

TR-IT-0174

類似検索を用いた情報検索システム
Information Retrieval System using Semantic Distance Calculation

美馬 秀樹 隅田英一郎 飯田 仁、
Hideki MIMA Eiichiro SUMITA Hitoshi IIDA

1996年6月21日

概要

近年のインターネットの普及や、今後期待されている電子図書館の実現等により、膨大な量の全文データベースが利用可能となるため、必要な情報を効率的に探し出す情報検索技術が望まれている。この要求に対し、用例検索を目的とした類似検索技術は一つの可能性を示すと考えられる。類似検索を文書データ等に対する情報検索に応用することにより、入力文と意味的に類似した文を含むテキストを効率的に検索できる。さらに、類似検索には従来技術と比べ次の特徴がある。1) ユーザは検索キーとして任意のフレーズや文をそのまま入力することができる。2) 意味距離計算により、入力文と検索対象との表層のばらつきをシソーラスの利用等により自動的に吸収できるため、ユーザは基本的に検索キーに使う語彙や活用等のバリエーションを工夫する必要がない。3) 意味距離によりマッチングの度合いを適宜調整しながら検索できる。本報告では、このような類似検索技術を用いた情報検索システムの実現方法について述べる。

ATR 音声翻訳通信研究所
ATR Interpreting Telecommunications Research Laboratories

©(株) ATR 音声翻訳通信研究所 1996
©1996 by ATR Interpreting Telecommunications Research Laboratories

目次

1	まえがき	1
2	類似検索を用いた情報検索の概要	3
2.1	用例を使ったインデックスの記述	3
2.2	意味距離計算	4
2.3	用例に対する検索対象文の決定処理	5
2.4	意味距離の総和による検索対象文の決定	6
2.5	意味距離の視覚化による目的情報の選択	6
3	類似検索システムの構成	8
3.1	類似検索システムの稼働環境	9
3.2	システム設計	10
3.2.1	モジュール構成	10
3.2.2	プロセス間ソケット通信	11
3.2.3	ソケット通信のプロトコル	13
3.3	利用方法	16
3.3.1	類似検索システム起動のためのシェルスクリプトの設定	16
3.3.2	類似検索システムのプロセスを削除するためのシェルスクリプトの設定	16
3.3.3	類似検索システムの稼働環境の設定	17
3.4	ユーザインタフェース	18
3.4.1	システムの起動方法	18
3.4.2	新聞記事ビューアー	19
3.4.3	検索入力	21
3.4.4	削除リスト	23
3.4.5	フレーズレベルの用例空間（意味距離の渦巻配列による視覚化）	24

3.4.6	フレーズレベルの用例空間（用例の各要素に対応する 2 次元意味距離視覚化）	26
3.4.7	検索要求に対するトータル用例空間	27
4	検索実験と実験結果に対する考察	29
4.1	プロトタイプシステムによる検索実験	29
4.2	実験結果に対する考察	30
4.2.1	検索結果	30
4.2.2	検索速度	31
5	まとめと今後の課題	32
A	類似検索システム実行ファイル及びツール類についての解説	34
A.1	類似検索システム実行ファイル	34
A.2	データファイル構築ツール	35
A.3	用例データのコーパスからの抽出	36
B	類似検索システムの操作例	39
B.1	入力文のサンプルと検索例	39
C	トラブル・シューティング	46
C.1	システム起動時の異常	46

表一覧

3.1	類似検索システムの稼働環境	10
3.2	類似検索処理部の稼働環境	10
4.1	検索実験に用いたターゲット・データベース（新聞記事8ヶ月分）の詳細	29
4.2	記事のインデックスとして用いた文（検索対象文）の詳細	30
4.3	抽出した用例の各言語表現別の内訳	30

図一覽

2.1	類似検索システム	7
3.1	類似検索を用いた情報検索システムの構成	9
3.2	類似検索システムのモジュール構成	11
3.3	類似検索システム起動時のソケット通信手順	11
3.4	類似用例検索時のソケット通信手順	12
3.5	類似検索システム終了時のソケット通信手順	12
3.6	類似検索システム起動時のソケット通信プロトコル	13
3.7	類似用例検索時のソケット通信プロトコル (意味距離の渦巻配列による視覚化)	14
3.8	類似用例検索時のソケット通信プロトコル (用例の各要素に対応する 2 次元意味距離視覚化)	15
3.9	類似検索システム終了時のソケット通信プロトコル	15
3.10	新聞記事閲覧ウィンドウ	20
3.11	検索入力ウインドウ	21
3.12	削除リストウインドウ	23
3.13	意味距離の渦巻配列による視覚化ウィンドウ	24
3.14	用例の各要素に対応する 2 次元意味距離視覚化ウィンドウ	26
3.15	トータル用例空間	27
A.1	corpus.head, article.head の作成	37
A.2	用例データのコーパスからの抽出	38
B.1	政府が提唱する	40
B.2	政府が提唱する, 空港の場所	40
B.3	検索された記事例 (政府が提唱する, 空港の場所)	41
B.4	政府が提唱する, 売上税に反対する	42
B.5	検索された記事例 (政府が提唱する, 売上税に反対する)	43

B.6	マスコミの問題	44
B.7	マスコミの問題, 問題が生じる	44
B.8	検索された記事例 (マスコミの問題, 問題が生じる)	45

第 1 章

まえがき

近年のインターネットの普及や、今後期待されている電子図書館の実現等により、膨大な量の全文データベースが利用可能となるため、必要な情報を効率的に探し出す情報検索技術が望まれている。データベースに対する検索では、大きく分けて、限定的かつ詳細な情報を要求する特定検索と、「情報散策」[有田 95]といった明確な目的の薄い大まかな情報集合への検索を要求する広域検索[武田 95]が考えられるが、特定検索においては基本的に全データを検索する必要があるため、検索対象が膨大になった場合、検索コストが極めて大きくなる可能性がある。

このような特定検索においては、従来より、検索キーやカテゴリによる検索指定と AND/OR 等の関係式よりなる検索式を用いるもの、自然言語による検索要求から検索キー等を抽出し検索式を作成するもの[加納 91]、さらに、それら検索キーに対し、シソーラスを用いた拡張検索[Pai 91]等が用いられてきた。しかし、検索式の作成は初心者には非常に困難である。また、自然言語による要求は、入力文よりユーザの検索要求を正確に抽出するのが難しい。さらに、シソーラスを用いた拡張検索では、検索のもれをある程度吸収できるが、検索キーとシソーラスのみの組み合わせでは、使用されている文脈における検索キーの意味が考慮できないなどの理由により、意図しない文脈におけるノイズ等を多く拾ってしまう可能性がある。

このような問題に対し、用例検索を目的とした類似検索技術[美馬 96]は、一つの解決の可能性を示すと考えられる。類似検索を文書データ等に対する情報検索に応用することにより、入力文と意味的に類似した文を含むテキストを効率的に検索できる。さらに、類似検索には従来技術と比べ次の特徴がある。1) ユーザは検索キーとして任意のフレーズや文をそのまま入力することができる。2) 意味距離計算により、入力文と検索対象との表層のばらつきをシソーラスの利用等により自動的に吸収できるため、ユーザは基本的に検索キーに使う語彙や活用等のバリエーションを工夫する必要がない。3) 意味距離によりマッチングの度合いを適宜調整しながら検索できる。

本稿では、このような類似検索技術を用いた情報検索システムの実現方法について述べる。ま

た、入力文と意味的に近い検索候補において、ユーザの意図する情報へ効率的に絞り込むため、意味距離計算により得られた意味距離を視覚化し、任意の意味距離にある情報を効果的に提示する枠組みを提案する。本手法では、検索入力文と当該用例を含むテキストとの意味距離を段階的に色分けし、ユーザに視覚的に提示する。ユーザはこの視覚化されたスペースの任意の位置を指し示すことにより、対応する意味的近さにあるテキストに動的にアクセス可能となる。本手法に基づく新聞記事の検索に対するプロトタイプシステムを作成し、実験によりその有効性を確かめた。

以下、第2章で、類似検索を用いた情報検索の概要、第3章で、類似検索を用いた情報検索システムの構成、第4章で、プロトタイプシステムによる検索実験と、実験に対する考察、第5章で、まとめと今後の課題について述べる。

第 2 章

類似検索を用いた情報検索の概要

類似検索に基づく枠組みで実現される情報検索では、入力語句と用例との間の意味距離計算という一貫した方法により、入力文に意味的に類似した用例を含む文（以下、検索対象文とする）を検索する。そして、これらの検索対象文に関連付けられた情報（雑誌や新聞記事等の文書データ、写真や音のマルチメディアデータ等；以下、目的情報とする）を表示することで、入力文に意味的に関連した情報をユーザに提供できる。また、意味距離計算は単純であり、高速な検索処理を実現できる。本章では、このような類似検索を用いた情報検索の概要を述べる。

2.1 用例を使ったインデックスの記述

類似検索におけるインデックスは、検索対象文の用例¹を意味的にまとまった単位で抽出したものである。用例に対する言語表現は、検索の鍵となる部分を持ち、この部分を具体化する語句を基に検索される検索対象文と、検索対象文に関連される目的情報が決定される。そこで、検索対象文の用例と、目的情報を示す検索レコード情報の組を、あらかじめ検索対象文となる実際のテキストデータから抽出し、検索の際のインデックス情報として次のように記述する。

用例の言語表現 =>
(E1, [検索レコード 1]),
:
(En, [検索レコード n])

