

データベース構築促進及び技術開発に関する報告書

書誌データベース用
ダイナミック・シソーラスの
可能性調査と実験

平成3年3月

財団法人 データベース振興センター
委託先 株式会社紀伊國屋書店

本報告書は、日本自転車振興会から競輪収益の一部である機械工業振興資金の補助を受けて作成したものである。

序

データベースは、わが国の情報化の進展上、重要な役割を果たすものと期待されている。今後、データベースの普及により、わが国において健全な高度情報化社会の形成が期待される。さらに海外に対して提供可能なデータベースの整備は、国際的な情報化への貢献および自由な情報流通の確保の観点からも必要である。しかしながら、現在わが国で流通しているデータベースの中でわが国独自のものは3割にすぎないのが現状であり、わが国データベースサービスひいてはバランスある情報産業の健全な発展を図るためには、わが国独自のデータベースの構築およびデータベース関連技術の研究開発を強力に促進し、データベースの拡充を図る必要がある。

このような要請に応えるため、(財)データベース振興センターでは日本自転車振興会から機械工業振興資金の交付を受けて、データベースの構築および技術開発について民間企業、団体等に対して委託事業を実施している。委託事業の内容は、社会的、経済的、国際的に重要で、また地域および産業の発展の促進に寄与すると考えられているデータベースの構築とデータベース作成の効率化、流通の促進、利用の円滑化・容易化などに関係したソフトウェア技術・ハードウェア技術である。

本事業の推進に当って、当財団に学識経験者の方々に構成されるデータベース構築・技術開発促進委員会(委員長 山梨学院大学教授 蓼沼良一氏)を設置している。

この「書誌データベース用ダイナミック・シソーラスの可能性調査と実験」は平成2年度のデータベースの構築促進および技術開発促進事業として、当財団が株式会社紀伊國屋書店に対して委託実施した課題の一つである。この成果が、データベースに興味をお持ちの方々や諸分野の皆様方のお役に立てば幸いである。

なお、平成2年度データベースの構築促進および技術開発促進事業で実施した課題は次表のとおりである。

平成3年3月

財団法人 データベース振興センター

平成2年度 データベース構築促進・技術開発依託課題一覧

分野	課題名	委託先
社会	1 形態学的コメントを含む病理データベースのフィージビリティ調査	(株)エス・ピー・オー
	2 災害情報データベース支援環境の構築	(株)防災都市計画研究所
	3 AV/MARCのための分類索引データベース構築	(株)ダイソメディアサービス
	4 気候情報データベースの構築	(株)エムデーエス雪氷研究所
	5 健康の自己管理と病気予防データベースの構築	(株)コンピュータコンビニエンス
	6 シルバーエイジの実態及び生活に必要な情報のデータベース構築のための調査研究	美崎高齢者福祉互助会美崎生活館
	7 交通事故調査データのデータベース化に関する調査研究	(財)日本自動車研究所
地域活性化 中小企業振興	8 アジア太平洋交流データベースの課題性の研究	(株)西日本新聞社
	9 戦略商圏レベルに細分化した地域データと分析・提案手法を統合化した企画支援システムデータベースの構築	パラシュート情報開発研究会 札幌凸版印刷(株)
	10 ネットワーク化された地域情報データベースの有効なマネジメントについての調査研究	セントラル開発(株)情報図書館 RUKIT
	11 徳島市中小商業振興データベースの構築	(株)ニューメディア徳島
	12 九州地域の人材情報データベース構築	(財)九州産業技術センター
海外	13 海外向け国内先端技術分野中堅企業情報英文データベース構築	コムラインインターナショナル(株)
	14 海外規格(ソ連邦国家規格)データベースの整備	日本電子計算(株)
	15 政府開発援助(ODA)に関するデータベースの構築調査	(財)日本国際協力システム
	16 専門用語データベースシステムの機能に関する調査研究	アイ・エヌ・エス(株)
	17 専門家データベース構築事業	(財)海外貿易開発協会
技術	18 VAN用データベース管理システムの開発	シャープ(株)
	19 レコードマネジメント用辞書管理システムの開発研究	(株)オフィス総研
	20 建築CAD用拡張可能データベースのプロトタイプ作成	三菱電機(株)
	21 先進複合材料データベース・プロトタイプの作成	(財)次世代金属・複合材料研究開発協会
	22 マイクロコンピュータのプログラマブル周辺デバイスのデータベース化	(社)日本システムハウス協会
	23 書誌データベース用ダイナミック・シソーラスの可能性調査と実験	(株)紀伊國屋書店

目 次

I 研究の目的	1
1 初期構想	1
2 具体的取り組み	1
2.1 現行書誌データベース検索の問題と打開技術の前提条件	1
2.2 適性技術の搜索と取組み	2
II 書誌データベース検索システム機能	6
1 索引語の機能	6
2 従来の研究と問題点	6
2.1 事前付与方式	7
2.2 検索時設定方式	8
III 文意解析処理システム機能	9
1 システムの前提	9
2 システム機能の概要	10
2.1 単位構文処理	10
付録 A 辞書データ例	17
IV システムの構築および実験・評価	27
1 構築システム	27
1.1 プロトタイプ処理方式概要	27
1.2 処理内容	28
2 処理結果	30
2.1 前処理	30
2.2 「要旨」からの抽出処理	31
2.3 「質問文」からの抽出処理	31
2.4 後処理	32
2.5 検索処理	33
2.6 結果の評価	35

付録 A	テストデータベース内容例	36
付録 B	「要旨」部分の前処理結果例	38
付録 C	「要旨」部分の処理結果例	40
付録 D	処理プログラムの概要	42
V	結論	43
1	ダイナミックシソーラス機構	43
2	検索システムとしてのダイナミック・シソーラス	43
2.1	検索手順の確認と評価	43
2.2	今後の課題	44
3	データベースとしてのダイナミック・シソーラス	44
3.1	ダイナミック・シソーラス搭載データベース構築の構想	44
4	トータルシステムとしてのダイナミック・シソーラス	45
5	今後の展望と抱負	45

I 研究の目的

1 初期構想

データベース検索をもっと楽に、効率良く、かつ効果的に行えないか。データベースをもっと楽に、効率良く、かつ効果的に作れないか。書誌データベースが自らの確に索引語句を特定・付与し、検索者が思うがままに綴る質問文をシステムの検索質問式へと変換できる方法はないか。データベース記事の自然言語(表題、要旨、抄録、原文など)から自動的に該当索引語句を発生できる方法はないか。即ち、自然言語による直接検索とデータベース記事索引語句の自動付与、および両者の自己学習的で相補・双方向・相乗的な更新・拡充を可能とするシステムをダイナミック・シソーラスと命名し、またそのシステムの作動機序をダイナミック・シソーラス・エンジン(またはシソーラス・エンジン)と命名し、そのシステム技術の可能性を調査・実験することが当委託研究の目的である。(図I.1を参照)

2 具体的取り組み

2.1 現行書誌データベース検索の問題と打開技術の前提条件

現行のデータベース検索の大半は索引語句(綴り)の照合と論理演算(+,*,-)によって行われ、精度向上のために分類・統制用語(項目、ファセット、コード、シソーラスなど)と近接演算((W),(N),(S),(L)など)がデータベース構築と検索の両方で活用されているが、両方とも長期学習と熟練および多大の経費を要し、データベース構築側では持続と均質保持が困難であり、検索側ではデータベースごと検索システムごとの特異性のため総合的熟練はとても不可能である。(図I.2を参照。)

言葉による意味の表現は表現対象を指摘する体言(名詞)とその動態とそれへの関係・関与度合いを限定する用言(動詞、形容詞)やその修飾・添加語(副詞、感動詞、助動詞)との組合せ、および両者の組合せの枠組み・区分・形態・方向・序列などを限定する接続詞・助詞の介入で成立している。(図I.3を参照)

これに対して現行の検索式(語句の照合と論理・近接演算)では語句の照合で主部と述部や体言と用言の区別がつかず、論理・近接演算では主部と述部や体言と用言の判別・組合せが粗雑であるために精密な意味表現が困難で、定義不足や定義過剰に陥りやすく、質問文の検索式への適正変換のためにはデータベースごとの記事編集特性に精通した名人芸が必要である。またデータベース構築側では現行検索式で検索される際の収載記事の意味の分裂と意味の深さの擦れ違いを防止するために用意した分類コードや統制索引用語集(シソーラス)に基づいて索引用語を付与し、検索者にもデータベース構築手順を知ってもらい同じ分類コードや統制索引用語集を検索式組立てに活用してもらうことで安定検索を期待・支援しているが、その労力と費用は甚大であり、商用データベースの成功を困難なものとしている。(図I.2を参照。)

よってこれらの困難を打開するためには、データベース構築技術としては記事の索引用語を精密(最適合・最網羅的)にかつ自動的に識別・発生できるアルゴリズム(ダイナミック・シソーラス)の開発、検索システム技術としては自然言語による質問文を検索システム特有の検索式へと精密(最適合・最網羅的)にかつ自動的に変換できるアルゴリズム(ダイナミック・シソーラス)の開発、そして両方とも(索引用要語句も検索式も)相補・双方向・相乗的に自動更新・自動拡充する自己組織的学習制御機能(ダイナミック・シソーラス・エンジン)の開発が必要となる。即ち、現行の書誌データベース記事の文章項目を材料とした記事の全体的・自動的な構造化と論理化(AI言語処理)が必要となる。

従って当ダイナミック・シソーラス構想の実現に必要な技術に対する要求機能としては、少なくとも以下の三条件が不可欠である。

- 1) 日本語自然言語解析(形態素分析と構文解析)機能
- 2) 日本語文の意味処理(意味を認知、判別、および優先順位判定する)機能
- 3) 日本語述部の汎用・自動的解析(述部の結合序列、役割、および動態を汎用・自動的に判別、分類、および体系化する)機能

2.2 適性技術の搜索と取組み

書誌データベースに対する理想的な検索機能は、検索質問内容およびデータベースの内容を自動解析し、直接検索を行うシステム機能である。当委託課題において、当システム機能を「ダイナミック・シソーラス」と位置付け、研究を行う。ダイナミック・シソーラスは、質問文および書誌データベース内の抄録を、文意解析処理を基に、それぞれの内容から有意味なキーワード抽出を行い、キーワード照合の基に検索するシステム機能である。そこで、

