文意解析処理に基づく主題索引語作成支援システム†

石川 徹也††

本研究の目的は，主題索引語作成支援（MAI：Machine-Aided Indexing）システム機能の開発にある．主題索引語は，“何々について報告する，あるいは何々について述べる”等，執筆者による資料の作成意図を表現する動詞に導かれる単語が相当する．そこで，資料の作成意図を表現する動詞を認識するための文意解析システム，その基で対象となる単語を主題表示キーワードとして抽出，あるいは主題表示キーワードが省略され存在しない場合に主題表示キーセンテンスとして抽出し出力するシステム機能の開発を行った．当システム機能の開発研究を行うために，日本語による科学技術論文の著者抄録を対象に解析を行い，システム機能の評価を行った．抽出主題表示キーワードの主題再現性を評価するために表題内のキーワードと比較，また実使用における有効性を評価するために JICST-DB に付与されている索引語と比較した．このことから，当システム機能により抽出される主題表示キーワードは，索引語作成に十分参考となる結果をもたらし，特に主題表示キーセンテンスの抽出は，主題索引語の作成に有効に機能することの結論を得た．

1. はじめに

文書，論文等資料を検索する場合，書誌データ（例：著者名，表題，出版者名等）を指示し検索する場合と，内容索引データ（例：分類番号，索引語）を指示し検索する場合とがある．そのために，資料情報データベース（以下 DB と記す）に，両者のデータを蓄積しておくことが行われている．分類番号は資料の配架場所を表示するために，索引語は資料の記述内容を表示するために作成される．大容量資料情報 DB の構築のために，書誌データおよび内容索引データの自動作成機能が必要になる．書誌データの自動作成には，例えば表題ページの自動認識および書誌データ記述フォーマットへの自動変換機能が，内容索引データの自動作成には，資料の記述内容の意味解析を行い設定する機能が必要になる．

現状での索引語作成は，索引語作成者が資料内容を解読し，シソーラス用語（統制索引語という）を用い設定し，統制索引語を補助するために表題もしくは抄録内のキーワードを自動抽出し，自然語レベルの索引語として設定している（日本における代表例：JICST-DB の構築[1]，国立国会図書館—雑誌記事索引の作成[2]）．前者においては大容量の処理に対応できず，後者においては資料の記述内容を解析した結果としての索引語設定にならない問題点がある．

本研究の目的は，索引語作成者が主題索引語を設定する際に，資料内容を解読しなくとも索引語を設定できる索引語作成支援（MAI：Machine-Aided Indexing[3),4)]）システム機能（図1参照）を開発することにある．資料内容を対象とする検索要求に，例えば“何々について述べている資料を知りたい”という主題検索要求がある．このことに対応する索引語を主題索引語という．主題索引語は，“何々について報告する，何々について述べる”等，資料の作成意図を表現する動詞に導かれる対象の単語が相当する．そこで“何々について報告する，何々について述べる”等の表現を認識するための文意解析システム，その基で対象となる単語を主題表示キーワードとして抽出，あるいは主題表示キーワードが省略され存在しない場合に主題表示キーセンテンスとして抽出し出力するシステム機能の開発を行う．当システム機能の開発研究を行うために，日本語による科学技術論文の著者抄録を対象に解析を行い，システム機能の評価実験を行う．

以下，2章でキーワード自動抽出システムについて従来の研究をレビューし，問題点について検討を行い，3章で筆者の方法（システム機能）について示す．4章で本システムの評価実験結果を示し，本システム機能の考察を行い，5章で結論を示す．

2. 従来の研究と問題点

索引語機能について検討をし，キーワードの自動抽出システムについて従来の研究をレビューし問題点を検討する．

2.1 索引語の機能

索引語機能に下記2点がある[5)]．

1）構造化索引語（String index term)…事前に設

† A Machine-Aided Subject Indexing System Based on Sentence Analysis by TETSUYA ISHIKAWA (Faculty of Library and Information Science, University of Library and Information Science).