この記述では、用例が E_i である時、検索レコード i に記述された情報が検索対象文や、関連す

¹本稿では、用例という言葉を用いて「ある言語表現の検索の鍵となる部分についての具体例」という意味で使用する。また、説明を簡潔にするために、用例の記述を省略することがある。

る目的情報であることを示す。尚、 E_i は語句の組である。検索処理において、入力表現の語句と N - ベスト (又は、意味距離のある閾値以下) でマッチする用例を意味距離計算により求め、用例が含まれる検索対象文と、関連される目的情報を検索結果として選択する。つまり、入力表現の語句に最も意味的に近い用例が E_i であれば、検索レコード i の記述を原言語表現に最も関連の大きい目的情報とみなすわけである。

2.2 意味距離計算

入力文の語句と用例との間の意味的な近さを求める方法として、本検索システムでは、隅田らの意味距離計算を採用している。この方法では、まず、シソーラス (類語辞典 [大野 84] に準拠) の概念階層における意味概念間の位置関係によって、入力文の単語 i と用例の単語 e の間の意味距離 $d(i, e)$ を計算する。意味距離は、0 から 1 までを値域とし、0 に近いほど i と e は意味的に類似していることを示す。意味距離はシソーラスの与え方によって値が変わるが、プロトタイプシステムで現在使用しているシソーラスに基づいた意味距離の例を示す。

$$d(\text{関西新空港}, \text{空港}) = 0.00.$$

$$d(\text{提唱}, \text{諮問}) = 0.33333.$$

$$d(\text{経済}, \text{美馬}) = 1.00.$$

シソーラス中の単語と exact-match しない複合語については、複合語の主とする意味が終端方向に表れる頻度が多いことを考慮して、単語間の左方向最長一致による partial-match を採用している。また、動詞等の活用語については、シソーラスとの整合を取るため終止形に変換してマッチングを行っている。

用例の言語表現に対する検索の鍵となる部分についての入力文と用例の表現を、それぞれ、 I と E とする。 I と E の間の意味距離は、 I と E を構成する単語間の意味距離を基にして計算する。 I および E が、次のように t 個の語句の組として構成されているとする。

$$I = (i_1, \dots, i_t)$$

$$E = (e_1, \dots, e_t)$$

I と E の間の意味距離を次のように計算する。

$$\begin{aligned} d(I, E) &= d((i_1, \dots, i_t), (e_1, \dots, e_t)) \\ &= \sum_{k=1}^t d(i_k, e_k) \cdot w_k \end{aligned}$$

w_k は、検索における k 番目の要素の重みを示し、0 から 1 までを値域とする²。

² w_k の値は単語の頻度情報に基づいて計算する等が考えられる。ただし、プロトタイプシステムでは、古瀬らの TDMT [古瀬 94] 同様、 w_k の値を一律に $1/t$ としている。

2.3 用例に対する検索対象文の決定処理

本システムでは検索対象文に現れる意味的役割を担う表現のパターンにより用例を分類してインデックス化し、検索入力のパターンに応じて、対応する用例との意味距離計算により検索対象文を絞り込む。

検索対象文に対し、文法属性を表現しない X のような記号（以下、可変部と呼ぶ）と表層語句とにより用例の言語表現を表わす。可変部は検索の鍵となる部分であり、可変部を具体化する語句により用例を記述する。例えば、次の用例は、検索対象文“運輸省が関西新空港の位置を諮問した”より得られた「 X が Y する」、「 X の Y 」のパターンを含む用例記述の例である。

X が Y する \Rightarrow
((運輸省, 諮問) [文番号 128]),
:
 X の Y \Rightarrow
((関西新空港, 位置) [文番号 128]),
:

(運輸省, 諮問) では、「 X の Y 」に対し、 X = 「運輸省」、 Y = 「諮問」による「運輸省が諮問する」という用例を表わす。可変部 X , Y を具体化する入力文の語句の組を I とする。例えば、入力文が「政府が提唱する」の場合、 I は (政府_[ガ], 提唱_[スル]) となる。

w_k の値を一律に 0.5 とすると、 I と用例 (運輸省, 諮問) の間の意味距離は次のようになる。

$$\begin{aligned} & d((\text{政府}_{[ガ]}, \text{提唱}_{[スル]}), (\text{運輸省}_{[ガ]}, \text{諮問}_{[スル]})) \\ &= d(\text{政府}, \text{運輸省}) \cdot w_1 + d(\text{提唱}, \text{諮問}) \cdot w_2 \\ &= 0.00002 \times 0.5 + 0.33333 \times 0.5 \\ &= 0.16667 \end{aligned}$$

また、 I と他の用例との距離として以下のようなものも考えられる。

$$\begin{aligned} & d((\text{政府}_{[ガ]}, \text{提唱}_{[スル]}), (\text{政府}_{[ガ]}, \text{提案}_{[スル]})) = 0.16. \\ & d((\text{政府}_{[ガ]}, \text{提唱}_{[スル]}), (\text{県議会}_{[ガ]}, \text{聞く})) = 0.29. \\ & : \\ & d((\text{政府}_{[ガ]}, \text{提唱}_{[スル]}), (\text{町長}_{[ガ]}, \text{話す})) = 0.67. \\ & d((\text{政府}_{[ガ]}, \text{提唱}_{[スル]}), (\text{企業}_{[ガ]}, \text{融資}_{[スル]})) = 1.00. \end{aligned}$$

以上のような計算により (運輸省_[ガ], 諮問_[スル]) を用例として持つ “運輸省が関西新空港の位置を諮問した” 等が意味的に近い検索対象文として検索され、入力「政府が提唱する」の検索結果として、関連づけられている新聞記事や画像データを得ることができる。

2.4 意味距離の総和による検索対象文の決定

入力文の言語表現と検索対象文に対する用例とのマッチングに複数の組み合わせ方が存在する場合、各検索対象文に対するトータルな意味距離に何らかの基準が必要である。例えば、「政府が新空港の場所を提唱する」は原言語表現「XのY」(下線)と「XがYする」(二重下線)の二つに対する用例とマッチするが、

- (1) “運輸省が関西新空港の位置を諮問した。”
- (2) “政府が提案している売上税の・・・・・・”
- (3) “現場はリーム空港の近くで・・・・・・”

のような検索対象文に対しての適切な意味的近さを定義しなければならない。このような場合、

- a) それぞれのパターンに対する意味距離の最小のものを検索対象文とのトータルな意味距離とする。
- b) 各検索対象文において、入力文との照合する用例が多いほどトータルの意味距離を小さくする。
- c) 文中での用例の現れる位置や、用例の出現する頻度などにより用例との意味距離に重みを設定する。

等の処理が考えられる。プロトタイプシステムでは、a)とb)に対して選択可能としている。

2.5 意味距離の視覚化による目的情報の選択

実際の新聞等のデータには様々な類義語や言い回しが存在するうえ、ユーザによる解釈の違い等により必ずしも意味距離の最小のものがユーザの意図する情報とは限らない。したがって、ユーザには目的情報がリアルタイムに提示され、かつ対話的に選択できることが望ましい。また、検索候補が大量になった場合にも、ユーザとの対話等により選択操作を適切に支援する必要がある。従来の検索システムでは、得られた候補の見出しやサマリーを列挙するものが中心であり、大量の候補に対してユーザへの適切な支援環境とは言い難い。このような問題に対し、情報の視覚化 (information visualization) が重要視されている [武田 95][有田 95]。本システムでは、検索手法として意味距離計算を用いるが、意味距離計算の特徴として、意味距離により入力文と検索対象文との意味的近さの度合いが得られる。よって、意味距離を視覚化することにより、意味

的な近さを基準としたユーザへの対話的な選択支援が行える。本報告では、このような意味距離を効果的に視覚化し、任意の意味距離にある情報をリアルタイムに提示する枠組みを提案する。本手法では、入力文と検索対象文との意味距離を用例空間として段階的に色分けし(図 2.1)、ユーザに視覚的に提示する。ユーザはこの色分けしたスペースの任意の位置を指し示すことで、対応する意味的近さにある目的情報に動的にアクセス可能となる。

文パターン

- ～が～
- ～が～する
- ～が～される
- ～は～
- ～は～する

入力

政府 が

提唱 する

用例空間

「アレイン」 「ソート」 ■ 濁

検索文 政府が提唱する

められ、40年代の初頭、関西に新しい空港をつくる動きが本格化していったようである。当時、成田空港建設は予定どおりには進んでいなかったらしい。

運輸省がこの審議会に、関西新空港の場所を諮問したのは16年秋のことであった。

大阪・泉州沖、神戸沖、淡路島などが候補に登ってきたようであったが、淡路島などの大阪から遠過ぎる場所には反対意見が出ていたようであった。神戸沖については、飛行機が低空で飛ぶと神戸港の運営に差しつかえるとの懸念もあったらしい。

入力

政府が提唱する

用例

運輸省が(諮問した)

距離

0.166672

同じく(C)

図 2.1: 類似検索システム

第 3 章

類似検索システムの構成

新聞記事の検索を対象とした本プロトタイプシステムを Sparc Station 10 上に実装した。本章では本システムの稼働環境、システム構成、使用方法について説明する。