文意解析処理システム機能について、図書館情報大学の石川徹也氏が研究・提案されているシステム機能¹⁾を基に開発を行い、ダイナミック・シソーラス機能の有効性を評価するために、質問文を収集、国内出版物の抄録を対象に評価実験を行う。

当研究の主な目標は、従来のキーワード指示検索方式に対し、質問文により直接検索を実行できることから検索システム機能として大きな改善を提案できるのではないかという可能性について、また、当システム機能は、データベース構築に多大な負荷となる人手による事前キーワード付与を不要とすることができて、データベース構築に多大の貢献をするのではないかという可能性について確認と見通しをつけることである。

以下、II章で書誌データベース検索システム機能について、データベース構築に於ける索引データ作成の課題および検索システムの課題を整理し述べ、III章で石川氏が提案されている「文意解析処理」技術に基づくキーワード抽出機能について解析を行い、IV章で当委託研究に於いて実現したダイナミック・シソーラス構築および実験・評価結果を示し、V章で当委託研究の結論を示す。尚、II章及びIII章については石川氏の転載許諾の元に文献1より引用した。

2 具体的取り組み

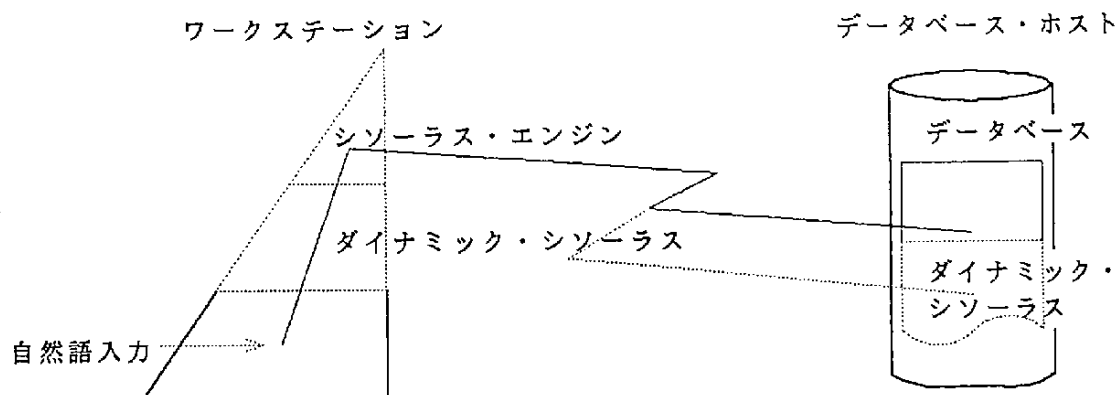
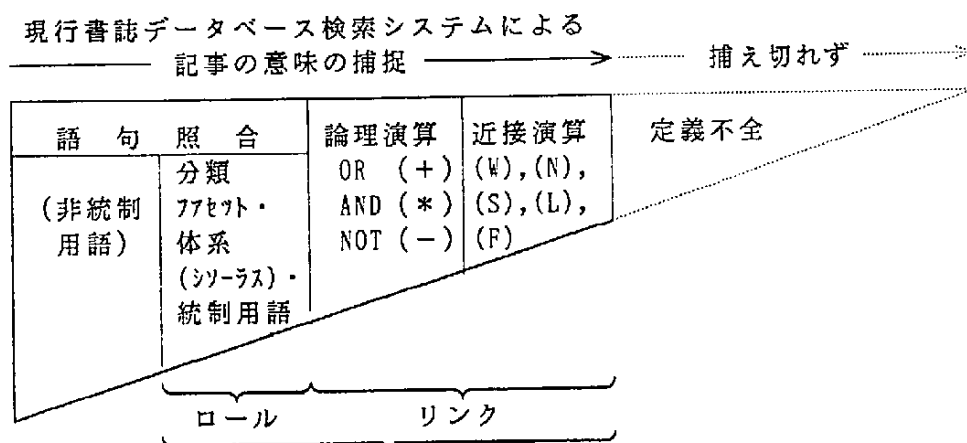
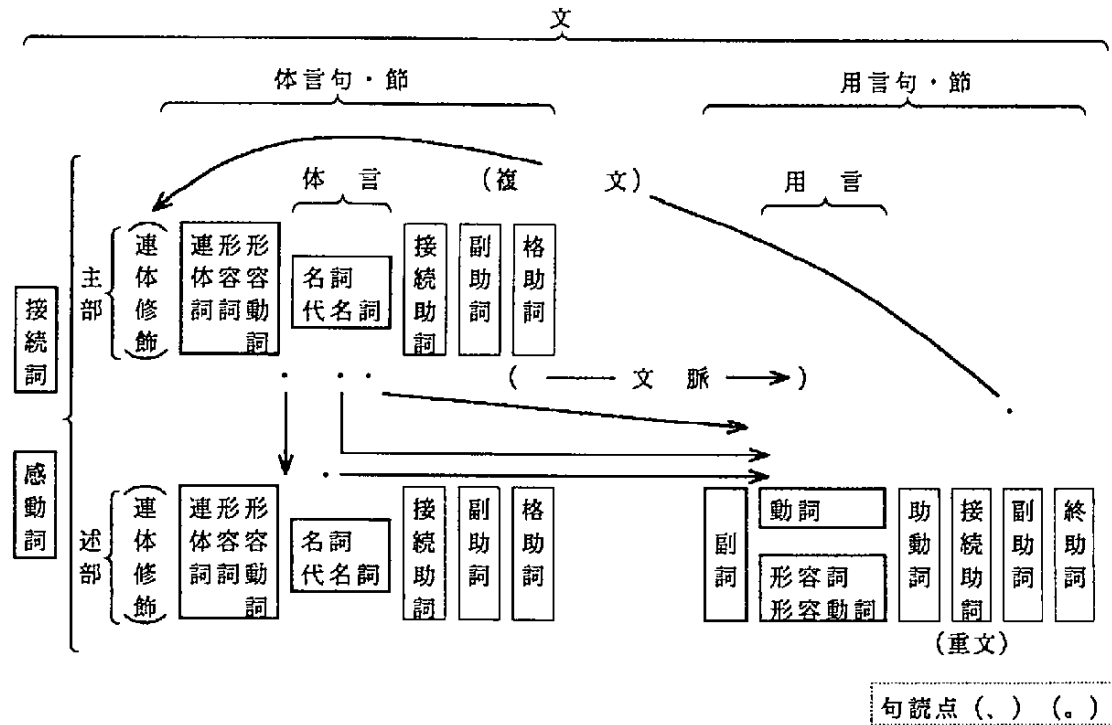


図 I.1: ダイナミック・シソーラス



(注) これらが作動する構造のデータベース記事の構築には多大の手間と費用を必要とし、かつ品質維持と均質化に難航している。
これらが作動するデータベース検索システムを使いこなすためには、データベースごとの深い学習と熟練を必要とし、かつ検索式の設定に難航（的外れ、定義不足、定義過剰など）している。

図 I.2: 現行書誌データベース検索システム



注1: 太枠は「自立語」

 細枠は「付属語」

注2: 当研究では用言の体系的意味解析に基づいて体言の意味・意図の範囲の自動的限定を試み、既に体系化が進んでいる体言の意味解析と連結して文全体の意味・意図を自動解析できる方法および体言の意味・意図を自動特定できる方法を探ることになる。

図 I.3: 日本語文の品詞・文節・文脈構成

II 書誌データベース検索システム機能

書誌データベース(以下、DBと記す)の構築に於ける索引データ自動作成システム及び検索システム機能について、技術的課題を中心に整理する。

1 索引語の機能

索引語機能に下記2点がある⁶⁾。

- 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。

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⁷⁾。

- 2) 主題索引語(Subject index term)…資料内容の主題、即ち、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^{8,9)}。

2 従来の研究と問題点^{10,11,12,13)}

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為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。

2.1 事前付与方式

2.1.1 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対象に解析し、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¹⁴⁾。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、例えば英文に於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とする。日本に於いて大規模キーワード方式の代表例に下記 ii) があり、ストップワード方式に二村氏らによるシステムがある¹⁵⁾。

ii) 単語間相関抽出法 複合語判定の為に、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全ての組み合わせを複合語とし、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、Stiles, H. E. により提案された¹⁶⁾。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、精度に於いて問題があり実用に至っていない。

2.1.2 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¹⁷⁾)、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。大規模キーワード・リストによる抽出法の日本に於ける実用例に、HAPPINESS を利用している前述雑誌記事索引及び新聞記事 DB 構築¹⁸⁾ 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。

2.1.3 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氏らの実用システム¹⁹⁾、構造化索引語抽出システムを目的とする Morita らの研究²⁰⁾がある。絹川氏らの方式は、5W1H に相当するキーワードに限定する為、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氏ら²¹⁾、形式動詞「である」に対応するキーワードを抽出する柴田氏ら²²⁾、字種により抽出する松崎氏ら²³⁾、「格」に対応するキーワードを抽出する木本氏²⁴⁾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 検索時設定方式

2.2.1 検索指示内容を直接利用する方法

検索語と並び例えば参考資料の表題を入力し、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照合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²⁵⁾。当方式は、指示される検索語若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、検索語若しくは抽出キーワードと DB 内の単語の照合に於いて欠落が生じる問題があり実用に至っていない。

2.2.2 シソーラス用語を利用する方法

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 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²⁶⁾。

この方式は、カテゴリ内の用語の範囲内に於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。

III 文意解析処理システム機能

情報検索システム及びデータベース構築の高度化機能として、情報要求者が検索語を設定したり、また索引語作成者が索引語を設定する際に、質問内容を解析したり、資料内容を解読しなくとも検索語及び索引語を設定可能な、例えば索引語作成支援 (MAI:Machine-Aided Indexing^{4, 5)}) システムが必要になる。

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。この種の要求に対して質問文及び例えば抄録内から主題キーワードを抽出し検索が出来なければならない。

主題キーワードは、“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”等の表現を認識し、その対象となる単語を抽出する文意解析処理システム機能を示す¹⁾。