†† 図書館情報大学図書館情報学部

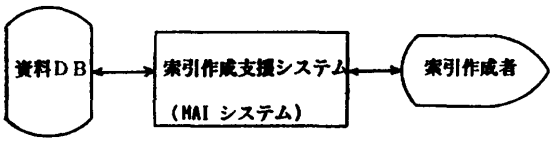

図1 索引作成支援システム機能
Fig. 1 The function of Machine-Aided Indexing system.

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．資料内容から索引項目に相当する索引語を解析し設定する．ファクト（事項）検索に対応する索引語といえる[6]．

2）主題索引語（Subject index term)…資料内容の主題，すなわち，執筆者の資料作成意図を表示する索引語．資料内容を解析し，主題を表示する索引語を設定する．主題検索に対応する索引語といえる[7),8)]．

2.2 従来の研究と問題点[9)~12)]

索引語の候補となるキーワードを自動的に抽出する方式の研究は，大容量資料情報 DB 構築の必要性から実務的に行われてきた．その流れに，資料情報 DB の構築時に個々の資料に索引語を付与するために抽出する事前付与方式と検索時に指示された検索語を対象に抽出する検索時設定方式の２通りの方式がある．以下，分けて従来の研究をレビューし問題点を示す．

2.2.1 事前付与方式

1）出現頻度解析抽出法

ⅰ）単一語抽出法

資料内に出現する単語を抽出し，出現頻度の事前設定値に対し対応する単語を索引語として直接利用する方式．この方式は，Luhn, H. P. により提案された[13)]．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すべての単語を対象資料の索引語とするストップワード方式により実用化された．しかし，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ことから，逆にキーワード・リストを事前に設定し利用するキーワード方式が提案された．当方式は，登録語の網羅性の保障を必要する．日本において大規模キーワード方式の代表例に下記2）があり，ストップワード方式に二村らによるシステムがある[14)]．

ⅱ）単語間相関抽出法

複合語判定のために，資料内に出現するキーワードを対象に，単語間相関頻度を基に解析し，相関頻度の事前設定値に対しすべての組合せを複合語とし，その複合語を上記ⅰ）によるキーワードを含め索引語とする方式．この方式は，Stailes, H. E.らにより提案された[15)]．当方式は，対象資料中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，精度において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，資料内に出現する単語を対象に照合し，その単語を索引語として直接利用する方式．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．この方式の例として，日本語による資料を対象とする例として HAPPINESS（(株) 平和情報センター[16)]），Free Base（(株) エム・シー・ワードセンター）等があり実用に供されている．当方式は，登録キーワードの網羅性を保障するために，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．大規模キーワード・リストによる抽出法の日本における実用例に，HAPPINESS を利用している前述雑誌記事索引および新聞記事 DB 構築[17)] 等がある．以上の方式は，主題索引語を抽出する方式である．

3）文構造解析による抽出法

資料の記述内容を対象に自然言語処理を基に文構造解析を行い抽出する方式．日本語資料を対象に，構造化索引語抽出となる，報道記事検索を目的に 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[18)]，構造化索引語抽出システムを目的とする Morita らの研究[19)] がある．絹川らの方式は，5W1H に相当するキーワードに限定するため，一般資料には適用できない問題がある．主題索引語抽出については形態情報による吉村ら[20)]，形式動詞「である」に対応するキーワードを抽出する柴田ら[21)]，字種により抽出する松崎ら[22)]，「格」に対応するキーワードを抽出する本本[23)]の研究があるが，主題を表示するキーワード以外のキーワードを抽出することになり，検索語を論理式に展開し指示しても“雑音”に相当する不適切な検索結果を生じさせる欠点がある．

2.2.2 検索時設定方式

1）検索指示内容を直接利用する方法

質問文と並び例えば参考資料の抄録を入力し，その中のキーワード抽出を行い，DB 内に出現する単語とのクラスタ関係により検索を行う方式．この方式は，Salton, G. により SMART システムとして提案された[24)]．当方式は，指示される質問文もしくは指示資料

から抽出されたキーワードを基に検索をすることから，不適切な検索結果の回避に寄与する利点はあるが，DB 内の単語の照合において欠落が生じる問題があり実用に至っていない．