図 3.1 に、本検索システムの構成を示す。

- 1) 言語処理部では、検索要求である入力文に対し、形態素解析、及び用例の言語表現とのパターン・マッチングにより意味距離計算の対象となるインデックスデータ（用例データの集合）を選択する。
- 2) 意味距離計算では、先に説明した計算法により、入力と検索対象文との意味距離を計算する。
- 3) 候補選択においては、本報告で提案する意味距離の視覚化による情報への動的アクセス法に基づき、ユーザとの対話的なインタフェースを用いて、検索候補よりユーザの意図する情報に絞り込む。
- 4) 検索・表示部では、3) によりユーザが指示した検索対象文に関連する情報をターゲット・データベース（プロトタイプシステムでは新聞記事を使用）より検索し、ユーザにリアルタイムに提示する。

以下に、本章の構成を示す。

3.1 節 類似検索システムの稼働環境 では、類似検索システムを利用する際のシステムの稼働環境について述べる。

3.2 節 システム設計では、本システムのモジュール構成、ソケット通信プロトコルの設計について述べる。

3.3 節 利用方法 では、類似検索システムの実行環境の設定方法について述べる。

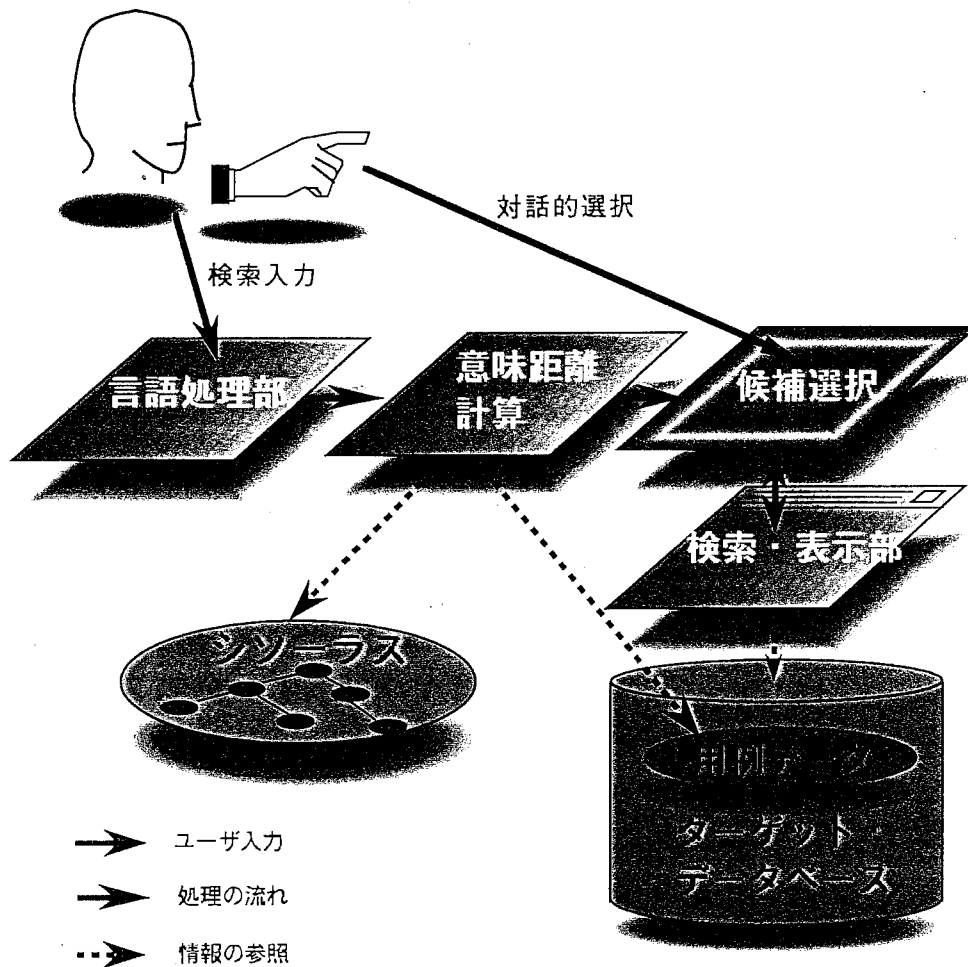


図 3.1: 類似検索を用いた情報検索システムの構成

3.4節 ユーザインタフェース では、本システムの起動方法、グラフィカル・ユーザインタフェースの使用法について説明する。

3.1 類似検索システムの稼働環境

本節では類似検索システムの稼働環境について説明する。尚、本検索システムは類似用例の検索部とユーザインタフェース部より構成される。以下では、類似用例の検索部をを類似検索処理部、ユーザインタフェース部を GUI 表示部と呼び、区別する。

表 3.1に類似検索システムの GUI 表示部の稼働環境を示す。

類似検索処理部と GUI 表示部は一体で稼働するが、それぞれ別プロセスであり、ソケットによ

表 3.1: 類似検索システムの稼働環境

マシン	Sun
開発言語	C (GUI 表示部) lisp (類似検索システム本体と GUI 表示部の間でソケット通信を行なうためのインターフェイス部)
稼働環境	Motif(IXI-Motif)
実行メモリ	約 32M バイト
ディスク容量	約 1.5M バイト

るプロセス間通信を利用するため、それぞれが任意のリモートマシン上で実行可能である。表 3.2に、類似検索処理部の稼働環境を示す。

表 3.2: 類似検索処理部の稼働環境

マシン	Sun
開発言語	Lucid Common lisp
実行メモリ	約 64M バイト
ディスク容量	約 35M バイト

3.2 システム設計

3.2.1 モジュール構成

本節では、本類似検索システムのプログラミングレベルでのモジュール構成について述べる。

図 3.2が本システムのモジュール構成である。尚、図中の () は、各モジュールの記述言語を示し、矢印はデータの流れを示している。

- システムの主な処理の流れを以下に示す。
 1. ユーザインタフェース部より入力文を読み込む
 2. 入力に対する用例を検索システム本体にソケット通信で送る
 3. ユーザインタフェース部に検索結果をソケット通信により返す
 4. 検索候補の視覚化とユーザによる選択、選択された検索結果を表示する

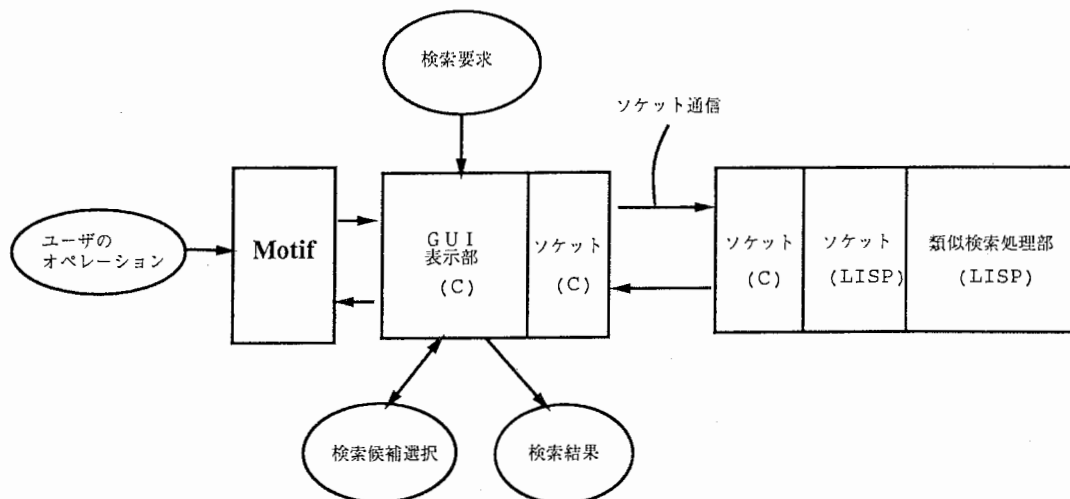


図 3.2: 類似検索システムのモジュール構成

3.2.2 プロセス間ソケット通信

GUI表示部と類似検索処理部のプロセス間ソケット通信は、システムの起動時（オプションにより選択）、及び類似用例の検索時、システムの終了時に行なう。

以下に、それらの通信手順を示す。

- システム起動時（ソケット通信により用例データを読み込むかどうかはオプションにより指定する）

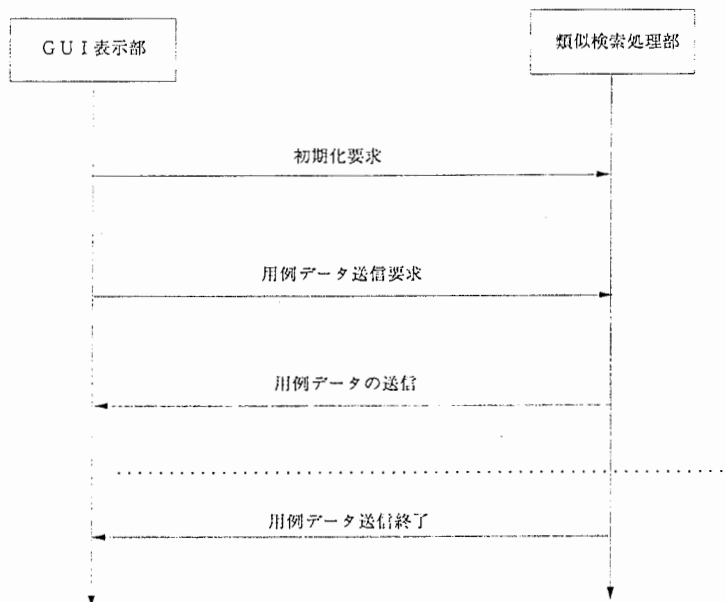


図 3.3: 類似検索システム起動時のソケット通信手順

- 類似用例の検索時

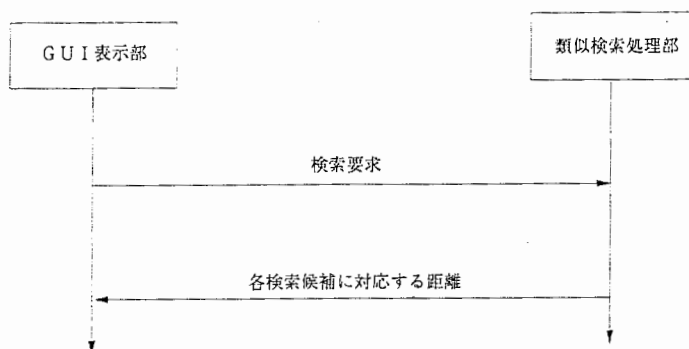


図 3.4: 類似用例検索時のソケット通信手順

- システム終了時



図 3.5: 類似検索システム終了時のソケット通信手順

3.2.3 ソケット通信のプロトコル

GUI表示部と類似検索処理部間で行なうソケット通信の各プロトコルを以下に示す。尚、図中の () 内には、送信データのバイト数を示してある。また、文字列による送信のように、バイト数が各々変化する場合は“(不定)”と表す。