1 システムの前提

- 1) 文意解析を行い、動詞の意味に注目し、目的とする名詞句(キーワード)あるいは記述文(キーセンテンス)を抽出・出力する。処理対象文章(テキスト)を VDT に表示、対応する処理結果のキーワードあるいはキーセンテンスを高輝度表示する。
- 2) 高輝度表示キーワードは自動的に、キーセンテンスに対し人間判断による単語をマウス・クリックし、両者を目的とするキーワードとしてシステム内に取り込む。
- 3) 名詞あるいは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、名詞辞書は持たない。動詞、助詞、副詞、接続詞のみの辞書により処理する。
- 4) 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。

2 システム機能の概要 (図III.1参照)

2.1 単位構文処理

文中に出現する動詞毎の構文(単位構文という)処理を行う。

・単位構文処理規則:

- 1) 1文を句点にて判別し取り出す。
- 2) 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、句接続詞及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。

例: ● 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。」

● 辞書データ例 (付録 A 参照。記号説明表参照。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))
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。但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 の結果スロットに埋める。以下、読点若しくは処理済みの単語まで判定し、その間の単

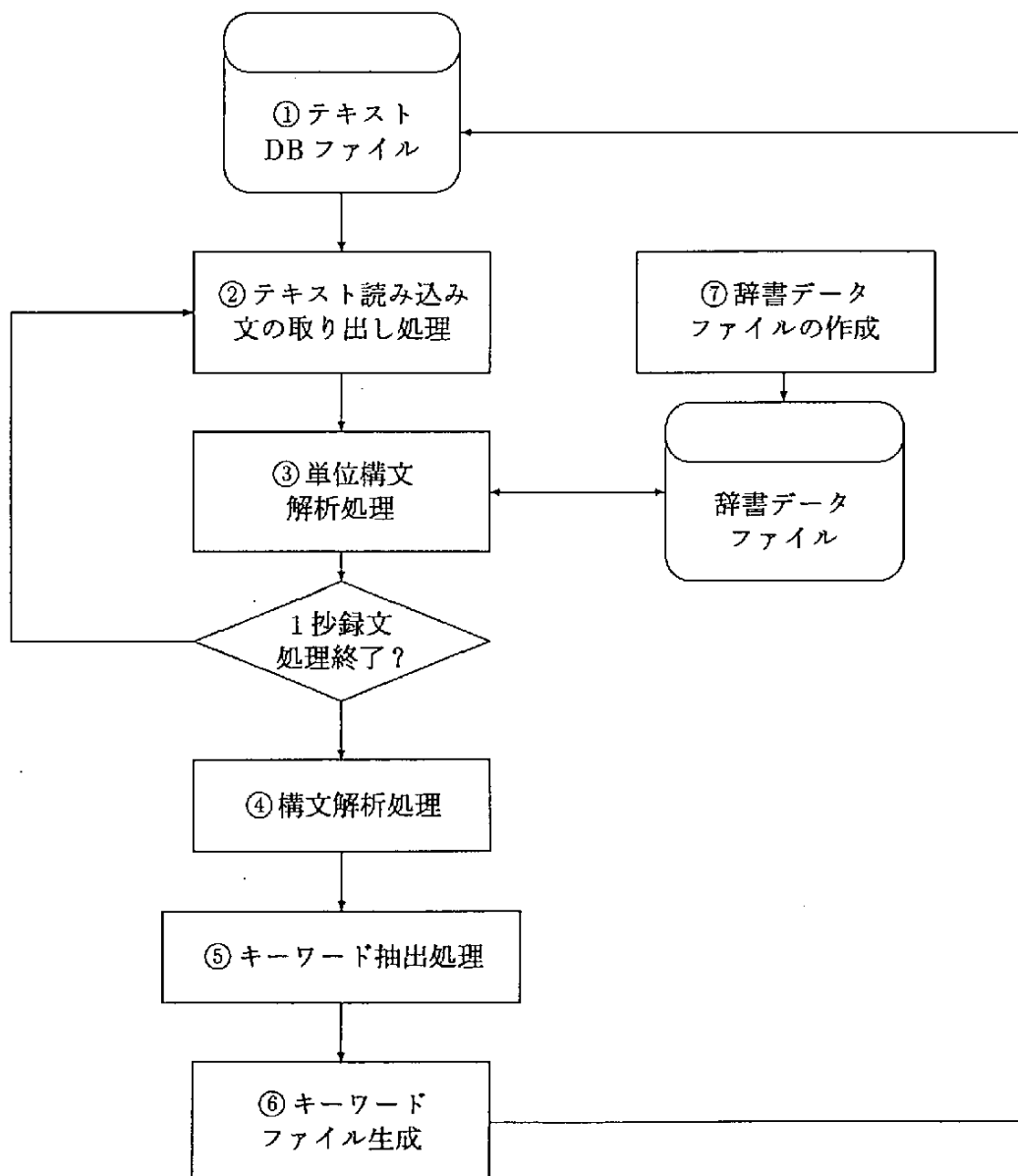


図 III.1: キーワード抽出システム機能図

語を助詞スロットの変項領域 (nil) に埋める。但し読点の直前が名詞の場合は格ラベルを (con) 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) 読点若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 (例:省略)。
- 4) 読点若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 (例:省略)。
- 5) 文接続詞は、その前後の処理済みスロットを () で囲む (例:省略)。
- 6) 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 (例:省略)。

2.1.1 構文処理

文毎の構文生成を行う為に、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。

・単位構文間埋め込み処理規則:

- 1) 連用修飾処理済みスロットを後出動詞処理済みスロットに埋め合成スロットを生成する。
- 2) 連体修飾処理済みスロットを後出の体言スロットに非交差規則を適用し埋め合成スロットを生成する。

・単位構文処理結果例

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。」

最初の動詞「用い」に対する処理結果

(ACTARE(用い)(連用中止形)(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)(連体形)(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)(終止形)(dur(で)(開発中))
(obj(を)(システム)))

・単位構文間埋め込み処理過程

処理1:文の最初の動詞「用い」は、連用中止形であるから処理規則1)を適用し、2番目の動詞「自動抽出する」の処理済みスロットに埋め合成スロットを生成する。

処理1の結果:

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

処理2:処理1の結果の合成スロットの最初の動詞「自動抽出する」は、連体形であるから処理規則2)を適用し、3番目の動詞「ある」の処理済みスロット内の体言「システム」のスロットに埋め合成スロットを生成する。

処理2の結果:

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

2.1.2 キーワード及びキーセンテンス抽出処理

構文解析処理結果を基に、キーワード及びキーセンテンス抽出処理を行う。

・キーワード及びキーセンテンス抽出処理規則(例)

- 1) 下記基本概念記号の動詞に対応する「obj」を主題表示キーワード若しくは「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」

運用動作動詞 (ACTAMA)…例:「行っている」

表現動作動詞 (ACTADE)…例:「述べる」

- 2) 資料毎に構文解析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」が無い場合は、文(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。

・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、情報処理学会全国大会講演論文集, Vol.27, No.2, pp.1067-1068(1983))

・構文処理結果

1 文めの解析結果

(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 文めの解析結果

(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 文めの解析結果

(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))))))

・主題表示キーワード及び主題表示キーセンテンス抽出処理結果

処理 1:1 文めの抽出処理結果…抽出処理非対象

処理 2:2 文めの抽出処理結果…抽出処理非対象

処理 3:3 文めの抽出処理結果…「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。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、自然索引語として「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。

付録 A 辞書データ例

● 動詞辞書

あげられる	(ACTADE (n i l) (T:p, M:p, A:a, C:終止))
上げる	(ACTACH (n i l) (T:p, M:a, A:a, C:終止))
与え	(ACTAMA (n i l) (T:p, M:a, A:a, C:連用中止))
与えた	(ACTAMA (n i l) (T:pt, M:a, A:a, C:連体))
与えられた	(ACTAMA (n i l) (T:pt, M:p, A:a, C:連体))
与えられている	(ACTAMA (n i l) (T:p, M:p, A:k, C:連体))
与えられない	(ACTAMA (n i l) (T:p, M:p, A:n, C:連体))
与える	(ACTAMA (n i l) (T:p, M:a, A:a, C:連体))
扱う	(ACTAMA (n i l) (T:p, M:a, A:a, C:連体))
扱っており	(ACTAMA (n i l) (T:p, M:a, A:k, C:連用中止))
表す	(ACTADE (n i l) (T:p, M:a, A:a, C:連体))
あり	(PRE (n i l) (T:p, M:a, A:a, C:連用中止))
ある	(PRE (n i l) (T:p, M:a, A:a, C:連体))
ある	(PRE (n i l) (T:p, M:a, A:a, C:終止))
いう	(CONCCI (n i l) (T:p, M:a, A:a, C:連体))
言えない	(CONCCI (n i l) (T:p, M:a, A:n, C:終止))
生かし	(ACTAMA (n i l) (T:p, M:a, A:a, C:連用))
至った	(CONCMO (n i l) (T:pt, M:a, A:a, C:終止))
至ったか	(CONCMO (n i l) (T:pt, M:a, A:?, C:終止))
位置づけ	(ACTAIN (n i l) (T:p, M:a, A:a, C:連用中止))
一般化された	(ACTAOP (n i l) (T:pt, M:p, A:a, C:連体))
一般化しすぎたり	(ACTAOP (n i l) (T:p, M:a, A:a, C:連用))
埋め込んだ	(ACTAOP (n i l) (T:p, M:p, A:a, C:連体))
英文解析できる	(ACTAIN (n i l) (T:p, M:a, A:ab, C:連体))
得た	(ACTARE (n i l) (T:pt, M:a, A:a, C:終止))
得ている	(ACTARE (n i l) (T:p, M:a, A:k, C:終止))
選び	(ACTAIN (n i l) (T:p, M:a, A:a, C:連用中止))
選ぶ	(ACTAIN (n i l) (T:p, M:a, A:a, C:連体))
得られた	(ACTARE (n i l) (T:pt, M:p, A:a, C:終止))
得られた	(ACTARE (n i l) (T:pt, M:p, A:a, C:連体))