2）シソーラス用語を利用する方法

検索システム内に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すべてを基に DB 内に出現する単語を対象に検索する方式．この方式の例として，英文資料を対象とする METAMOLPH (Thunderstone/EPI 社) がある[25]．この方式は，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，根本的には同様な問題点がある．

3. システム機能

システムの目標と機能を示す．

3.1 システムの目標

索引語作成者に主題索引語の設定に参考となる主題表示キーワードおよび主題表示キーセンテンスを提示するシステムとする．主題表示キーワードおよび主題表示キーセンテンスとは，執筆者の資料作成意図に対応する資料内の名詞句あるいは記述文をいう．このために，下記システム機能とする．

1）動詞の意味に注目し文意解析を行い，資料作成意図に対応する名詞句あるいは記述文を抽出・出力する．

2）名詞句に未知語が生じるのでは意味をなさないことから，名詞辞書を持たない動詞，助詞，副詞，接続詞のみの辞書により処理するシステムとする．

3.2 システム機能の概要

1）処理対象文に対する前修正処理

i）名詞連続表記（例：「科学技術，政治，経済」）は中黒「・」表記とする．

ii）格助詞「が」の直後の読点（例：「それらが，」）は削除する．

iii）接続助詞「と」「や」等の直後の読点は削除する．

iv）接続詞の後に読点を挿入する．

2）単位構文処理

文中に出現する動詞ごとの構文（単位構文という）処理を行う．

●単位構文処理規則：

i）1文を句点にて判別し取り出し，右線形処理により動詞辞書データと照合し，適合した動詞の直前の助詞を発見し，適合した助詞の直前の単語（接続助詞，副助詞，句接続詞および形容詞を含む名詞句）を助詞スロットに埋め，出現順に動詞スロットに埋める．ただし，①読点もしくは処理済みの単語を越え存在する未処理語は，直後の動詞スロットの対象として処理する．②読点もしくは処理済みの単語を越え存在する句接続詞は，直後の動詞の処理済みスロットの後に独立スロットとして位置付ける．③文接続詞は，その前後の処理済みスロットを（ ）で囲む．④文の冒頭の文接続詞は，その文の最後の動詞の処理済みスロットの後に独立スロットとして位置付ける．下記に例を示す．

●例文：「本報告ではキーワード自動抽出をターゲットとしたキーワード用のシソーラス概念構造について報告する．」（出典：下記4）の抄録例）

●辞書データ例（図2参照．記号付表参照，以下同様）

動詞：

した（ACTAOP (nil) (T：pt, M：a, A：a, C：連体)）

報告する（ACTADE (nil) (T：p, M：a, A：a, C：終止)）

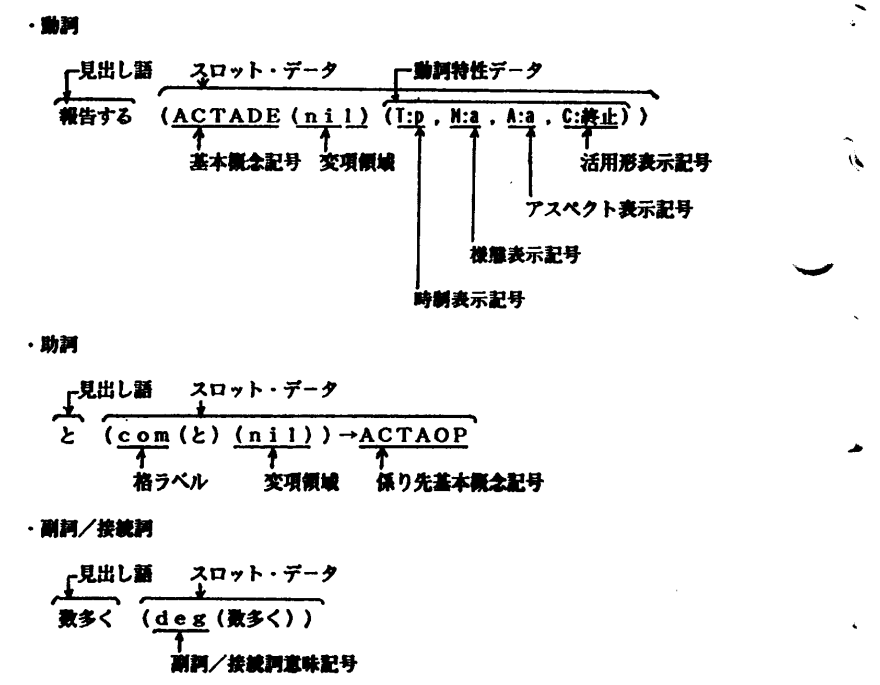

図2 辞書データ・フォーマット（記号：付表参照）
Fig. 2 Dictionary data format.