- システム起動時のソケット通信プロトコル

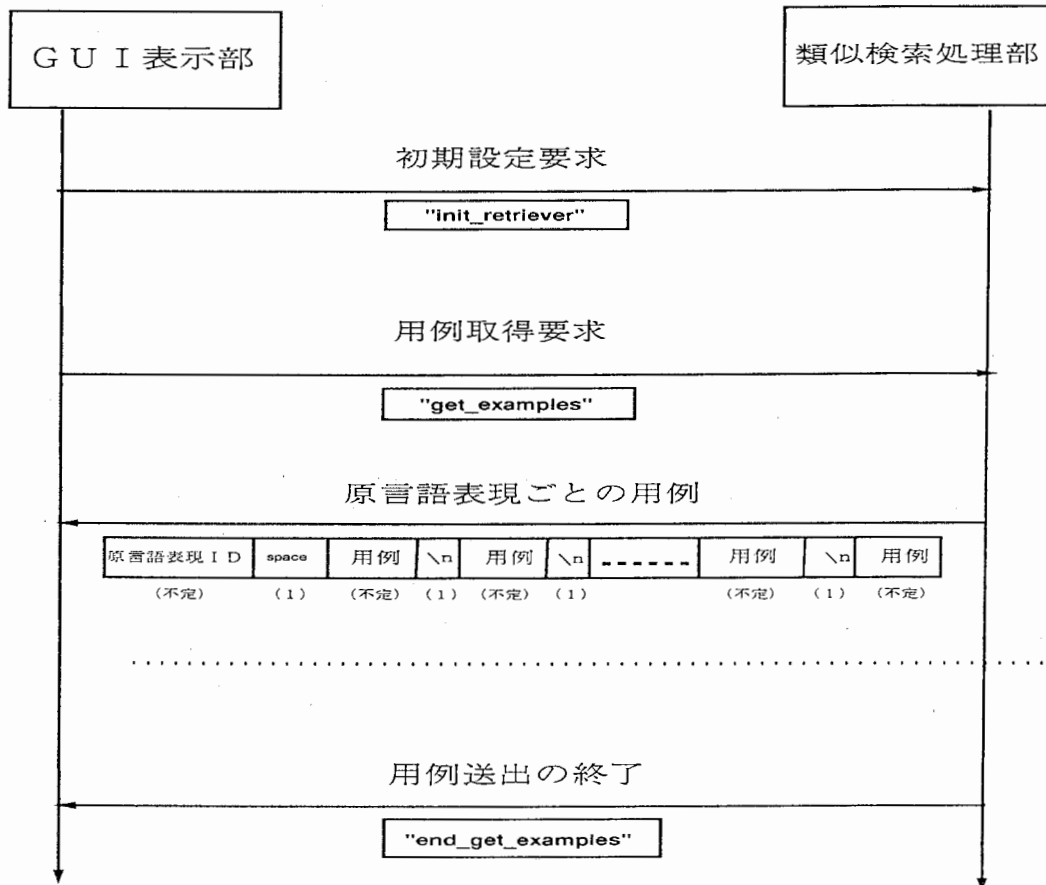


図 3.6: 類似検索システム起動時のソケット通信プロトコル

- 類似用例検索時のソケット通信プロトコル（意味距離の渦巻配列による視覚化）

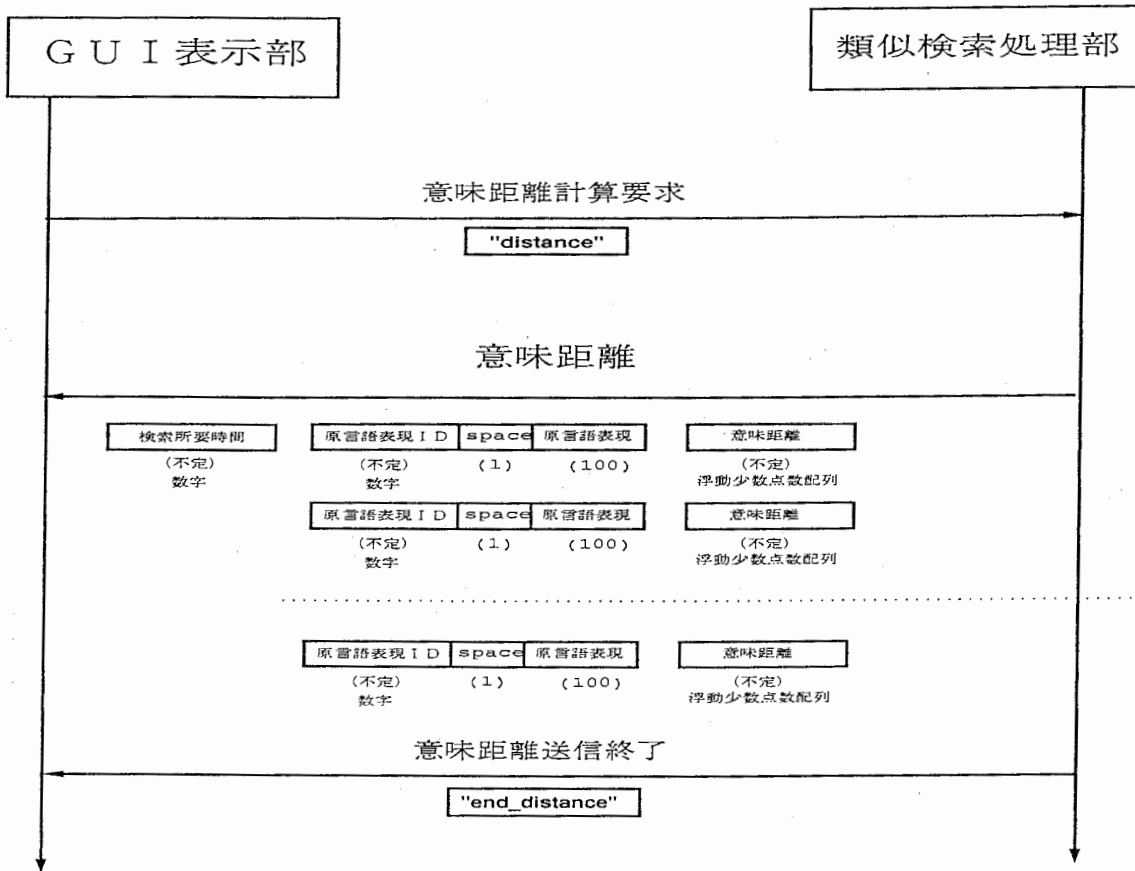


図 3.7: 類似用例検索時のソケット通信プロトコル（意味距離の渦巻配列による視覚化）

- 検索時のソケット通信プロトコル (用例の各要素に対応する 2 次元意味距離視覚化)