● 助詞辞書

か	(c o m (か)) → A C T A I N
が	(s u b (が)) → P R E
が	(s u b (が)) → A C T A D E
が	(s u b (が)) → A C T A C O
が	(s u b (が)) → A C T A I N
が	(s u b (が)) → A C T A R E
が	(s u b (が)) → A C T A O P
が	(s u b (が)) → A C T A M A
が	(s u b (が)) → A C T A C H
が	(s u b (が)) → A C T A C R
が	(s u b (が)) → C O N C I N
が	(s u b (が)) → C O N C C I
かの	(c o n (かの)) → A C T A C O
かは	(c o n (かは)) → C O N C I N
から	(c o m (から)) → P R E
から	(c o m (から)) → C O N C C I
から	(c o m (から)) → A C T A C R
から	(c o n (から)) → A C T A D E
から	(o r i (から)) → A C T A C O
から	(o r i (から)) → A C T A C R
から	(o r i (から)) → A C T A R E
から	(o r i (から)) → A C T A I N
から	(p u r (から)) → C O N C C I
から	(s f r (から)) → A C T A M O
から	(s f r (から)) → C O N A I N
から	(s o u (から)) → A C T A C H
から	(s o u (から)) → C O N A I N
から	(t f r (から)) → A C T A I N
から	(v i e (から)) → A C T A C O
からの	(o p p (からの)) → P R E

● 副詞辞書

3 - 3 0 倍	(d e g (3 - 3 0 倍))
あらかじめ	(t i m (あらかじめ))
新たに	(d e g (新たに))
いかにして	(m a n (いかにして))
大きく	(d e g (大きく))
多く	(d e g (多く))
主に	(i m p (主に))
必ず	(i m p (必ず))
かなり	(d e g (かなり))
近年	(t i m (近年))
効率よく	(d e g (効率よく))
効率良く	(d e g (効率良く))
これまで	(t t o (これまで))
盛んに	(d e g (盛んに))
さらに	(a n d (さらに))
若干	(d e g (若干))
精度良く	(d e g (精度良く))
そのまま	(s t a (そのまま))
第一に	(o r d (第一に))
第二に	(o r d (第二に))
例えば	(e x s (例えば))
直接	(s t a (直接))
次いで	(o r d (次いで))
通常	(s t a (通常))
次ぎに	(o r d (次ぎに))
続いて	(o r d (続いて))
とくに	(i m p (とくに))
特に	(i m p (特に))

● 接続詞辞書

文接続詞

あるいは	(/ (あるいは))
しかし	(b u t (しかし))
したがって	(r e s (したがって))
そこで	(r e s (そこで))
そして	(a n t (そして))
また	(a n t (また))

句接続詞

および	(a n t (および))
及び	(a n t (及び))
が	(b u t (が))
かつ	(a n t (且つ))
ため	(f o r (ため))
ために	(f o r (ために))
と	(r e s (と))
なお	(r e s (なお))
ならびに	(a n t (ならびに))
のみならず	(n o t (のみならず))
または	(/ (または))

記号表 (例)

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	变化状态動詞
CONCCI	状況状态動詞
CONCIN	意図状态動詞

動詞の時制記号 (形式 T : x)

記 号 (x=)	意 味
p t	過去 (past)
p	現在 (present)
f	未来 (future)

動詞の様態記号 (形式 M : y)

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

動詞のアスペクト記号 (形式 A : z)

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

動詞の活用形記号 (形式 C : w)

記 号 (w =)	意 味
連体	連体形
連用	連用形
終止	終止形
連用中止	連用中止形

格（助詞・助詞相当語意味）ラベル記号

記 号	意 味
sub	動作・行為の主体 (subject)
obj	動作・行為の対象 (object)
rec	受け手 (recipient)
ori	与え手 (origin)
par	相手1 (partner)
opp	相手2 (opponent)
tim	時 (time)
tfr	時・始点 (time-from)
tto	時・終点 (time-to)
dur	時間 (duration)
spa	場所 (space)
sfr	場所・始点 (space-from)
sto	場所・終点 (space-to)
sth	場所・経過 (space-through)
sou	始状態 (source)
goa	終状態 (goal)
att	属性 (attribute)
cau	原因・理由 (cause)
too	手段・道具 (tool)
mat	材料 (material)
com	構成要素 (component)
man	方式 (manner)
con	条件 (condition)

副詞意味記号

記 号	意 味
and	追加 (and)
ant	添加 (and then)
deg	度合 (degree)
exs	实例 (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)

IV システムの構築および実験・評価

1 構築システム

書誌データベース用ダイナミック・シソーラスの可能性を確認するために、実験システムを開発する。実験のために構築するシステムは図書館情報大学の石川徹也氏がアルゴリズムを開発された「文意解析手法」を参考にし、新たに実験プログラムを作成する。以下、この文意解析手法を「文意解析処理方式」と称する。

1.1 プロトタイプ処理方式概要

評価・実験を行なうシステムは UNIX 上で作成する。プロトタイプ作成に使用したシステム構成を示す。

システム	機種	OS 等
ワークステーション	SONY-NEWS/3860	NEWS-OS 3.91R
ワークステーション	SUN SPARCstation 2	SUN-OS 4.1.1
プリンタサーバ	IMAGEN 7320 Turbo	OS 4.1T

実験システムで扱う入力と出力の概念を次に示す。

例題:

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。

最終結果 (S 式風)

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

最終結果 (テキスト体)

1 構築システム

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

上記の2つの最終結果は、どちらも意味的には等価である。

今回の実験では、テスト用に作成した書籍データベース1万件(全5万項目)のデータについて、助詞・副詞などについての予め文例を調査の上、実験用プログラムを作成した。1万件のデータのうち「要旨」をもつものは約4,000件である。

処理概要を下記に示す。

- 1) 「は、」のような読点を削除する。(前処理)
- 2) 「。」で処理単位を分割する。
- 3) 分割した処理単位毎に動詞・助詞を目印にして分解・解析処理を行なう。
- 4) 今回の目的とする抽出処理対象でないものは削除する。
- 5) 抽出処理対象のキーワードを出力する。

1.2 処理内容

1.2.1 前処理

前処理では次のような処理を行なう。

- 1) 次のようなパターンをもつ文字列の読点(、)を削除する。
と、 や、 が、 は、 を、
- 2) 次のようなパターンを持つ読点(、)を中黒(・)に変換する。
[漢字またはカタカナ]、[漢字またはカタカナ]

1.2.2 分解・解析処理

- i) テスト用書籍データベース テスト用書籍データベースの要旨からキーワードを抽出する。動詞として認識したものは次のものである。

あった	検証し	紹介し、	分析する
ある	行なわれた	紹介する。	保証されている

つくる	参加した	進む	補完する
できなかった	惨敗した	整理した	明確化する
とり上げ	使いやすい	生かした	問う
わかる	至る	設け	役立てる
解説した	取り入れて	先駆けて	予定されている
解説している	収めてある。	託して	利用して
解説する	収めている。	統合した	理解する
概観している	集めました。	把握する	剔抉する
拡張された	出題された	分析し	

これらの動詞の内、次の動詞に注目して、各々の書籍を的確に表現しているキーワードを抽出する。

- 1) 解説した
- 2) 解説している
- 3) 解説する
- 4) 収めている
- 5) 収めてある
- 6) 紹介し
- 7) 紹介する
- 8) 集めました
- 9) 検証し
- 10) 明確化する
- 11) 剔抉する

ii) 質問文 質問文に対しては次の動詞を抽出する。

- 1) 知りたい。
- 2) 可能であろうか。

2 処理結果

3) もたらすのか。

4) 影響するのか。

5) 何であろうか。

1.2.3 抽出処理

前記の動詞について意味情報を抽出し、書籍の内容を表現するにふさわしい名詞をキーワードとして抽出する。また質問文からは、その質問の核心に触れる単語を抽出し、データベース検索時のキーワードとして採用する。

1.2.4 後処理

「文意解析処理方式」で抽出したキーワードは、文章の一部分を切りだしたものであるの
で、検索用のキーワードとして使うために、更に一般化処理が必要となる。今回はこのため
に、データベース用語管理システムである「NICE システム」を利用した。この NICE シス
テムは日外アソシエーツが開発したシステムである。

2 処理結果

今回評価のために作成したデータベースの内容の一部を付録 A に示す。データベースの
内容は下記の内容が含まれている。

整理番号＝データベース作成時の整理番号

書籍書名＝書籍の名称

出版社名＝出版社の名称

著者名前＝著者・執筆社・編集社などの名前

要旨概要＝書籍の要旨・概要。キーワード作成の原データとなる。

I S B N＝国際書籍識別番号

2.1 前処理

テスト用データベースに対して前処理を行なった結果の一部を付録 B に示す。結果の表
示は次のようになっており、出版社や ISBN 等対象とならない部分は省略してある。

DTS=(整理番号) 書籍書名

前処理を行なった要旨概要部分

2.2 「要旨」からの抽出処理

テストデータベースの要旨部分からキーワードを抽出処理する。次に整理番号 01724 についての結果を示す。

2.2.1 原文

DTS=(01724) 書物の現在

映像の優位・活字文化の衰退という俗説を排し、グーテンベルクから電子出版に至る《書物》の歴史と現在をグローバルに検証しつつ、現在のそして未来の文化にとっての《書物》の揺るぎない重要性を2人の研究者（浅沼／清水）が明確化すると共に、今日の出版、とりわけ雑誌出版の問題を『試行』と『リュミエール』の2人の《編集者》（吉本／蓮実）が鋭く剔抉する。文学・出版・印刷関係者必読の書。

排し	俗説(o:を)	映像の優位・活字文化の衰退(という)
至る	電子出版(に)	グーテンベルク(から)
	検証し	グローバル(に) 書物の歴史と現在(o:を)
	明確化する	2人の研究者(浅沼／清水)(が)
		書物の揺るぎない重要性(o:を) →
	剔抉する	試行とリュミエールの2人の編集者(吉本／蓮実)(が) →
		とりわけ雑誌出版の問題(o:を)