助詞：

と (com(と)(nil))→ACTAOP

を (obj(を)(nil))→ACTAOP

について (obj(について)(nil))→ACTAOP

では (spa(では)(nil))→ACTAOP

●単位構文処理過程：

処理1：文の最初の動詞「した」に対する辞書データ内のスロット・データ（以下スロットと記す）を取り出し，スロット内の変項領域 (nil) に動詞「した」を埋める．ただし体言止めの場合は，陳述動詞 (PRE) にする．

処理1の結果：

(ACTAOP(した)(T：pt, M：a, A：a, C：連体)▼)

処理2：動詞「した」の直前の助詞「と」に対する辞書データ内のスロットを処理1の結果スロット（上記 ▼ の部分）に埋める．以後，読点もしくは処理済みの単語まで判定し，その間の単語を助詞スロットの変項領域 (nil) に埋める．ただし読点の直前が名詞の場合は格ラベルを (com) にする．

処理2の結果：

(ACTAOP(した)(T：pt, M：a, A：a, C：連体)(com(と)(ターゲット))(obj(を)
(キーワード自動抽出))(spa(では)(本報告)))
2番めの動詞「報告する」に対する処理結果
(ACTADE(報告する)(T：p, M：a, A：a, C：終止)(obj(について)(キーワード用のシソーラス概念構造)))

3）構文処理

文ごとの構文処理を行うために，単位構文間の埋め込み処理を単位構文処理結果の順に行う．

●埋め込み処理規則：

i）連用形スロットを後出動詞スロットに埋める．

ii）連体形スロットを後出体言スロットに非交差規則を適用し埋める．下記に例を示す．

●例文：「筆者らは，シソーラスを用い，新聞記事からキーワードを自動抽出するシステムを開発中である．」（出典：下記4）の抄録例）

●単位構文処理結果（以下動詞特性データのうち，活用形表示記号以外省略，以下同様）

最初の動詞「用い」に対する処理結果
(ACTARE(用い)(連用中止)(obj(を)
(シソーラス))(top(は)(筆者ら)))
2番目の動詞「自動抽出する」に対する処理結果
(ACTACR(自動抽出する)(連体)(obj(を)
(キーワード))(ori(から)(新聞記事))◆)
3番目の動詞「ある」に対する処理結果
(PRE(ある)(終止)(dur(で)(開発中))
(obj(を)(システム)▽))

●単位構文埋め込み処理過程：

処理1：文の最初の動詞「用い」は，連用中止であるから処理規則 i）を適用し，2番目の動詞「自動抽出する」のスロット（上記 ◆ の部分）に埋める．

処理1の結果：

(ACTACR(自動抽出する)(連体)(obj(を)
(キーワード))(ori(から)(新聞記事))
(ACTARE(用い)(連用中止)(obj(を)
(シソーラス))(top(は)(筆者ら))))

処理2：処理1の結果の最初の動詞「自動抽出する」は，連体であるから処理規則 ii）を適用し，3番目の動詞「ある」のスロット内の体言「システム」の後（上記 ▽ の部分）に埋める．

処理2の結果：

(PRE(ある)(終止)(dur(で)(開発中))
(obj(を)(システム)
(ACTACR(自動抽出する)(連体)
(obj(を)(キーワード))
(ori(から)(新聞記事))
(ACTARE(用い)(連用中止)
(obj(を)(シソーラス))(top(は)(筆者ら))))))