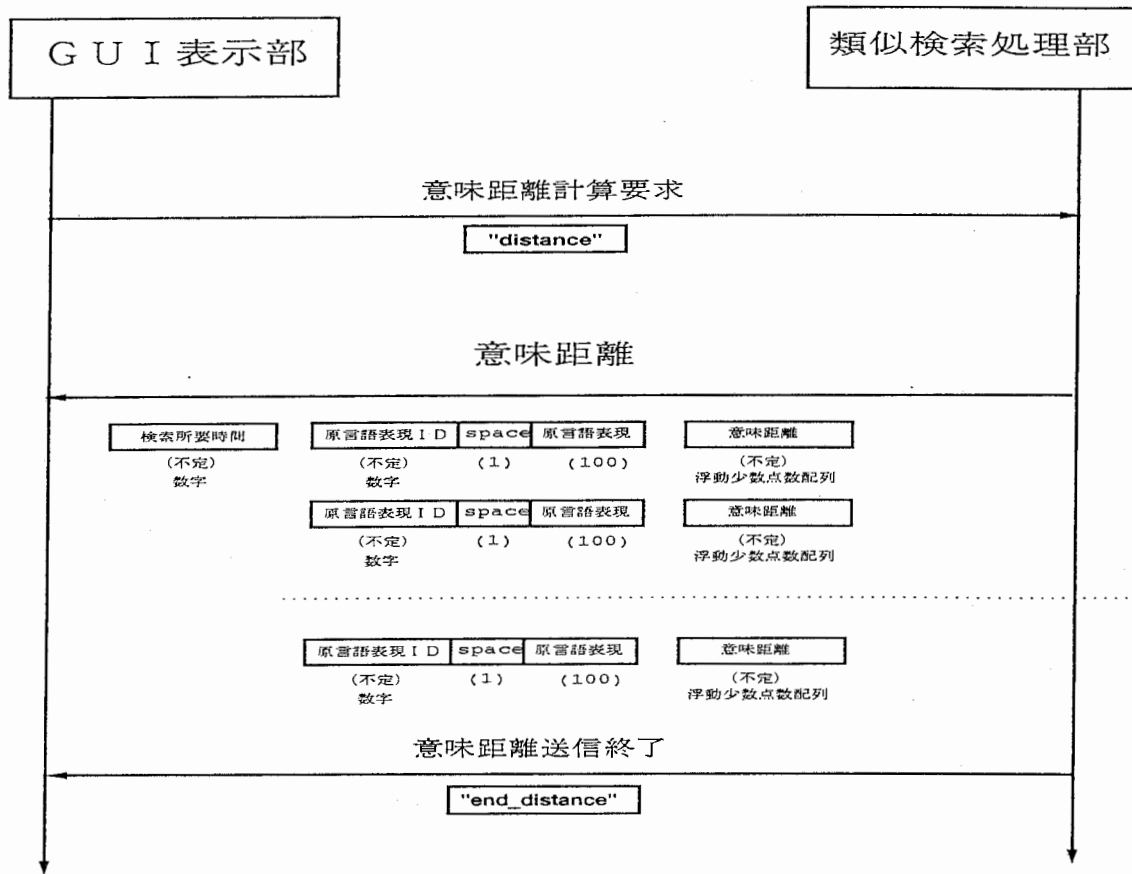


図 3.8: 類似用例検索時のソケット通信プロトコル (用例の各要素に対応する 2 次元意味距離視覚化)

- システム終了時のソケット通信プロトコル

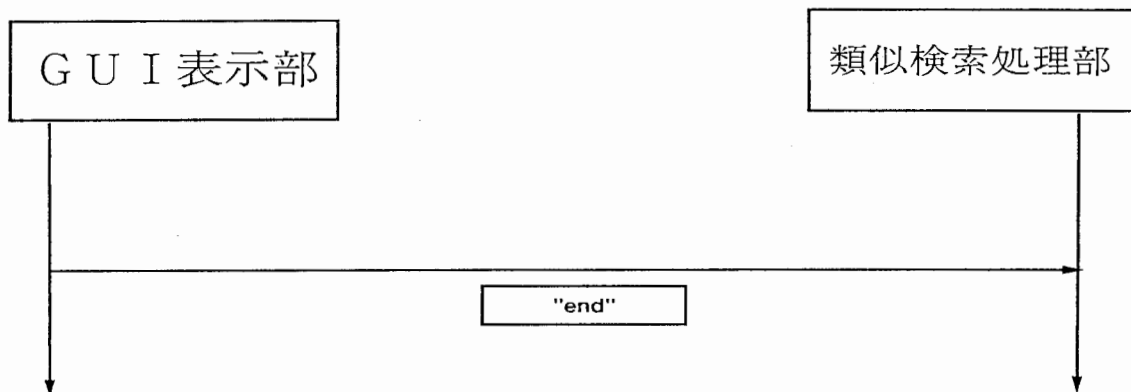


図 3.9: 類似検索システム終了時のソケット通信プロトコル

3.3 利用方法

本節では、類似検索システムの利用方法について説明する。

3.3.1 類似検索システム起動のためのシェルスクリプトの設定

1. システム起動用のシェルスクリプトファイルの

```
run.sbdr
```

をエディットして、ホスト名とパス名、ソケット名等を設定する。以下に、定義例を示す。

```
##### 表示システム (GUI 表示部) が稼働するホスト
set display_host = as31
##### 表示システム (GUI 表示部) が置かれているパス名
set display_path = /home/as31/mima/work/SBDR/DEMO
##### 本体システム (類似検索処理部) が置かれているパス名
set sbdr_path = /home/as31/mima/work/SBDR/DEMO
##### プロセス間通信で用いるソケット名
set seq_socket = SBDR-SEQ
```

3.3.2 類似検索システムのプロセスを削除するためのシェルスクリプトの設定

1. システムのプロセスを削除するために、シェルスクリプトファイル

```
kill.proc
```

に対し、ホスト名とパス名、ソケット名等を設定する。以下に、定義例を示す。

```
##### 表示システム (GUI 表示部) が稼働するホスト
set display_host = as31
##### 本体システム (類似検索処理部) が稼働するホスト名
set sbdr_host = as44
##### プロセス間通信で用いるソケット名
set seq_socket = SBDR-SEQ
```

3.3.3 類似検索システムの稼働環境の設定

- 本類似検索システムは、前述のように、グラフィカル・ユーザインタフェースを用いた GUI 表示部と類似検索処理部の 2 つのプロセスより構成され、ソケットを用いたプロセス間通信により協調的に稼働する。したがって、GUI 表示部と検索処理部間でのソケット通信に用いるソケットを環境設定ファイル

```
/etc/services
```

にあらかじめ定義しておく必要がある。以下に定義例を示す。

```
例: SBDR-SEQ 5035/tcp
```

3.4 ユーザインタフェース

3.4.1 システムの起動方法

1. デモシステムのディレクトリへ移動する

```
cd ディレクトリ  
実行例: cd ~ mima/work/SBDR/DEMO
```

2. システム起動

```
run.sbdr [英日] [用例の読み込み方法] [本体の起動方法]  
実行例: run.sbdr j file
```

(a) 英日

- e ... 英語表示
- j ... 日本語表示
- デフォルトは e

(b) 読み込み方法

- file ... 用例をファイル (付録 A 参照) から読み込む
- socket ... 用例を類似検索処理部から読み込む
- デフォルトは socket

(c) 本体システムの起動方法

- rsh ... 類似検索処理部をリモートマシン上で起動
- sh ... 類似検索処理部をカレントマシン上で起動
- デフォルトは sh

3.4.2 新聞記事ビューアー

システム起動時に、図 3.4.2 に示す新聞記事閲覧ウィンドウが表示される

1. プルダウンメニュー (ファイル (F), 編集 (E), ツール (T))
ユーザから本表示システムに処理を行なわせるためのメニューバー
 - (a) ファイル ... ファイル入出力.
 - (b) 編集 ... 記事のカット&ペースト.
 - (c) ツール ... 類似用例検索等のツール.
2. 見出しウィンドウ
検索した新聞記事の見出しが表示される.
3. 記事表示ウィンドウ
検索した新聞記事の本文が表示される.
4. 文章頭ボタン
記事の先頭にフォーカスを移動する.
5. トピックボタン
記事中の検索入力に関連する文にフォーカスを移動する.
6. 文章末ボタン
記事末にフォーカスが移動する.
7. 終了ボタン
検索システムの終了.