これからこの書籍の中心になるキーワードは

- 書物の歴史と現在
- 書物の揺るぎない重要性
- とりわけ雑誌出版の問題

が検出される。

2.3 「質問文」からの抽出処理

質問文として次の文章を使った。

2 処理結果

- 1) ビデオが普及し、CD-ROM や CD-I などの光ディスクによる電子出版が実用化され、テレマティークなどの電子図書館の研究開発も進んでいるが、これらは従来の書籍や雑誌へどう影響するのか。
- 2) 脳とは何であろうか。並列学習情報処理の原理を核とした脳の原理の解説と、ニューロコンピュータを含む応用例はないか。

最初の文についての解析結果は次のようになる。

普及し	ビデオ (が)	
実用化され	電子出版 (が)	CD-ROM や CD-I などの光ディスク (による)
	進んでいる	テレマティークなどの電子図書館の研究開発 (も)
	影響するのか	従来の書籍や雑誌 (o:へ) これら (は)

これから、次の質問に関するキーワードが抽出される。

- 従来の書籍や雑誌

2.4 後処理

「文意解析処理方式」でキーワードを抽出した結果に対し、NICE システムで同義語・上位語を付加して実際の検索用キーワードを生成する。以下「文意解析処理方式」とは NICE 処理を含んだものを言う。

2.4.1 「要旨」の後処理

「要旨」から抽出されたキーワードを更に NICE システムで処理を行ない、次の最終結果を得る。

書物	シヨモツ
歴史	レキシ
雑誌	ザツシ
出版	シュツパン

2.4.2 「質問文」の後処理

「質問文」から抽出されたキーワードを更に NICE 処理を行なうと、次の検索用キーワードを得る。

書籍	シヨセキ
出版	シユツパン
図書	トシヨ
本	ホン
雑誌	ザツシ

2.5 検索処理

検索は次の4つの方式について実験を行ない検討する。

キー作成方式	質問文	書籍データベース
方式1	文意解析処理方式	文意解析処理方式
方式2	文意解析処理方式	全文検索方式
方式3	文意解析処理方式	索引候補語リスト方式
方式4	人手分析	索引候補語リスト方式

2.5.1 方式1

データベースのキー作成と質問文の処理に「文意解析処理方式」を行なった場合の結果を示す。

DB キー (整理番号 1724)		質問文キー	
表記	読み	表記	読み
書物	シヨモツ	書籍	シヨセキ
歴史	レキシ	出版	シユツパン
雑誌	ザツシ	図書	トシヨ
出版	シユツパン	本	ホン
		雑誌	ザツシ

この結果、テストデータベースの整理番号=1724 のデータ項目は、太字のキーが一致しているので質問文によって検索されることになる。

2.5.2 方式2

質問文に対し「文意解析処理方式」を適用して得られた結果でデータベースの要旨部文を全文一致検索した場合、方式1の質問文キーを含む全ての項目が検索される。

テストデータベースの整理番号=1724のデータ項目は、

映像の優位・活字文化の衰退という俗説を排し、グーテンベルクから電子出版に至る《書物》の歴史と現在をグローバルに検証しつつ、現在のそして未来の文化にとっての《書物》の揺るぎない重要性を2人の研究者（浅沼／清水）が明確化すると共に、今日の出版、とりわけ雑誌出版の問題を『試行』と『リュミエール』の2人の《編集者》（吉本／蓮実）が鋭く剔抉する。文学・出版・印刷関係者必読の書。

であり、太字の部分が一致する。

2.5.3 方式3

質問文に対し「文意解析処理方式」を適用して得られた結果を、データベースの要旨部分に対し「索引候補語リスト方式」で作成したキーで検索する。

テストデータベースの整理番号=1724のデータ項目に対する「索引候補語リスト方式」の処理結果は次のようになる。

エディター	エディター
グロ	グロ
グロテスク	グロテスク
ハス	ハス
印刷	インサツ
映像	エイゾウ
活字文化	カツジブンカ
芸術	ゲイジユツ
検証	ケンシヨウ
研究者	ケンキュウシヤ
雑誌*	ザツシ
史	シ
出版*	シュツパン
書物	シヨモツ
水	ミズ
清	シン
中国史	チュウゴクシ
電子出版	デンシシュツパン
文化	ブンカ
文学	ブンガク
編集者	ヘンシユウシヤ
未来	ミライ
明	ミン
歴史	レキシ
連	ハス

「索引候補語リスト方式」で得られたテストデータベースの整理番号=1724のデータ項目の内*印が付いているキーワードと、「文意解析処理方式」で得られたキーワードが一致する。

2.5.4 方式 4

これは、現在までの一般的な方式である。質問文に対して人間が人手でキーワードを選択し、「索引候補語リスト方式」で付されたキーワードに対して検索を行なう。質問文に対する検索用キーワードの選定は個人的な力量に依存する。

今回はかなり高度な能力を持ったサーチャーが選定されると思われるキーワードを採用した。

質問文: ビデオが普及し、CD-ROM や CD-I などの光ディスクによる電子出版が実用化され、テレマティークなどの電子図書館の研究開発も進んでいるが、これらは従来の書籍や雑誌へどう影響するのか。

検索キーワード: 電子出版 or 電子図書館

この結果、検索キーワードの「電子出版」がデータベースのキーワードと一致する。

2.6 結果の評価

結果の評価として、「文意解析処理方式」は対象とする文章からその意味内容を自動的にかなりの的確に拾いだしている。データベース項目の要旨を処理した場合を例にとれば、方式 1 と方式 3 の結果を比較すると、「文意解析処理方式」では 4 件のキーワードが抽出されたのに対し、従来の「索引候補語リスト方式」では 25 件のキーワードが抽出されている。この場合数の多少よりは、その文章の意味内容を現していないノイズの方が多くなっている。

「索引候補語リスト方式」は、できる限り多くのキーワードを付与しようとするのに対し、「文意解析処理方式」はその文章の核心部分を現すキーワードのみを付与しようとする試みであることが、この実験からも確認できた。

付録 A テストデータベース内容例

整理番号=01724

書籍書名=書物の現在

著者名前=吉本/隆明; 蓮実/重彦; 清水/徹; 浅沼/圭司【著】

要旨概要=映像の優位、活字文化の衰退という俗説を排し、グーテンベルクから電子出版に至る《書物》の歴史と現在をグローバルに検証しつつ、現在のそして未来の文化にとっての《書物》の揺るぎない重要性を2人の研究者(浅沼/清水)が明確化すると共に、今日の出版、とりわけ雑誌出版の問題を『試行』と『リュミエール』の2人の《編集者》(吉本/蓮実)が鋭く剔抉する。文学、出版、印刷関係者必読の書。

ISBN=ISBN4-89176-224-1

整理番号=01730

書籍書名=准看護婦(士)試験

出版社名=国家試験シリーズ(41)

要旨概要=本書は過去に行われた准看護婦資格試験の問題を、それぞれの科目に分けて提出し、これに1題1題、きわめて適切でいねいな解説をつけ、これによって正解を導いています。

ISBN=ISBN4-7703-1115-X

整理番号=06399

書籍書名=図書館—その本質・歴史・思潮

著者名前=岡田/温【著】

要旨概要=図書館は時の流れと共にさまざまに変化してきました。そしていま、かつては夢であった電子図書館さえも誕生しようとしています。それでこの「図書館」は、どのようにして生まれてきたのでしょうか。本書では、文字文化、記録文化の初期といわれる古代オリエン特時代にまで遡り、図書館という社会現象がどのようにして現われ変遷してきたかを、今日までの内外の文化史に沿って興味深く解説しています。

ISBN=ISBN4-621-03362-X

整理番号=06995

書籍書名=インタプレス—電子出版のためのページ記述言語

著者名前=Harrington, Steven J.; Buckley, Robert R. 【著】; 富士ゼロックス【訳】

要旨概要=インタプレスは、世界で初めての最も実用的なページ記述言語である。プロの仕事のようなドキュメントを、ネットワーク上のあらゆるプリンタに、美しく、素早く、そして効率的に印刷させることができる。読みやすさと、徹底した技術的正確さによって、読者は本書を通して基本的な概念を最も進んだ使い方を段ることができ、インタプレスを内と外から理解することができる。

ISBN=ISBN4-621-03350-6

整理番号=07021

書籍書名=高性能コンピュータアーキテクチャ

著者名前=Stone, Harold S. 【著】; 斉藤/忠夫; 発田/弘【監訳】

要旨概要=コンピュータアーキテクチャはその時々のデバイス技術に大きく影響され、現在普及しているマシンは必ず古くさいものになってしまう。本書は、このようなマシンの変遷の中で一貫して不変のものである設計の方法論と評価手法を集中して扱い、さらに実際の設計過程を豊富にとり入れた演習問題によって設計のこの2つの柱を完全に理解させるものである。大形高性能機のアーキテクチャ設計は全ての設計の土台となるものであり、本書はその肝要を十分に解説した待望の一書である。

ISBN=ISBN4-621-03354-9

整理番号=13402

書籍書名=並列処理マシン

著者名前=富田/真治; 末吉/敏則【共著】

出版社名=コンピュータアーキテクチャシリーズ

要旨概要=数値計算や人工知能分野などでのコンピュータの高速化要求に伴い、ますます期待が高まっている並列処理マシンについて、構成方式から超並列処理への展望まで具体的に解説。

ISBN=ISBN4-274-07501-X

整理番号=13403

書籍書名=クリストファー・アレグザンダー—建築の新しいパラダイムを求めて

著者名前=Grabow, Stephen 【著】(Grabow, Stephen); 吉田/朗; 長塚/正美; 辰野/智子【訳】

要旨概要=ポストモダンの建築の旗手クリストファー・アレグザンダー。「パタン・ランゲージ」「センタリング・プロセス」などのキーコンセプトを次々と提出し、建築パラダイムの再構築を図るアレグザンダーの全

映像を余すところなく描き出し、建築界のみならず、現代思想全般に衝撃を与える傑作評伝。

ISBN=ISBN4-87502-153-4

整理番号=13406

書籍書名=ぶきっちょさんのフルーツのお菓子

著者名前=森山/サチ子【著】

出版社名=ぶきっちょさんのCOOKINGNOTE-FRUITDESSERTS

要旨概要=フルーツのお店やデパートのフルーツコーナーには、1年中、色とりどりのいろいろな形をしたたくさんのフルーツ達が顔を並べています。フルーツとお菓子の関係は非常に古く、その相性は昔から認められており、焼き菓子から冷たいお菓子、そしてジャムなどの保存食にいたるまで、それぞれの特性と美しい色や形を活かして用いられています。さあ、みなさんもフルーツを使つて、様々なお菓子作りを始めてみて下さい。