4）主題表示キーワードおよび主題表示キーセンテンス抽出処理

構文処理結果を基に，主題表示キーワードおよび主題表示キーセンテンス抽出処理を行う．

●抽出処理規則：

i）下記基本概念記号の動詞に対応する「obj」の単語を主題表示キーワードもしくは「obj」が存在しない文を主題表示キーセンテンスとする．

変成動作動詞(ACTACH) …例：「改良した」
生成動作動詞(ACTACR) …例：「構築した」
表現動作動詞(ACTADE) …例：「述べた」
確認動作動詞(ACTREC) …例：「確かめた」

ただし，動詞特性データについて，“時制”に関して“過去と現在”，“様態”に関して“能動と受動”，“アスペクト”に関して“肯定”を対象とする．

ii）資料ごとに構文処理結果の先頭の動詞の基本概念記号を判定し，「obj」内の名詞句（主題表示キー

ワード）を高輝度表示する．「obj」がない場合は，主題表示キーセンテンス全体を高輝度表示する．索引作成者は，この結果を参考にする．以下に例を示す．

- 抄録例：「筆者らは，シソーラスを用い，新聞記事からキーワードを自動抽出するシステムを開発中である．前回報告したシステムでは，データベース検索用のシソーラスを流用しており，シソーラスの概念構造が検索用であるためキーワード自動抽出には不十分である．本報告ではキーワード自動抽出をターゲットとしたキーワード用のシソーラス概念構造について報告する．」（出典：森崎正人，中園薫：キーワード自動抽出をねらいとした構成法について，第27回情報処理学会全国大会論文集，27-2，pp. 1067-1068 (1983)）
- 構文処理結果：

  1文めの処理結果
  (PRE(ある)(終止)(dur(で)(開発中))
  (obj(を)(システム)
  (ACTACR(自動抽出する)(連体)
  (obj(を)(キーワード))
  (ori(から)(新聞記事))
  (ACTARE(用い)(連用中止)
  (obj(を)(シソーラス))(top(は)(筆者ら))))))
  2文めの処理結果
  (PRE(ある)(終止)(att(で)(不十分))
  (obj(には)(キーワード自動抽出))
  (for(ため)(PRE(ある)(終止)(com(で)(検索用))
  (sub(が)(シソーラスの概念構造))
  (ACTARE(流用しており)(連用中止)
  (obj(を)(データベース検索用のシソーラス))
  (top(では)(システム)(ACTADE(報告した)
  (tim(前回)))))))
  3文めの処理結果
  (ACTADE(報告する)(終止)(obj(について)
  (キーワード用のシソーラス概念構造)
  (ACTAOP(した)(連体)(com(と)(ターゲット))
  (obj(を)(キーワード自動抽出))
  (spa(では)(本報告)))))
- 抽出対象例：表現動作動詞（ACTADE）
- 主題表示キーワードおよび主題表示キーセンテンス抽出処理：

  処理1：1文めの抽出処理…抽出処理非対象
  処理2：2文めの抽出処理…抽出処理非対象
  処理3：3文めの抽出処理…「キーワード用のシソーラス概念構造」
- 抽出主題表示キーワードの理解：

  資料例の記載内容は，“固有名詞シソーラスの自動拡張”にあり，その“シソーラス構造”について言及している．ゆえに，当資料の主題索引語には，「シソーラス構造」が必要になる．例えばJICST-DBの同資料データには，「シソーラス，キーワード，件名索引語，形式文法，件名索引，データベース管理システム，文献検索，自動言語処理，自動索引法，ニュース，新聞，検索効率」の12個の索引語が付与されているが，唯一対応する「シソーラス」は“シソーラス構造”に限定する索引語にはならない．当システムの抽出結果は，「キーワード用のシソーラス概念構造」という主題表示キーワードを提供しており，主題索引語として「シソーラス構造」を採用するに可能なデータを提供しているといえる．

4. 評価実験

抽出実験を行い，主題索引候補語となり得るか否かの適性を評価し，本システム機能の有効性を評価する．

4.1 実験対象抄録

- 実験対象抄録：情報処理学会自然言語処理研究会報告，Vol. 86, No. 25 (1986)～Vol. 87, No. 32 (1987)（約1年分）の33抄録（164文）を対象に行う．

抄録を対象とする理由…抄録は，特に科学技術論文において研究の目的，研究方法，結論の3点を中心に本文を紹介する目的のもとに圧縮したテキストであり，索引語作成の対象になることから．