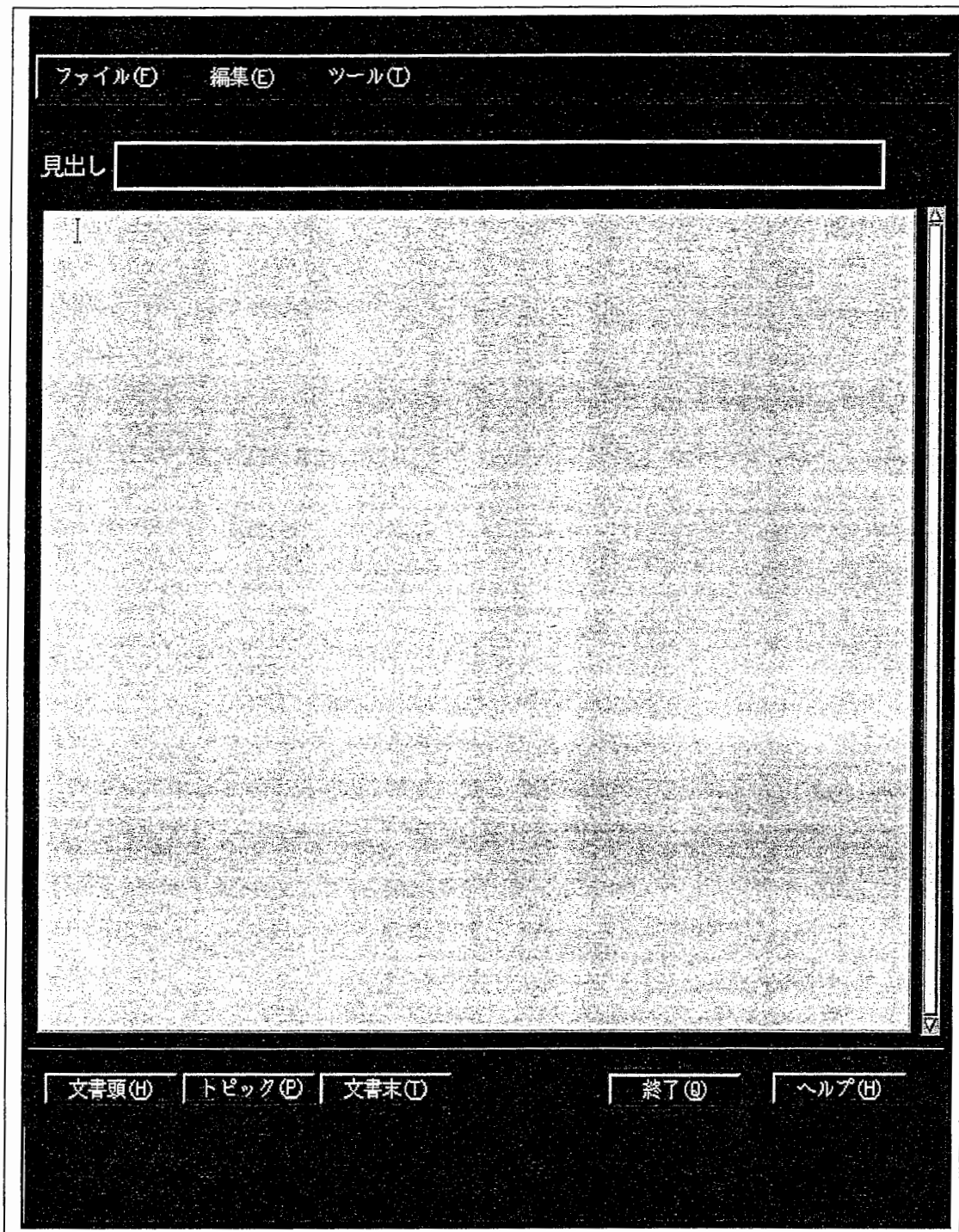


図 3.10: 新聞記事閲覧ウィンドウ

3.4.3 検索入力

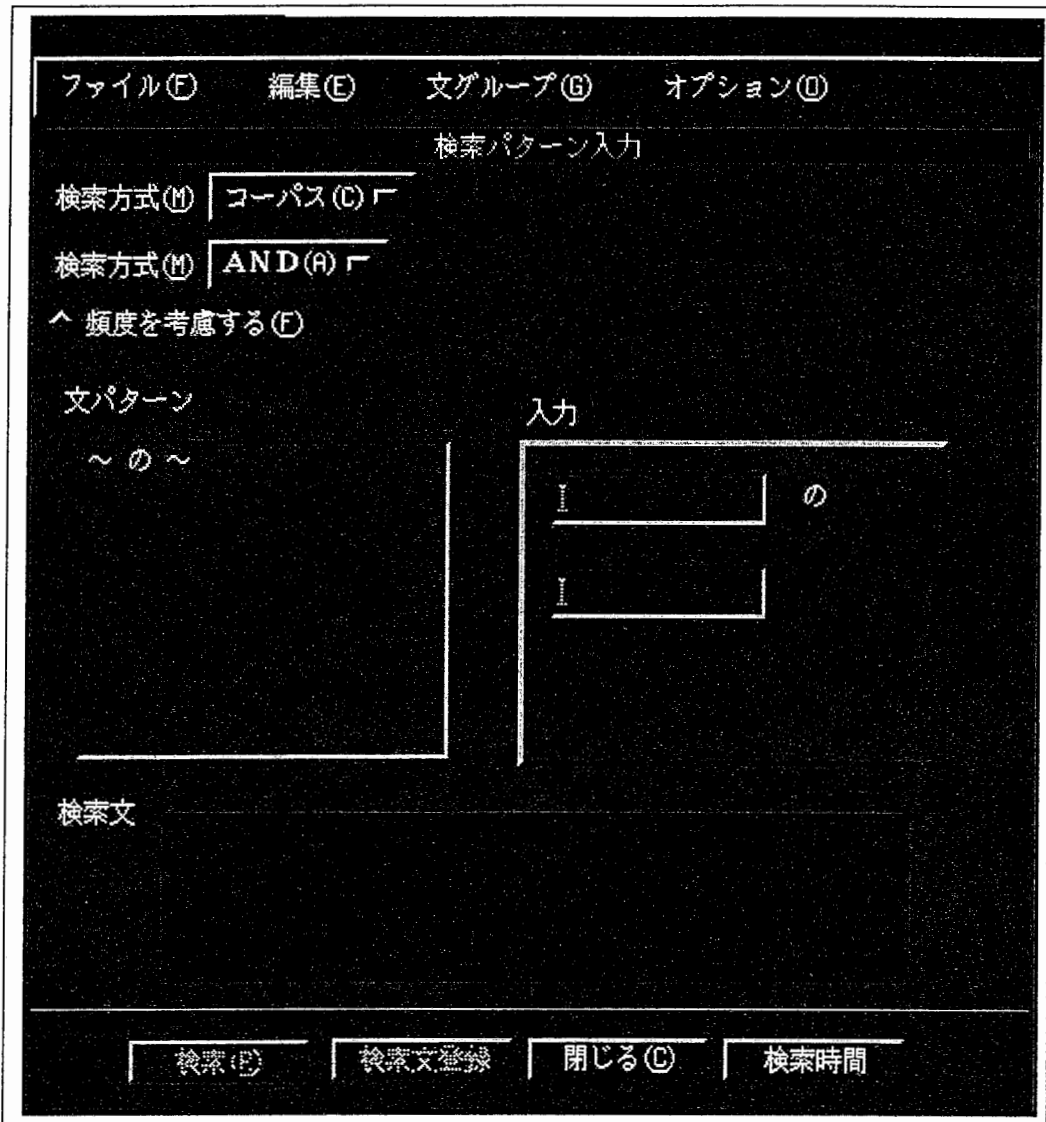


図 3.11: 検索入力ウィンドウ

1. プルダウンメニュー

ファイル, 編集は 3.4.2 と同様.

(a) 文グループボタン 文のグループを指定する

(b) オプション 検索結果をクリア, ヒストリーの消去

2. 検索方式ボタン

検索結果のトータル計算における AND/OR 演算に対し, 記事全体を対象にか, もしくはコーパス内のみを対象にするかを選択する.

以前の検索結果、もしくは複数の検索入力に対し、トータル計算方式を AND、あるいは OR 演算から選択する。

3. 頻度を考慮するボタン

トータル計算において、入力とヒットする用例の頻度を考慮する。

4. 文パターンウィンドウ

文パターンをクリックする。(入力ウィンドウも対応する)

パターンを変更する時は、マウスの右ボタンを押しながら、選択したいグループを選ぶ。メニューバーの文グループと同じ。

5. 入力ウィンドウ

マウスで、入力ウィンドウをクリックしフォーカスを行なう。

検索する語彙を入力する。

日本語を入力する時は kinput2¹ を使う。マウスの右ボタンを押すと、ヒストリーメニューがポップアップする。

6. 検索文リストウィンドウ

入力した検索文が表示される。

一度、検索した文はキャッシュされており、ダブルクリックにより瞬時に検索された状態へ戻る。

削除する時は、

(a) 削除するフレーズをマウスでクリックし反転させる。

(b) マウスの右ボタンを押して、削除ボタンを選択する。

マウスの右ボタンを押して、削除リストを選択すると、削除リストウィンドウ (3.4.4参照) がポップアップする。

7. 検索ボタン

検索を開始する。

8. 検索文登録ボタン

検索文を入力後にボタンを押して登録する。

検索文が1つの場合は登録しなくてもよい。

¹開始, 終了は Shift+Space.

9. 閉じるボタン

検索入力ウィンドウを閉じる。

10. 検索時間ボタン

検索にかかった時間を表示する。

3.4.4 削除リスト

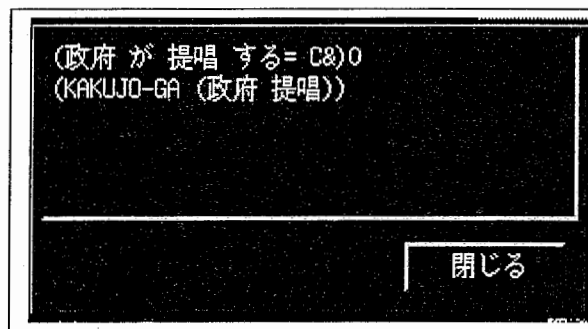


図 3.12: 削除リストウインドウ

検索文ウィンドウで削除された文がここに表示される。また、検索の際に用いた検索方式もここに表示される。

削除した検索文はこのリスト内に登録されるが、削除された検索結果を再び出す、もしくは検索文として登録する場合は、リスト内の対応する検索文をダブルクリックする。

リスト内の記号の説明 “C |”, “C&”, “A |”, “A&”, “1”, “0”

1. “C |” : 検索方式がコーパスでORである。
2. “C&” : 検索方式がコーパスでANDである。
3. “A |” : 検索方式が記事でORである。
4. “A&” : 検索方式が記事でANDである。
5. “1” : 頻度情報を考慮している。
6. “0” : 頻度情報を考慮していない。