ISBN=ISBN4-277-61313-6

整理番号=13407

書籍書名=インテリアコーディネーターの基礎知識

著者名前=古谷/彰子【著】

要旨概要=インテリアコーディネーターとして必要な知識、特に女性に身近なインテリアエレメント関係に重点を置いてわかりやすく解説。新時代の職業を目指す女性を応援します。

ISBN=ISBN4-88470-690-0

整理番号=13423

書籍書名=雑誌新聞総かたろぐ〈1989年版〉

要旨概要=本誌は、毎年1回刊行いたします。本号に収録してあるデータは、2月現在のものです。全ての定期刊行物の内容・特色までを完全に調査してあります。

ISBN=ISBN4-89554-020-0

整理番号=13446

書籍書名=PC・9800シリーズアプリケーション百科〈1989年版〉

著者名前=新紀元社編集部；知識計画【編】

要旨概要=PC・9800シリーズアプリケーションの最新情報を網羅。「用途で探す、内容で比べる、比較・検討のためのソフトウェア情報」を、一冊に集大成。

ISBN=ISBN4-915146-11-1

整理番号=13449

書籍書名=実用漢方処方集

著者名前=日本漢方協会【編】

要旨概要=現在使用されている漢方処方集に記載されている薬方をすべて収載することに努めた（合計約1,600）。

ISBN=ISBN4-8407-1008-2

整理番号=13457

書籍書名=がんの予防

著者名前=小林/博【著】

出版社名=岩波新書〈71〉

要旨概要=がんは、いまや死に至る病ではない。がんの実体はかなり究明され、診断から治療までの医療技術も格段に向上し、予防に関する科学研究も進んだ。ここでは、がんの予防を、細胞レベルでのがん化の阻止、早期発見・早期治療、転移・再発の予防と3段階に分け、それぞれ対策を考える。がん研究の最前線からの最新の報告。

ISBN=ISBN4-00-430071-1

付録 B 「要旨」部分の前処理結果例

DTS=(01724) 書物の現在

映像の優位・活字文化の衰退という俗説を排し、
グーテンベルクから電子出版に至る《書物》の歴史と現在をグローバルに検証しつつ、
現在のそして未来の文化にとっての《書物》の揺るぎない重要性を2人の研究者（浅沼／清水）が明確化する
と共に、
今日の出版、
とりわけ雑誌出版の問題を『試行』と『リュミエール』の2人の《編集者》（吉本／蓮実）が鋭く剔抉する。
文学・出版・印刷関係者必読の書。

DTS=(01730) 准看護婦（士）試験

本書は過去に行われた准看護婦資格試験の問題を、それぞれの科目に分けて提出し、
これに1題1題、
きわめて適切でいねいな解説をつけ、
これによって正解を導いています。

DTS=(01732) 味吉兆で学んだ日本料理

味吉兆に通うこと10年。
9人の料理研究家が聞き取った献立、
その数100有余。
ていねいな作り方の記載と詳細な分量の表示。
今・第一級の料理書として見事に結実。
日本料理の粋がここにある。

DTS=(01736) 実習 パソコン BASIC入門

本書の実習内容に従って、
プログラムを処理することによって、
知的生産の1つの道具としてパソコンが使えるようになることを目標にし、
BASICの文法およびプログラミング技術が確実に習得できるように企画した。

DTS=(01756) 外国車カタログ〈'89〉

おしゃれで価格もリーズナブルな欧州車。
パワーを誇るアメ車。
'89年の外国車をすべてラインナップした魅力の総合アルバムが完成。

DTS=(01775) コンピューターシステムの落とし穴—企業がとるべき安全対策と危機管理

あなたのシステムが災害・事故・エラー・犯罪・人正行為等の“落とし穴”に落ちる前に打つべき対策を多くの
事例を通して警告する。

DTS=(01778) みんなの健康宣言—医療改革への提言

心身の健康は個人で解決できる問題ではない。
「みんなの健康はみんなで」という立場に立ち、
市民の力を結集する中に、
健康で快適な明日の社会をひらく展望がある。
みんなの健康をつくるための政策提言集！慢性病・病院・介護・保険・高負担社会へ向けた提言。

DTS=(01786) ストレスくん、

さようなら
ストレスにおしつぶされそうになる、
そんなあなたを救う現代版ストレス解消の“傾向と対策”が本書です。
敵の正体をつきとめて、
どうやっつけるか、
あるいはストレスをどうやって味方にしてしまうか、
これさえ身につけておけば、
ストレスなんかちっともこわくありません。

DTS=(01792) なにわの味師 24人

食い倒れの町で味一筋に生きる名匠たちの人柄・料理哲学から人生観まで。

DTS=(01800) ツルの渡る日

冬枯れの出水平野（鹿児島県）に、
毎年訪れるツルたち、
その数 8 0 0 0 羽あまり。
彼らの生態、
おおらかで優美な姿を詳さに描きながら、
「渡り」の不思議に迫る。

DTS=(01807)MACの達人—紀田\&松田のFAX 交遊録
深遠なるアップルの精神に魅了された文芸評論家と商社マン。
知の開発メディアとも言えるマックに彩られた彼らふたりの日常生活を3年間に及ぶFAX通信を通して綴る
紀田&松田のマッキントッシュ・ワールド。

DTS=(01809)ワープロ通信で情報通になる本—はじめての人でもらくらくアクセス
本書ははじめての人がワープロ通信を行なう前に、
知っておきたい基礎知識を平易に紹介し、
NECのPC・VANを例に設定方法などを誌上で体験していただく。
また、
各種ネットワークのサービス内容をはじめ、
具体的な活用事例をも広範に紹介する。

DTS=(01820)著作権のノウハウ
広範囲にわたる著作権の問題点を62問にまとめ、
わかり易く解説。
第3版増補では昭和63年の著作権法改正による修正を行ったほか、
最近・議論になっている「新しい創作物への対応並びに著作物の新しい利用形態への対応について」の章を加え、
さらに内容の充実をはかりました。

DTS=(01831)2級建築士試験問題と解説〈平成元年版〉
本書は昭和60年においては全国47都道府県が出現された問題のうち、
非常に優れていると考えられる問題を各科40題ずつ選び、
昭和61年～63年においては出題された全問題を収録し、
それに解説を加えたものです。

DTS=(01835)国立国会図書館百科
わが国唯一の国立図書館である国立国会図書館は昭和23年に設立された。
議会図書館として、
また中央図書館としての機能を持ち、
文化財の蓄積と利用を目的にすべての国内出版物を収集する納本図書館である。
本書は同館開館40周年を記念して出版するもので、
その沿革・資料管理・各種利用法・所蔵資料の解題・国会サービス・支部図書館制度・納本制度等々羅網的に要約・解説したものである。

DTS=(01839)新版 電気工事関係法規集
本書は電気工事士を目指す方および電気工事の作業に従事する方が必ず知っておかなければならない電気工事士法ならびに電気工事業の業務の適正化に関する法律、
いわゆる電気工事二法を中心に関係法規を集録したものである。

付録 C 「要旨」部分の処理結果例

DTS=(B8900081)UNIX System5 リリース 4.0 ソフトウェアデベロッパコンファレンス'88 セミナーテキスト
本書はSVR4.0の発表(a:に)(x:先駆けて)このシステムのパワフルな機能と(x:使いやすい)ユーザインタ
フェースOPEN LOOKの詳細(o:を)(v:紹介し)SVR4.0発表と同時に(a:に)タイムリーなリリース
(o:を)可能にする

obj=ユーザインタフェースOPENLOOKの詳細
obj=タイムリーなリリース

DTS=(B8900084)4訂食品成分表 1989

この冊子(a:は)前半(a:に)科学技術庁資源調査会編四訂日本食品標準成分表(o:を)収め後半(a:に)保健栄
養(a:に)関する諸表(o:を)(v:収めている)

obj=科学技術庁資源調査会編四訂日本食品標準成分表
obj=関する諸表

DTS=(B8900092)水理学 2

2巻(a:は)広範な応用面をもつ水理学(a:が)将来さらに分化進展することを念頭におき題材としてやや高度
なものを含め重要かつ基本的テーマ(o:を)選び最新の知識とデータ(o:を)(x:取り入れて)(v:解説している)
obj=将来さらに分化進展することを念頭におき題材としてやや高度なものを含め重要かつ基本的テー
マ

obj=選び最新の知識とデータ

DTS=(B8900237)いま病院で看護してみたい

新版1級建築士受験講座(a:は)単(a:に)試験の正解(o:を)(x:理解する)だけでなく建築(o:を)全般にわ
たって広く学ぶために試験問題となりやすい点(a:に)ウェイト(o:を)置きその周辺(a:に)関しても幅広く
(v:解説している)

obj=試験の正解
obj=だけでなく建築
obj=ウェイト

DTS=(B8900378)解析学の基礎

すなわち常微分方程式関数論フーリエ解析偏微分方程式ベクトル解析(o:を)中心にその主要部分(o:を)丁寧
(a:に)(v:解説している)

obj=すなわち常微分方程式関数論フーリエ解析偏微分方程式ベクトル解析
obj=中心にその主要部分

DTS=(B8900426)物理学への招待

本書(a:は)文科系一般教育向きに書かれたテキスト(a:で)(x:ある)が理科系学生の使用も考慮し物理学全般
の内容(o:を)体系立てて興味深く(v:解説している)

obj=が理科系学生の使用も考慮し物理学全般の内容

DTS=(B8900457)ISDN(サービス総合デジタル通信網)の基礎を知る事典

本書(a:は)88年4月にわが国(a:で)実用サービス(a:が)開始されたISDNサービス総合デジタル通信
網の基礎知識(o:を)豊富な図版をまじえて平易(a:に)(v:解説した)事典型式の入門書です

obj=開始されたISDNサービス総合デジタル通信網の基礎知識

DTS=(B8900469)現場で役立つシーケンス回路

この本(a:は)現場の技術者管理者(o:を)対象(a:に)機械の加工組み立てラインにおける加工忘れ加工ミス
誤組み欠品組立ミスなどを防止する装置ボカヨケというを作るためのシーケンス回路の基礎知識と具体的な応
用例(o:を)多数(v:紹介し)た実務書です

obj=現場の技術者管理者
obj=機械の加工組み立てラインにおける加工忘れ加工ミス誤組み欠品組立ミスなどを防止する装置ボ
カヨケというを作るためのシーケンス回路の基礎知識と具体的な応用例