4.2 辞書データ・ファイルの作成

1）見出し語の設定

上記実験対象文を対象に，筆者により動詞，助詞，副詞，接続詞を抽出し設定．

2）スロット・データの設定

i）動詞基本概念記号の設定

サ変動詞974語を対象に筆者により分析・設定した概念分類体系に対し[26]，上記実験抄録内に出現した動詞317語（サ変動詞146語，和語動詞171語）に対し検証・分析を行い，特に和語動詞を中心に再検討を行い修正し設定（付表参照）．

ii）格ラベルの設定

科学技術庁の機械翻訳プロジェクト（Muプロジェクト）の格ラベルを利用し，上記実験抄録内に出現し

た助詞 65 語に対し設定.

ⅲ) 助詞，接続詞，副詞意味記号の設定

上記実験抄録内に出現した単語を元に筆者において新規に設定（付表参照）.

4.3 評 価 法

抽出された主題表示キーワードおよび主題表示キーセンテンスから判定できる主題表示キーワードの主題再現性および実使用における有効性を評価するために，下記の評価を行う.

1) 主題再現性の評価…抽出主題表示キーワードと表題中の主題表示キーワードと比較する.

表題と比較する理由…表題は，執筆者による本文の主題内容を表す句もしくは文である．ゆえに表題中のキーワードの中に，主題表示キーワードがあると考えることから（ただし本文の内容と異なる表題もある．ゆえに表題を対象にキーワードを抽出することは危険である）.

2) 実使用における有効性評価…実験対象抄録について JICST-DB に付与されている索引語を調査し比較する.

JICST-DB に付与されている索引語と比較する理由…JICST-DB に付与されている索引語は，実使用に供されていることから．そこで筆者において実験対象資料の本文をも解読し JICST-DB に付与されている索引語の再評価を行い，抽出主題表示キーワードと比較し，特にフリー・キーワードとして付加すべきか否かを判定し，抽出主題表示キーワードの実使用における有効性を評価する．以下に評価例を示す.

評価例（実験対象抄録 Id-No.: 86-58-1）:

- 表題:「質問応答における意図の把握と話題の管理」
- 表題内の主題表示キーワード:「意図の把握と話題の管理」(「質問応答」は，"分野" を表示するキーワードであって，主題を表示するキーワードではない．以下関連キーワードと呼ぶ)
- JICST-DB の索引語:「質問応答システム，知識表現，人工知能，意味論，自動分類，文章，文書処理，構文分析，知識ベース，エキスパートシステム」

  フリー・キーワード: なし
- 抽出主題表示キーワード:「意図の把握と話題の管理，対話モジュールとフェーズの概念，方式→入力の意図の判定方式（主題表示キーセンテンス内において判定)」
- 主題再現性の評価:((表題内主題表示キーワード数)−(表題内主題表示キーワードとの一致抽出主題表示キーワード))=0（意味: 完全一致)

  評価結果:「意図の把握と話題の管理」は一致．ゆえに主題再現性の評価値=0．ゆえに再現性ありと判定する（完全一致のみ判定．理由: 完全一致以外の値については，表題および抄録の "質" を評価しなければならなくなる．当システムの目的は，原著データに対し機能させることを目的としていることから，完全一致において有効と見る).
- 実使用においての有効性評価: フリー・キーワードを含め JICST-DB の索引語と抽出主題表示キーワードを比較し，特にフリー・キーワードとして付加すべきか否かを判定する.

  評価結果:「意図の把握と話題の管理」は「知識表現」および「知識ベース」に相当すると判定.「対話モジュールとフェーズの概念」は関連キーワードとして「質問応答システム」に相当，「意図の判定方式」を本文の内容からフリー・キーワードとして設定する必要ありと判定．ゆえに抽出主題キーワードすべてが実使用に有効であると判定する (JICST-DB の索引語を評価することが目的ではないことから，他の索引語については評価しない).