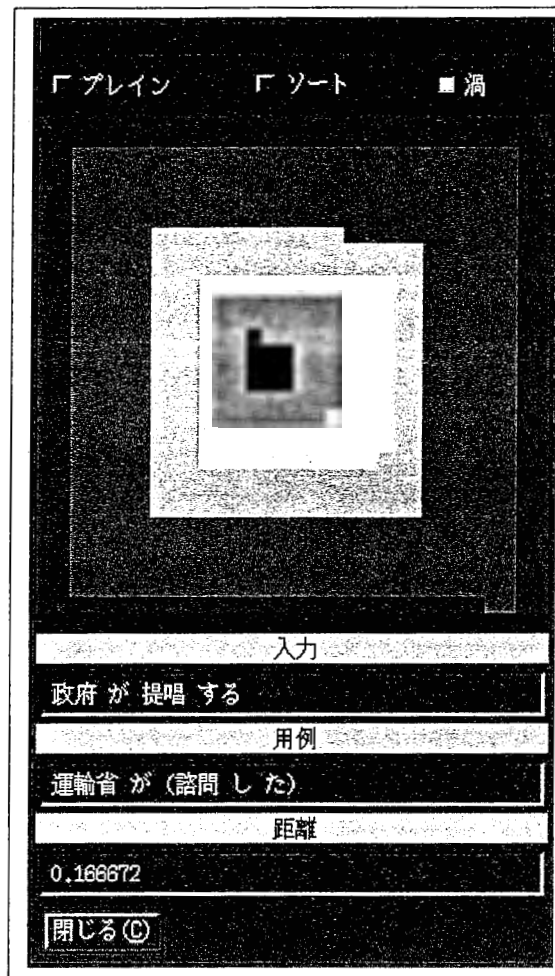


図 3.13: 意味距離の渦巻配列による視覚化ウィンドウ

3.4.5 フレーズレベルの用例空間 (意味距離の渦巻配列による視覚化)

1. ソートトグルボタン (プレイン/ソート/渦)
用例空間内の意味距離のソート方法を指定する。
2. 用例空間
意味距離を色によるメタファで表している。緑→黄→赤の順に意味的な距離が遠くなる。
用例空間内の任意の位置をマウスでクリックすると対応する意味距離にある用例と意味距離が表示され、記事表示ウィンドウのテキストもリアルタイムに切り替わる。
3. 入力ウィンドウ
入力された検索文が表示される。

4. 用例ウィンドウ
指示した意味距離にある用例が表示される。
5. 距離ウィンドウ
選択された用例との意味距離が表示される。
6. 閉じるボタン
用例空間ウィンドウを閉じる。

3.4.6 フレーズレベルの用例空間（用例の各要素に対応する2次元意味距離視覚化）

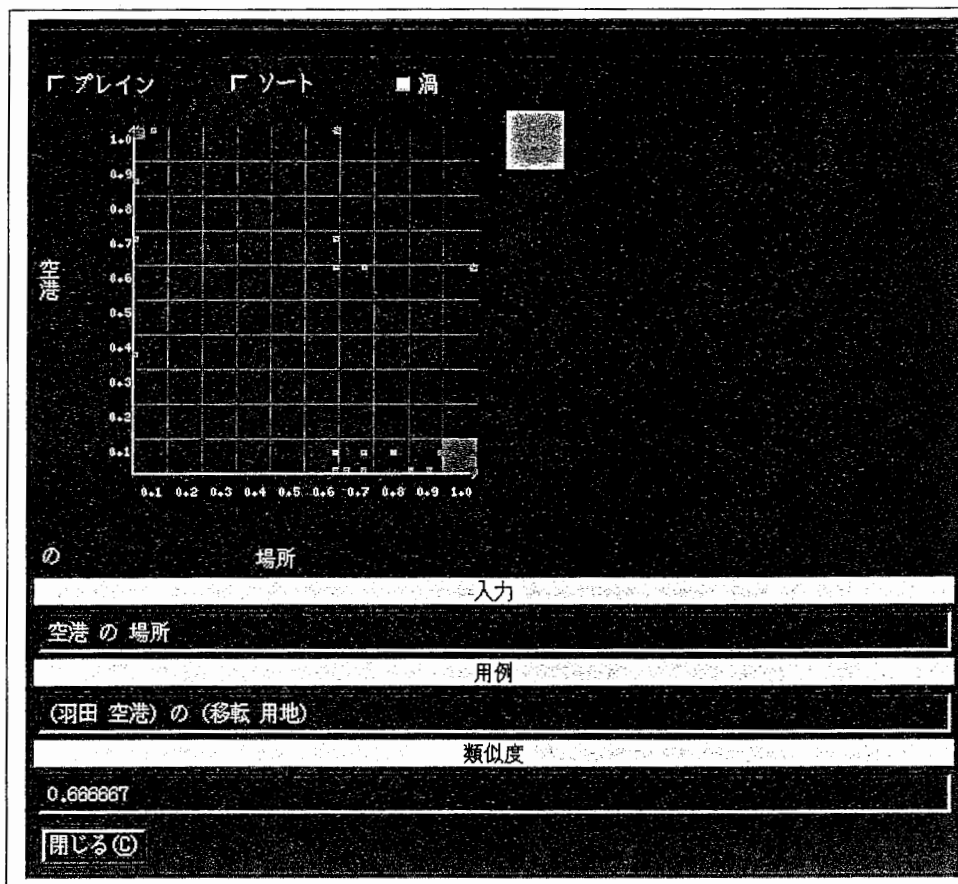


図 3.14: 用例の各要素に対応する2次元意味距離視覚化ウィンドウ

1. ソートトグルボタン（ブレイク／ソート／渦）
用例空間内の距離のソート方法を指定する。
2. 用例との二次元距離空間ウィンドウ（2次元格子）
検索入力語彙と用例の各要素との意味距離を2次元空間上に配列した視覚化ウィンドウ。格子上をクリックすると、その格子内の対応する用例が格子右側の用例空間上に視覚化される。
3. 用例空間（2次元格子右）
意味的な距離を色によるメタファで表している。緑→黄→赤の順に意味的な距離が遠くなる。
用例空間内の任意の位置をマウスでクリックすると対応する意味距離にある用例と意味距離が表示され、記事表示ウィンドウのテキストもリアルタイムに切り替わる。

4. 入力ウィンドウ
入力された検索文が表示される.
5. 用例ウィンドウ
該当する用例が表示される.
6. 類似度ウィンドウ
該当する用例との類似度²が表示される.
7. 閉じるボタン
意味距離視覚化ウィンドウを閉じる.

3.4.7 検索要求に対するトータル用例空間

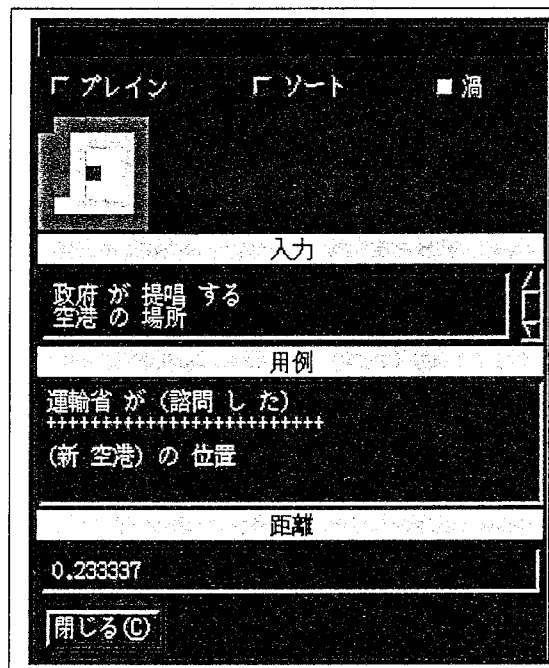


図 3.15: トータル用例空間

1. 用例空間の表示ラベル
検索方法を表示している.

²意味距離 α に対し, $\frac{1}{\alpha+1}$ により計算

2. ソートトグルボタン

用例空間内の意味距離のソート方法を指定する。

3. 用例空間

意味距離を色によるメタファで表している。緑→黄→赤の順に意味的な距離が遠くなる。

用例空間内の任意の位置をマウスでクリックすると対応する意味距離にある用例と意味距離が表示され、記事表示ウィンドウのテキストもリアルタイムに切り替わる。

各用例に対するフレーズレベルの用例空間ウィンドウを見るには、マウスの右ボタンを押し、ポップアップしたメニューより選択する。

4. 入力ウィンドウ 入力された検索文が表示される。

5. 用例ウィンドウ 該当する用例が表示される。

各検索入力に対応する用例がデリミター ++++++ で区切られ表示される。

6. 距離ウィンドウ 該当する用例との意味距離が表示される。

7. 閉じるボタン トータル用例空間ウィンドウを閉じる。

検索される用例が一つも無い場合は、用例空間ウィンドウには何も表示されない。

第 4 章

検索実験と実験結果に対する考察

4.1 プロトタイプシステムによる検索実験

前章での構成に基づき作成したプロトタイプシステムにおいて、新聞記事の検索を対象とする検索実験を行った。実験には「XのY」、「XでY」、「X{が,は,に,を,で}Yする」なる用例の言語表現に対し、約4万文の検索対象文(EDR 日本語コーパス [EDR93] を使用)より抽出した約15万の用例を用意し、記事数約2万のデータに対して任意のフレーズを入力し、逐次アルゴリズムを用いた意味距離計算により検索を行った。表 4.1に、実験に使用したターゲット・データベース(新聞記事8ヶ月分)の詳細データを示す。また、表 4.2には、記事のインデックスとして表 4.1のデータ中より用いた検索対象文のデータ、表 4.3には、表 4.2の文より抽出した用例の、各言語表現別の内訳を示す。

表 4.1: 検索実験に用いたターゲット・データベース(新聞記事8ヶ月分)の詳細

文数(見出し除く)	494,857
文字数(見出し除く)	23,596,836
記事数	38,115

尚、現在のプロトタイプシステムでは、意味距離計算以降の処理の実験評価を目的とし、実験には言語解析済みのフレーズを入力とした。以下に、検索実験の入力例と検索された用例(下線)、その用例を含む文を示す。

入力: 「日本の経済」

検索結果: “ニューヨークの株式暴落をきっかけに、世界同時に大暴落となっています。”