DTS=(B8900478)MSX2+パワフル活用法

本書はMSX2の(x:拡張された)機能(o:について)(v:解説する)とともにその機能(o:を)具体的(a:に)活
用したプログラム(o:を)(v:紹介し)ています

obj=機能
obj=とともにその機能
obj=活用したプログラム

DTS=(B8900496) 電磁気学

前半 (a:では) 電磁気学 (a:で) 基本的 (a:に) 重要な静電場静磁場 (o:について) 豊富な図解 (o:を) 用いて丁寧 (a:に) (v:解説する)

obj=重要な静電場静磁場

obj=豊富な図解

DTS=(B8900623) 一般教養 生命の科学

本書はこうした現状をふまえ分子レベルや細胞レベル (o:を) 中心とした現代生物学 (o:を) (x:理解する) のにかかせない主要な知識をわかりやすく (v:解説した) もので (x:ある)

obj=本書はこうした現状をふまえ分子レベルや細胞レベル

obj=中心とした現代生物学

DTS=(B8900663) 物理学最前線 23

酸化物超伝導体 (a:に) 共通 (a:に) 見られる結晶構造の特徴 (o:を) 平易 (a:に) (v:解説する) とともにこれらの超伝導体の開発 (a:に) 携わる研究現場の姿 (o:を) 生々しいエピソード (o:を) 混じえて紹介高温超伝導体の結晶構造

obj=見られる結晶構造の特徴

obj=携わる研究現場の姿

obj=生々しいエピソード

DTS=(B8900819) 男をパワーアップさせる秘薬秘術中国四千年の歴史が教える即効の健康強精法

手軽 (a:に) 手に入る材料 (o:を) 使ったマル秘漢方処方と健康食 (o:を) 一挙 (a:に) (v:紹介し) 昼の気力と夜の活力 (o:を) 取り戻す啓示の書男 (o:を) 鍛え充実の人生 (o:を) 送るためのノウハウ満載

obj=手に入る材料

obj=使ったマル秘漢方処方と健康食

obj=昼の気力と夜の活力

obj=取り戻す啓示の書男

obj=鍛え充実の人生

DTS=(B8901089) 初めて学ぶ 図解ツーバイフォー工法

本書 (a:は) 梓組壁工法教材研究会 (a:が) 昭和55年度労働省主催職業訓練教材コンクール (a:で) 受賞した作品をもとに本研究会の鈴木秀三氏 (o:を) 中心とした執筆者 (a:が) 改訂増補したもので初学者 (a:に) 必要とされるツーバイフォー工法の構造材料施工 (a:に) 関する知識 (o:を) 系統的 (a:に) (v:解説した) もので (x:ある)

obj=受賞した作品をもとに本研究会の鈴木秀三氏

obj=関する知識

DTS=(B8901121) 図解 建築法規事典 1989年版

新しい建築基準法とその関連法規 (o:を) 敷地道路建築物の高さ構造設備工作物 (a:という) 建築の実務 (a:に) 即したユニークな分類法により再整理し最新の資料とわかりやすい図版 (o:を) 豊富 (a:に) 使って具体的 (a:に) (v:解説した) 法規解説書の決定版

obj=新しい建築基準法とその関連法規

obj=即したユニークな分類法により再整理し最新の資料とわかりやすい図版

DTS=(B8901179) カロリー計算なしで確実にやせるコツ

この本ではこの難問 (o:を) 解決するための画期的な方法を (v:紹介する) ことにしました

obj=この本ではこの難問

DTS=(B8901223) 機械材料学

本シリーズ (a:は) 大学短大高専における機械系諸学科の学生のために基礎的必修的な学科目の知識と技術 (o:を) 体系的 (a:に) 習得してもらうことを目的として図表をできるだけ豊富 (a:に) 盛り込み基礎的内容 (o:を) 簡潔平易 (a:に) (v:解説する) 編集方針 (a:で) 企画されている

obj=大学短大高専における機械系諸学科の学生のために基礎的必修的な学科目の知識と技術

obj=盛り込み基礎的内容

DTS=(B8901250) 中枢神経疾患の画像診断

本書 (a:は) 中枢神経の画像診断の初心者これらの多岐にわたる画像診断 (o:を) 各検査法ごとにその特徴をわかりやすく解説し検査法をどのように選択し組み立てすすめるかを (v:解説した) 入門書 (a:で) (x:ある)

obj=中枢神経の画像診断の初心者これらの多岐にわたる画像診断

付録 D 処理プログラムの概要

```
#!/bin/sh
#      DPC ダイナミック・シソーラス評価用システム

VSEL=vdb.sed
VERB=verb.sed
OBJC=obj.sed
ETCS=aux.sed

case $# in
1) ;;
*)      echo 1>&2 "usage: proc file"
        exit 1;;
esac

cat $1 |
prepro |
sed 's/[、-__○--□]//g' |
sed -f $VSEL |
sed -f $VERB |
tr '\\ ' '\012' | tee /tmp/p1 |
egrep '^DTS=|(v:' |
sed -f $OBJC |
sed -f $ETCS | tee /tmp/p2 |
awk -F')' '
    '/^DTS=/ {printf("\n");}
    {print;
     for(i=1; i<=NF; i++) {
       if(split($i,a,"(")>=2)
         if(jsubstr(a[2],1,2)=="o:")
           printf("\tobj=%s\n",a[1]);
     }}' |
awk '
    '{if(substr($0,1,4)=="DTS=") dts=$0;
     else {if($0!="" && dts!="") {print dts; dts="";}
     print }}' | uniq
```

V 結論

1 ダイナミックシソーラス機構

第IV章1の「文意解析処理方式」手順により、データベース処理手順(IV.2～IV.2.2.1)としては下図(図V.1)の①②③④通り、質問文処理手順(IV.2.3～IV.2.4.2)としても下図の①②③④通り進め、検索処理手順として④の照合(IV.2.5.1 方式1: 質問文、データベースとも「文意解析処理」し、同義語辞書で語彙拡充してキーワード作成)が高い適合性能(質問文とデータベースから自動的に重要意味語句を抽出し、ノイズを排除)を発揮することを確認した。またこれらの処理手順が原理的に自動実行できることから、当研究で目的としたダイナミック・シソーラスの基本的作動機構を成立させることが出来た。

2 検索システムとしてのダイナミック・シソーラス

2.1 検索手順の確認と評価(図V.2を参照)

①(文意解析処理)で質問文の主題用語を的確に自動捕捉できるためには、質問者の願望・質疑を表現した用言辞書の拡充および質問者の願望・質疑に該当する用言が無い体言留め文体(箇条書きや表題文調)の採集など質問表現捕捉性能(網羅・適合性)の強化が必要である。当実験ではこれが不十分なため、②1(主題向体言の自動抽出)で既に②1b(願望・質疑の用言辞書の不全による漏れ)が生じている。

②(重要文・語句の抽出)での処理としては②1aだけでなく、本来は②2(用言動態の分類による体言の自動仕分け)が更に重要であり、

言葉の意味を電算システムとして把握する手順の中核である。用言動態の分類とは石川方式で「単位構文処理(動詞の意味に注目した文意解析処理)」と称し、いわゆる日本語文法で「コトの類型」をさし、動詞を相(Aspect)、時制(Tense)、態(Modality, Voice)、法(Mood)、活用(Conjugation)および結合し得る助詞で分類し、連結する名詞や文の意味を限定・特定する方法として構築された記号体系である。だが今回の研究ではこの一部を②1として活

3 データベースとしてのダイナミック・シソーラス

用できる可能性の確認に留めたため、②1で抽出された文・句に含まれる自立語は無作為に検索用語として採用するので、②2b(ノイズ)を検索用語編成段階で発生させている。よって投入される検索用語③には③2(ノイズ)が既に混入しており、②1bが原因で③3(漏れ)も出ている。従ってこれらの検索用語を統制用語辞書で拡充するとノイズも少しばかり増幅される。とは言え、原則として、当検索方法により自然語質問文によるノイズの少ないデータベース検索が可能であることを確認した。

2.2 今後の課題

- 1) 主題向き体言をより網羅的に抽出できるようにするために、願望・質疑を表わす用言の辞書を充実させる。
- 2) 抽出された用言を更に解析して、ノイズを発生する語句を除去し、また接続詞や接続助詞から論理演算が必要な語句の組み合わせも判定できるようにする必要がある。
- 3) 文意解析処理されてS式に展開された質問文を願望・質疑・意図の部分、主題部分、および不要部分に分解したリストとして電子ファイルしておき、願望・質疑の用言辞書の更新と主題部分の今後の組織化・構造化研究(用言シソーラスや項目自動付与など)に備える。

3 データベースとしてのダイナミック・シソーラス

3.1 ダイナミック・シソーラス搭載データベース構築の構想

基本的に、当データベース構築の手順は検索質問処理と同じで、処理対象が質問文から記事の文章に代わり、捕捉対象用言が「筆者の意図・見解・願望・抱負など」となる。

だがここでは用言系のシソーラスが本格構築され、従来の体言系シソーラスと連動したシステムとの自然語による自在な対話と正確なデータベース検索が可能となる。(図 V.3 を参照)

当然、自動索引付与が作動するし、更には自動文意把握によるシステム上の論理化・組織化・構造化が発達するので用言系のシソーラスの充実に付随して体言系のシソーラスの自動更新も可能となろう。

4 トータルシステムとしてのダイナミック・シソーラス

図 V.3 を参照の通り、検索側もデータベース側も本当の意図が均質に照合し合い、認知し合うので、データベース制作作業と制作内容が画期的に変貌し、また検索と言うよりもシステムとの気楽な対話が実現し、従来のデータベース構築作業と検索システム運用上の諸問題は払拭されることが出来よう。