4.4 評 価 結 果

1) 主題再現性の評価結果

主題表示キーワードとして直接抽出できたのは 13 抄録，主題表示キーセンテンスから判定したのは 17 抄録，指定基本概念記号がなく抽出・判定できなかった抄録は 3 抄録（評価対象外とする）であった．抽出主題表示キーワードと表題中のキーワードとの比較において 96% 一致した.

2) 実使用の有効性の評価結果

抽出主題表示キーワードと JICST-DB の索引語 (フリー・キーワード既付与件数: 6 抄録) との比較において JICST-DB の索引語で関連キーワードを含め充足しているのは 17 抄録，充足していないのは 13 抄録であった．抽出主題表示キーワードをフリー・キーワードとして付与すべきか否かの判定の結果，すべて付与すべきと判定できた.

以上のことから，当システム機能により抽出される主題表示キーワードは，索引語作成に十分参考となる結果をもたらし，特に主題表示キーセンテンスの表示は，主題索引語の設定に十分参考になるものと考えられる.

4.5 当システム機能の課題

当研究において辞書データを実験対象抄録を基に作成し利用した．当システム機能の実運用のために下記2点の課題を解決する必要がある．

1）辞書データの充足

2）概念分類体系の最適化・詳細化

科学技術論文において，述語はサ変動詞による表現が圧倒的に多いが，形容詞終止形による表現もある．今回設定した概念分類体系は，研究の目的からサ変動詞，和語動詞を中心に当概念分類体系を適用し，形容詞終止形は「形容詞終止形＋のである」（例：「広い」→「広いのである」）とし「陳述」（PRE）に一意に設定し処理を行った．主題表示キーワードを抽出する目的においては問題はないが，当システム機能を，例えば構造化索引語の抽出に利用するならば，形容詞を対象とする概念分類体系の設定研究を行う必要がある．

5. おわりに

本研究の目的は，資料を対象とする主題検索要求に対応する主題索引語を自動的に設定するシステム機能を供することにある．評価実験から，当初の目的を達成する機能として利用可能であると考える．

これまでの資料情報 DB は，書誌データの提供を目的に構築，供されてきた．資料情報 DB を検索し，必要な書誌データが分かっても，原資料を入手しなければならず，繁雑な作業を伴う．このことに対し，近年，資料情報 DB の質的変革が求められてきている．特に新しい知見（例：方式）等を資料情報 DB から直接求める期待が出てきている．科学技術論文の利用の目的は，主に下記の点にある．1）新しい知見を得る．2）研究内容と研究者あるいは研究機関（所在情報という）を得る．1)のことに応えるには資料内容から新しい知見等を抽出し，2)のことに応えるには資料内容から研究対象項目を抽出し，必要な書誌データと組み合わせ提供する必要がある．このことに関して，構造化索引語の設定が必要になる．当研究成果は，主題表示キーワードを抽出するシステム機能だけではなく，このことに応用できると考える．今後は，これまでの成果を発展させ，構造化索引システムの研究を行う．

謝辞 電子技術総合研究所坂本義行主任研究官，シャープ(株)情報システム研究所塚田康博主任らに，助言，支援をいただいた．ここに記し感謝を申し上げます．

参考文献

1) 五味淵亘：インデクシングの現場(3)，JICSTにおけるインデクシングの実際，情報の科学と技術，Vol. 39, No. 3, pp. 99-109 (1989).
2) 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).
3) Martinez, C. et al.: An Expert System for Machine-Aided Indexing, *J. Chem. Inf. Comput. Sci.*, Vol. 27, No. 4, pp. 158-162 (1987).
4) Humphrey, S. W.: A Knowledge-Based Expert System for Computer-Assisted Indexing, *IEEE Expert*, Vol. 4, No. 3, pp. 25-38 (1989).
5) Cleverland, D. B. and Cleverland, A. D.: *Introduction to Indexing and Abstracting*, p. 209, Libraries Unlimited, Inc., Colorado (1983).
6) Craven, T. C.: *String Indexing*, p. 246, Academic Press, Inc., London (1986).
7) Foskett, A. C.: *The Subject Approach to Information*, 4th ed., p. 574, Clive Bingley, London (1982).
8) Dym, E. D. ed.: *Subject and Information Analysis*, p. 498, Marcel Dekker, Inc., NY (1985).
9) Stevens, M. E.: *Automatic Indexing: A State-of-the-Art Report*, p. 290, National Bureau of Standards, Washington D. C. (1965).
10) 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).
11) 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).
12) Chan, L. M. et al. ed.: *Theory of Subject Analysis, A Sourcebook*, p. 415, Libraries Unlimited, Inc., Colorado (1985).
13) Luhn, H. P.: A Statistical Approach to Mechanized Encoding and Searching of Literary Information, *IBM J. Res. Dev.*, Vol. 1, No. 4, pp. 309-317 (1957).
14) 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).
15) Stailes, H. E.: The Association Factor in Information Retrieval, *J. ACM*, Vol. 8, No. 2, pp. 271-279 (1961).
16) 染谷浩司：日本語データベースの構築を目指して，キーワード展望，第22回情報科学技術研究集会発表論文集，Vol. 22, pp. 37-45 (1986).
17) 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).
18) 絹川博之ほか：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).

