ◎品詞リスト

- 助詞 (関連格) : について についても  
                  については に関して
- 助詞 (目的格) : に (動詞「光をあてる」  
                  「焦点を当てる」に対して)  
                  を (その他の動詞に対して)
- 助詞 (主格) : が
- 助詞 (疑問格) : かを なのかを のかを
- 助詞 (形式) : と に
- 助詞 (始点格) : から より
- 形式動詞 : なる ある

- 抽象名詞 : 研究 基礎 原理 こと 意味理論
- 形式名詞 : その この それらの これらの  
                  それ これ それら これら

#### ◎動詞辞書の構成

使用データに現れた全ての動詞は、次の情報を  
付与された形で動詞辞書にリストされている。

- ①基本概念: 表現動作動詞 (ACTADE)  
                  その他 (空欄)
- ②時制: 過去 現在 未来 語幹
- ③様態: 能動 受動
- ④アスペクト: 肯定 否定 仮定 継続 可能  
                  希望 義務
- ⑤活用形: 終止形 連体形 連用形 連用中止形

## 4. 評価実験

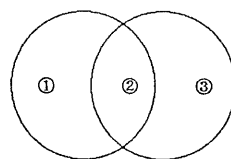
### 4.1. 評価方法

以下の3つのキーワードを比較する。

- 1) 人間が付与したキーワード
  - 2) 単独のキーワード抽出システムによる抽出キ  
ーワード
  - 3) 当システムによる抽出キーワード
- 評価は、1)と2)、1)と3)をそれぞれ比較すること  
によって行なう。1)、2)、3)の3者の関係は、次  
のようにまとめることができる。

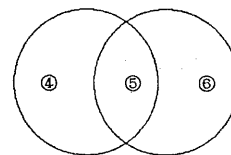
#### 1)と2)の関係

人間が付与した      単独のキーワード抽出シ  
キーワード              テムが抽出したキーワード



#### 1)と3)の関係

人間が付与した      当システムが抽出した  
キーワード              キーワード



②、③、⑤、⑥の数をカウントし、キーワード抽  
出システムが抽出したキーワードの適合率と当シ  
ステムが抽出したキーワードの適合率を計算する。

|            | 単独の<br>キーワード抽出システム   | 当システム  |
|------------|--|--|
| 適合率<br>(%) | $\frac{\textcircled{2}}{\textcircled{2} + \textcircled{3}} \times 100$ | $\frac{\textcircled{5}}{\textcircled{5} + \textcircled{6}} \times 100$ |

#### 4.2. キーワードの照合法

照合法には、形態一致と意味的一致がある。形態一致は、表記が完全一致する場合であるのに対して、意味的一致は、表記は異なるが、指示内容が同じ場合である。

本研究では、形態一致のみとし評価した。

#### 4.3 評価結果

②、③、⑤、⑥の数はそれぞれ、384、1036、278、515であった。上の式に基づいてキーワード抽出システムと当システムの適合率を算出する。

キーワード抽出システムの適合率

$$\frac{\textcircled{2}}{\textcircled{2} + \textcircled{3}} \times 100 = \frac{384}{384 + 1036} \times 100 = 27.0 \%$$

当システムの適合率

$$\frac{\textcircled{5}}{\textcircled{5} + \textcircled{6}} \times 100 = \frac{278}{278 + 515} \times 100 = 35.1 \%$$

単独のキーワード抽出システムの適合率(27.0%)よりも当システムの適合率(35.1%)の方が大きいことから、当システムは、単独のキーワード抽出システムよりも、35.1% - 27.0% = 8.1%の分だけ有意なキーワードを効率的に抽出できると言える。

#### 4.4. 問題点と対策

当システムのもっとも大きな問題点は、DTEが抽出するキーフレーズの範囲が広すぎて、無駄な部分を含んでいることが多いことである。キーフレーズの中からより重要な部分のみを抽出するように、ルールを強化する必要がある。

例えば、データ #28は第1グループのルール11によって要旨の全体がキーフレーズとして切り出されている。しかし、波線部(〰)の表現動作動詞に注目し、さらに細かな分析を進めることによって、実線部(—)の真のキーワードをより効果

的に抽出できる可能性がある。

#28 【標題】 『意味』の地平へーレヴィ・ストロース、柳田国男、デュルケーム

【要旨】 {主としてレヴィ・ストロースの構造主義理論を現代的な視点から平明に解説し柳田国男、デュルケームとの意外性と示唆に富んだ知のトリアーデを構成するなかで人が生きる『意味』をそして世界の『意味』を探りだそうとする知的冒険の書}。

データ#24では、第1グループのルール11によって{ }内がキーフレーズとして切り出されているが、「に関する」という助詞を鍵にして、真のキーワード「人工知能」だけを切り出すルールの可能性を探る必要がある。