運転と更新も双方から随時自動的に行なわれてダイナミック・シソーラス・エンジンが成立し、例外的な処理不能リストだけを専門家が検討して通常処理へ組み入れていくこととなるろう。

5 今後の展望と抱負

上記の見通しを持つダイナミック・シソーラスは人間の記憶・思考・創作活動をその代読・記憶・整理の面で支援・誘導し、人間の知的活動を保全・強化するマインドツールの基本技術ともなれる可能性がある。

よって当研究を筆者の意図と質問者の願望の用言の実用的整備、及び用言シソーラス試作の面から更に深め、この構想を更に現実化したい。

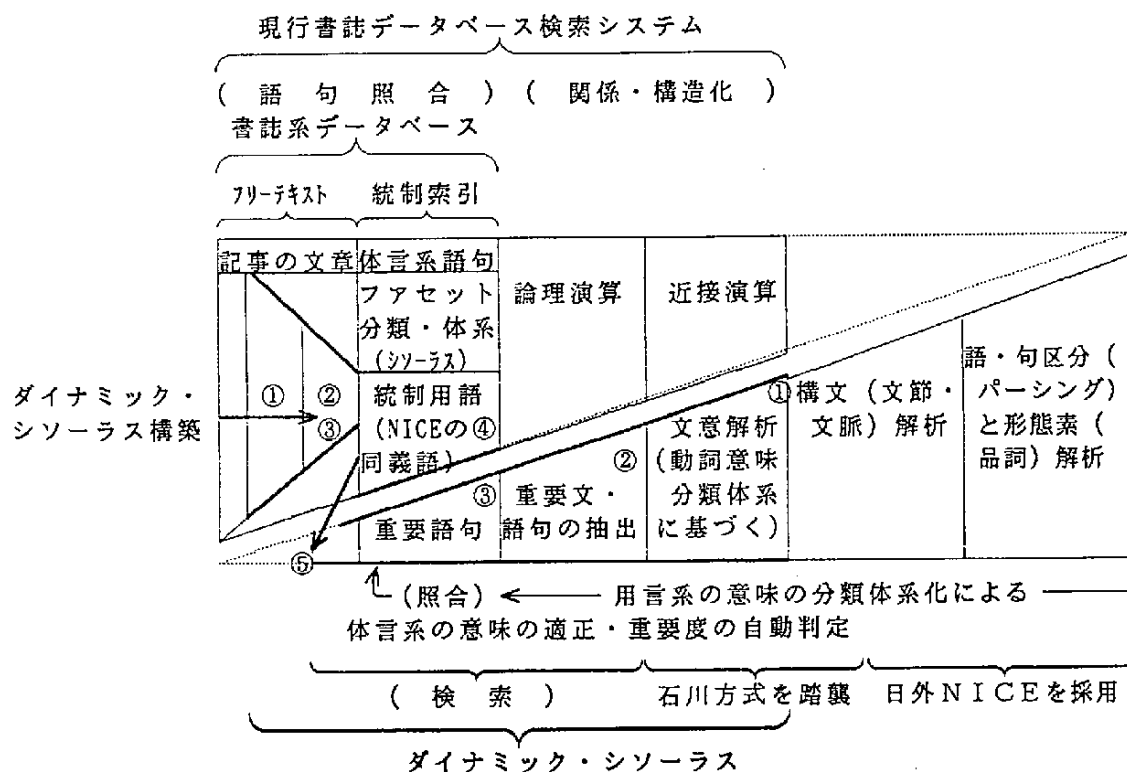


図 V.1: ダイナミック・シソーラスの機構

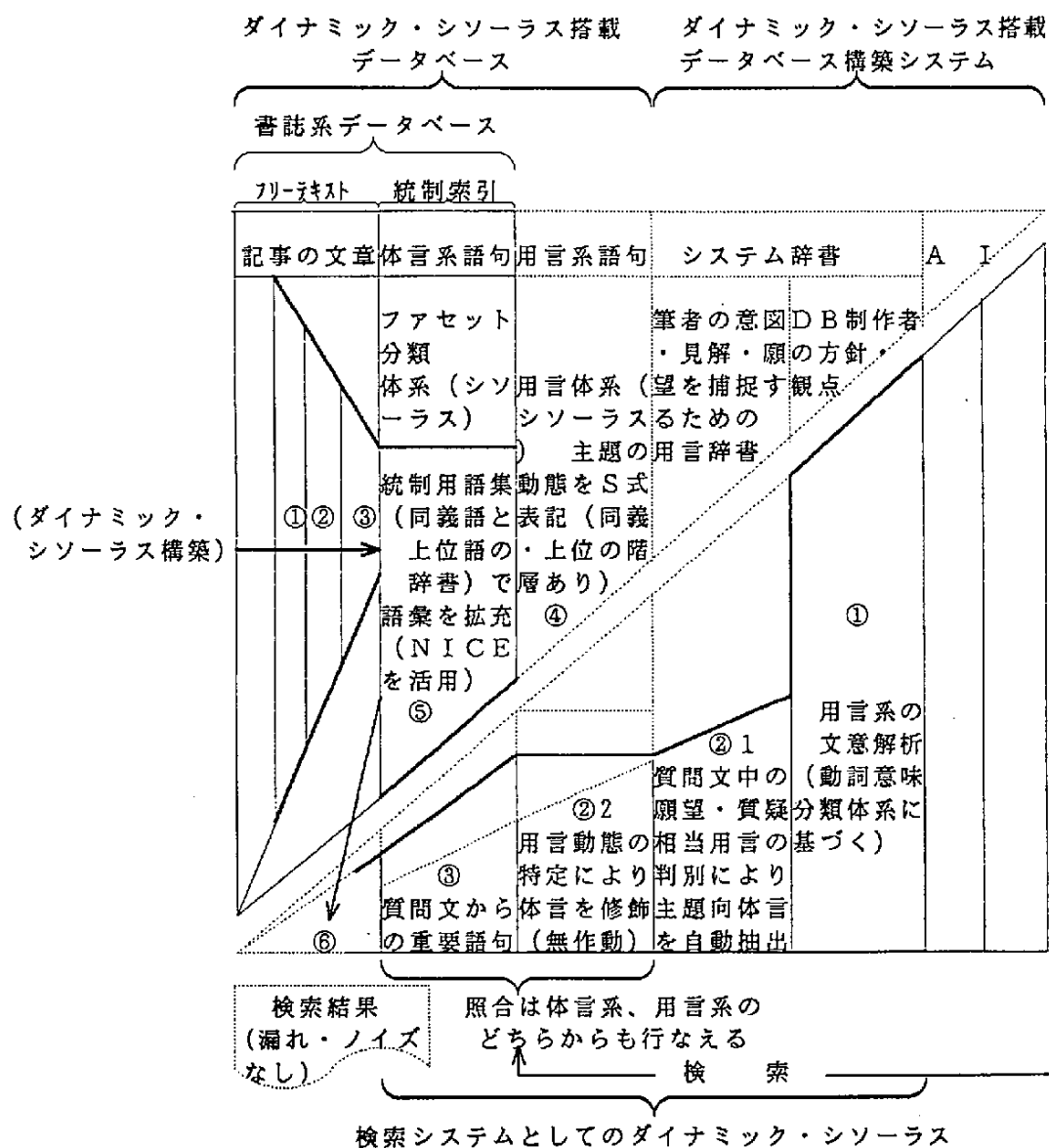


図 V.3: データベースとしてのダイナミック・シソーラスの概念図

参考文献

- [1] 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.
- [2] 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).
- [3] 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).
- [4] Martinez, C., et al: An Expert System for Machine-Aided Indexing, J. Chem. Infor. Comput. Sci., Vol.27, No.4, pp.158～162(1987).
- [5] Humphrey, S. W. : A Knowledge-Based Expert System for Computer-Assisted Indexing, IEEE Expert, Vol.4, No.3, pp.25～38(1989).
- [6] Cleveland, D. B. and Cleveland, A. D. : Introduction to Indexing and Abstracting, p.209, Libraries Unlimited, Inc., Colorad(1983).
- [7] Craven, T. C. : String Indexing, p.246, Academic Press, Inc., London(1986).
- [8] Foskett, A. C. : The Subject Approach to Information, 4th ed., p.574, Clive Bingley, London(1982)
- [9] Dym, E. D., ed. : Subject and Information Analysis, p.498, Marcel Dekker, Inc., NY(1985).
- [10] Stevens, M. E. : Automatic Indexing : A State-of-the-Art Report, p.290, National Bureau of Standards, Wachington(1965).
- [11] 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).

- [12] 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).
- [13] Chan, L. M., et al ed. : Theory of Subject Analysis, A Sourcebook, p.415, Libraries Unlimited, Inc., Colorado(1985).
- [14] Luhn, H. P. : A Stastical Approach to Mechanized Encoding and Searching of Literary Information, IBM J., Oct., pp.309-317(1957).
- [15] 二村 一、松尾文：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).
- [16] Stailes, H. E. : The Association Factor in Information Retrieval, J. ACM., Vol.8, No.2, pp.271～279(1961).
- [17] 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、情報科学技術研究集会発表論文集、Vol.22, pp.37～45(1986).
- [18] 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).
- [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] 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).
- [22] 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).
- [23] 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).
- [24] 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.64pp.1～8(1987).
- [25] Salton, G. and McGill, M. J. : Introduction to Modern Information Retrieval, p.448, McGraw Hill International Book Co., Auckland(1983).

- [26] 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ェア,
オンライン検索、Vol.10, No.3, pp.117～126(1989).

————— 禁無断転載 —————

平成 3 年 3 月 発行

発 行 財団法人 データベース振興センター
東京都港区浜松町二丁目 4 番 1 号
世界貿易センタービル 7 階
TEL. 03-3459-8581

依 託 先 株式会社 紀伊國屋書店
東京都世田谷区桜ヶ丘五丁目 38 番 1 号
TEL. 03-3439-0123

印 刷 所 有限会社 日製工業写真社
東京都大田区大森北一丁目 23 番 8 号