19) 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).
20) 吉村賢治ほか：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).
21) 柴田浩一ほか：科学技術文献からの専門用語情報の自動抽出，第35回情報処理学会全国大会論文集，35-2, pp. 1283-1284 (1987).
22) 松崎浩一ほか：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).
23) 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).
24) Salton, G. and McGill, M. J.: *Introduction to Modern Information Retrieval*, p. 448, McGraw-Hill International Book Co., Auckland (1983).
25) 米田健二：メタモルフ (METAMORPH)，テキストを理解する人工知能ソフトウエア，オンライン検索，Vol. 10, No. 3, pp. 117-126 (1989).
26) 石川徹也，坂本義行：動詞意味機能に基づく日本語格フレームの生成，情報処理学会自然言語処理研究会報告，Vol. 89, No. 27, pp. 71-73 (1989).

付表：記号表

基本概念記号

動詞 CCP 名	名　　称
PRE	陳述
ACTACO	結合・分割動作動詞
ACTAMO	移動動作動詞
ACTACH	変成動作動詞
ACTACR	生成動作動詞
ACTAMA	運用動作動詞
ACTARE	参考動作動詞
ACTADE	表現動作動詞
ACTREC	確認動作動詞
ACTAIN	意図動作動詞
ACTAOP	状態化動作動詞
CONCMO	移動状態動詞
CONCMN	変化状態動詞
CONCCI	状況状態動詞
CONCIN	意図状態動詞

動詞の時制記号

記　　号	意　　味
pt	過去 (past)
p	現在 (present)
f	未来 (future)

動詞の様態記号

記　　号	意　　味
a	能動 (active)
p	受動 (passive)

動詞のアスペクト記号

記　　号	意　　味
a	肯定 (affirmation)
n	否定 (negation)
c	仮定 (condition)
k	継続 (continuous)
ab	可能 (ability)
w	希望 (want)
m	義務 (must)
?	疑問 (interrogate)

副詞意味記号

記　　号	意　　味
and	追加 (and)
ant	添加 (and then)
deg	度合 (degree)
exs	実例 (example sentence)
imp	強調 (importance)
man	方式 (manner)
ord	順序 (order)
sta	状態 (state)
tim	時 (time)
tto	時・終点 (time-to)
wit	同伴 (with)

句接続詞意味記号

記　　号	意　　味
ant	添加 (and then)
not	追加 (not only)
/	並列

文接続詞意味記号

記　　号	意　　味
ant	添加 (and then)
but	逆接 (but then)
exp	説明 (explanation)
for	目的 (for)
res	順接 (response)

接続助詞意味記号表

記　号	意　味
and	並列 (and)
but	逆接 (but then)
res	順接 (response)
sup	仮定 (supposition)

(平成2年9月4日受付)
(平成2年11月13日採録)

石川　徹也（正会員）

昭和18年生．昭和46年度慶応義塾大学大学院修士課程（図書館情報学専攻）修了．富士写真フィルム(株)足柄研究所入社，図書館短期大学を経て現在，図書館情報大学図書館情報学部助教授．情報管理システム機能の高度化の研究に従事．米国計算機言語学会，人工知能学会，オフィスオートメーション学会，日本経営工学会各会員．