#24 【標題】 図解 人工知能入門

【要旨】 本書は、{人工知能に関する初心者向けの解説書}です。特に、図を多くし、文章をやさしくして、初心者にもわかりやすいように心がけています。

#### 4.5. 欠落キーワードの原因と対策

人間が付与したキーワード(②または④+⑤)が384であるのに対して、当システムが抽出したキーワード(⑤)は278である。つまり、384 - 278 = 106のキーワードが欠落していることになる。

現在のルールは、動詞の基本概念として表現動作動詞のみを考慮している。表現動作動詞が真のキーワードを抽出するのに有効であることは確認できたが、表現動作動詞を含まない要旨も存在する。表現動作動詞以外の動詞を整理し、相対的に重要なキーフレーズを切り出すための新たなルールを追加することが必要である。

また、BOOKの要旨の中には、連体止めの文が多い。第1グループのルール13でも、主に書籍に関する連体止めの表現を扱っているが、その他の連体止めに関するルールを作る必要がある。

#### 5. おわりに

キーワード・リストのみによる従来のキーワード抽出システムでは全ての単語を抽出するが、これらの単語の全てが目的とするキーワードとは必ずしもならない。検索精度を上げるためには、人間による索引語作成作業のように、文意に基づいて主題該当語を抽出する必要があり、我々はこの抽出機能を「ダイナミック・シソーラス・エンジ

ン」)として実現した。このことによって我々は、適合率を8.1%上げることができるという成果を見た。だが当開発実験の結果、4で述べたような幾多の問題点も判明したので、今後の課題としたい。

#### 参考文献

- 1) Cleverland, D.B. and Cleverland, A.D.: Information to Indexing and Abstracting, p.209, Libraries Unlimited, Inc.(1983).
- 2) Craven, T.C.: String Indexing, p.246, Academic Press, Inc.(1986).
- 3) Dym, E.D., ed.: Subject and Information Analysis, 498p, Marcel Dekker, Inc.(1985).
- 4) Stevens, M.E.: Automatic Indexing: A State-of-the-Art Report, p.290, National Bureau of Standards(1965).
- 5) 石川徹也: 索引語の定量分析実験、図書館短期大学紀要、Vol.7, pp.43~81(1973).
- 6) 諸橋正幸: 自動索引付け研究の動向、情報処理、vol.25, No.9, pp.918~925(1984).
- 7) Martinez, C., et al: An Expert System for Machine-Aided Indexing, J.Chem.Infor.Comput. Sci., Vol.27, No.4, pp.158~162(1987).
- 8) Humphrey, S.W.: A Knowledge-Based Expert System for Computer-Assisted Indexing, IEEE Expert, Vol.4, No.3, pp.25~38(1989).
- 9) Luhn, H.P.: A Statistical Approach to Mechanized Encoding and Searching of Literary Information, IBM J., Oct., pp.309-317(1957).
- 10) 二村祥一、松尾文碩、英文科学技術文献情報に対する不要語除去法による自動索引、情報処理学会論文誌、Vol.28, No.7, pp.737-747(1987).
- 11) Stailes, H.E.: The Association Factor in Information Retrieval, J.ACM, Vol.8, No.2, pp.271~279(1961).
- 12) 山口義一、杉山時之: 国立国会図書館の雑誌記事索引システムにおける自然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、Vol.32, No.4, pp.31~40(1988).
- 13) 神尾達夫: 新聞記事データベースにおけるキーワード自動抽出、情報管理、Vol.32, No.4, pp.283~293(1988).
- 14) 吉村賢治、他: 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、情報処理学会論文誌、Vol.27, No.1, pp.33~40(1986).
- 15) 柴田浩一、他: 科学技術文献からの専門用語情報の自動抽出、情報処理学会全国大会講演論文集、Vol.35, No.2, pp.1283~1284(1987).
- 16) 松崎浩一、他: 日本語名詞の自動抽出について、九州大学工学集報、Vol.60, No.3, pp.293~298(1987).
- 17) 木本春夫: キーワード自動抽出の重要度評価、情報処理学会研究報告、Vol.87, No.64, pp.1~8(1987).
- 18) 石川徹也: 文意解析処理に基づく主題索引語作成支援システム、情報処理学会論文誌、Vol.32, No.2, pp.220~228(1991).
- 19) 絹川博之、他: 日本語情報検索システムにおけるキーワード自動抽出、日立評論、Vol.64, No.5, pp.25~38(1989).
- 20) Morita, Y., et al: An Indexing Scheme for Terms using Structural Superimposed Code Words, ICOT 研究論文、No.383, pp.1~9(1988).
- 21) Salton, G. and McGill, M.J.: Introduction Information Retrieval, p.448, McGraw Hill International Book Co.(1983).
- 22) 根岸正光: 解説 フルテキスト・データベースの応用動向、情報処理、Vol.33, No.4, pp.413~420(1992).
- 23) 米田健二: メタモルフ (METAMORPH) ーテキストを理解する人工知能ソフトウェア、オンライン検索、Vol.10, No.3, pp.117~126(1989).