入力: 「マスコミの問題」

表 4.2: 記事のインデックスとして用いた文 (検索対象文) の詳細

文数	34,667
単語数	823,190 (延べ) 40,946 (表記異なり) 45,000 (表記, 品詞共異なり)
抽出した用例数	154,280

表 4.3: 抽出した用例の各言語表現別の内訳

用例の言語表現	数
XのY	43,132
XでY	12,198
XはYする	25,516
XがYする	22,272
XにYする	26,140
XをYする	25,022
計	154,280

検索結果: “ ～エイズをめぐる 報道とプライバシーの問題 などが次々起こり, ～. ”

入力: 「問題が生じる」

検索結果: “ 韓国労働省の調べによると, ～新たに 労使紛争が発生, 84件が解決した. ”

4.2 実験結果に対する考察

4.2.1 検索結果

類似検索を用いた検索では, 漠然とした検索要求に対しても思いつくフレーズを並べるだけで検索が行えるため, 比較的安易に検索要求を満たすことができる場合が多い。

しかし, 「政府が提唱する」と「政府の提唱」の例のように, 原言語表現が異なっても同義と考えられるものや, “で”格と“に”格のような場合によっては交換可能な格については,

現状の固定化された用例に対する意味距離計算のみではカバーできない。したがって、事前にパラフレーズ等のフィルタを通す必要がある。

また、類似検索については、意味距離の定義がシソーラスの構造によって変化し、さらにユーザの意図や状況等によっても動的に変化すると考えられるが、このようなゆれの問題に対して複雑なシソーラスの探索や、文脈状況の解析を行なうことは、計算コストの面を考慮すると必ずしも有効な手法とは言えない。しかし、シソーラスにおける概念の抽象度を比較的大きくとり（階層を浅く一様に分布させる）こと及び検索候補に対する視覚化の効果により、ある程度このようなゆれの問題を回避できると考えられる。類語辞典 [大野 84] によるシソーラス¹を用いた実験においても良好な結果を得た。ただし、検索候補が多くなる場合についてはシソーラスの選択等にさらなる考察の余地があろう。例えば、大井ら [大井 96] は、大規模シソーラスを用い、単語の類似度、関連度、頻度に基づいた検索手法を提案し、実験を行っている。また、前後 50 語の単語による文脈を考慮することで、文書中の単語の意味的曖昧性を解消する実験を行い、検索精度が向上することを確認している。

意味距離の視覚化による効果については、意味距離の同じ候補が多くなった場合の距離空間の細分化等に考察の余地がある。しかし、1) 意味的近さという比較的直感的な基準による視覚化、2) 距離空間上の任意の位置へのマウス・クリックによるダイレクトなインタフェース、3) 目的情報へのリアルタイム・アクセスによる対話的操作環境、などを実現することで、検索要求に対し概ね快適で良好な選択が可能となった。検索候補を列挙するのみの選択操作と比較しても、候補の絞り込みは直接的かつ迅速に行えるため、本手法は、候補選択の困難さに対する一解決策になりえると考えられる。

4.2.2 検索速度

本検索システムでは、検索手法として入力と目的言語表現との意味距離計算を用いているが、検索対象文（目的情報）が大規模になった場合には処理時間が膨大になってしまうという懸念がある。しかし、インデクシングやクラスタリングなどの高速化手法が適用でき、さらに、超並列計算による高速化の効果が報告されている [Sum95]。よって、検索対象が大量になった場合でも常に高速な検索処理を実現することが可能である。

¹約 6 万語の語彙に対し三階層、約 1100 の抽象概念より構成される。EDR 概念体系辞書（日本語約 20 万語に対して約 457,000 概念で構成、階層の深さ不定）と比較しても抽象度が非常に大きいと考えられる。

第 5 章

まとめと今後の課題

本報告では、類似検索を用いた情報検索システムの実現手法について述べた。類似検索では、意味距離計算を用い、入力に意味的に類似した言語表現を検索する。したがって、入力に意味的に類似した言語表現に付随する情報を検索することが可能になる。また、検索結果に対する効率的な候補の選択方法として、意味距離を視覚化し、任意の意味距離にある情報を提示する枠組みを提案した。本手法により、ユーザは任意の意味距離にある情報への動的アクセスが可能となる。

本システムで採用した意味距離計算は単純であり、大規模なデータベースに対する特定検索においても高速な検索が可能である。また、近年の用例に基づく言語翻訳においても入力文に類似した用例の検索手法として有効であるとの報告もなされている [古瀬 94]。本手法に基づくプロトタイプシステムを作成し、実験によりその有効性を確かめた。今後の課題としては、ユーザモデルの利用 [加納 91] によるユーザの個別要求への対応等が考えられる。また、アプリケーションとして、音声入力との結合や、インターネット上で特に必要とされる多言語入力に対する検索システムの実現等も興味深い課題である。

参考文献

- [EDR93] 日本電子化辞書研究所. EDR 電子化辞書仕様説明書. Japan, 1993.
- [Pai91] Chris D. Paice. A thesaural model of information retrieval. *Information Processing & Management*, Vol. 27, No. 5, pp. 433-447, 1991.
- [Sum95] Eiichiro Sumita, Kozo Oi, Osamu Furuse, Hitoshi Iida, and Tetsuya Higuchi. Example-based machine translation using associative processors. *Journal of Natural Language Processing*, Vol. 2, No. 3, pp. 27-48, 1995.
- [有田 95] 有田英一, 安井照昌, 津高新一郎. 情報空間の可視化による「情報散策」方式. In *Int. Sym. of Info-Tech '95*, pp. 20-26, Japan, 1995.
- [大井 96] 大井耕三, 隅田英一郎, 飯田仁. 単語間の意味的類似度に基づく文書検索手法. 言語処理学会第2回年次大会, No. A5-2, pp. 109-112, 1996.
- [大野 84] 大野晋, 浜西正人. 類語新辞典. 角川書店, Japan, 1984.
- [加納 91] 加納康男, 岸野文朗. ユーザモデルを用いた知的文献検索インタフェース. *信学論 (D-I)*, Vol. J74-D-I, No. 8, pp. 567-576, 1991.
- [武田 95] 武田浩一, 建石由佳. 情報の可視化と自然言語処理. In *Proceedings of the 50th Annual Convention of IPSJ*, pp. 57-64, 1995.
- [古瀬 94] 古瀬蔵, 隅田英一郎, 飯田仁. 経験的知識を活用する変換主導型機械翻訳. *情処学論*, Vol. 35, No. 3, pp. 414-425, 1994.
- [美馬 96] 美馬秀樹, 隅田英一郎, 飯田仁. 類似検索を用いた情報検索システム. 言語処理学会第2回年次大会, No. A5-3, pp. 113-116, 1996.

付録 A

類似検索システム実行ファイル及びツール類についての解説

A.1 類似検索システム実行ファイル

1. 類似検索システムの実行ディレクトリ

~ *mima/work/SBDR/DEMO/* 意味距離の渦巻配列による視覚化バージョン

~ *mima/work/SBDR/DEMO2/* 用例の各要素に対応する2次元視覚化バージョン

2. 各ファイルの説明

run.sbdr	類似検索システム実行シェルスクリプトファイル
kill.proc	類似検索システムプロセス削除シェルスクリプトファイル
sys_display	類似検索システム GUI 表示部本体
sbdr_sparc_client.image	類似検索処理部本体 (lisp イメージ)
sbdr_display.jrc	GUI 表示部: Motif リソースファイル (日本語表示版)
sbdr_display.erc	GUI 表示部: Motif リソースファイル (英語表示版)
examples.sbdr	表示用の用例を格納したファイル

corpus.head	用例番号から用例情報（新聞記事番号，用例を含む検索対象文の記事中の位置，検索対象文の文字長，記事サイズ）を得るテーブルファイル
article.head	新聞記事番号から記事情報（新聞データファイル番号，新聞データファイル中のオフセットアドレス）を得るテーブルファイル

3. 新聞データファイル

本デモシステムでは，検索の対象となるターゲット・データベースとして，朝日新聞記事データを用いている．以下が，検索に用いる新聞データファイルの格納ディレクトリである．

```
~ mima/work/ASAHI/
```

新聞データファイルの格納ディレクトリの変更や，データファイルの追加の際は，ソースファイル「`~ mima/work/SBDR/{DEMO, DEMO2}/disp_corpus.c`」の以下の部分を変更し，再コンパイルする．

```
#define NEWSPAPER_FILE_PATH "/home/as31/mima/work/ASAHI/"

char *DocumentFileName[]={ "T87A", "T87B", "T87AEUC", "T87BEUC" };
char *DocumentName[]={ "朝日新聞 ", "朝日新聞 ", "朝日新聞 ", "朝日新聞 " };
```

A.2 データファイル構築ツール

本デモシステムでは，実行の事前に次のデータファイルを作成しておく必要がある．

examples.sbdr

corpus.head

article.head

以下に、その作成ツールと使用方法を説明する。

1. 「examples.sbdr」

作成方法：類似検索システムの実行時、スクリプト「run.sbdr」の第2引数 [用例の読み込み方法] に、"socket" を指定する。

実行ディレクトリ：~ mima/work/SBDR/{DEMO, DEMO2}/

実行ファイル名：run.sbdr

書式：3.4節 参照。

実行例：run.sbdr j socket

2. 「corpus.head」, 「article.head」

作成方法：「makecdata」を実行する (図 A.1)。

実行ディレクトリ：~ mima/work/SBDR/tools/

実行ファイル名：makecdata

書式：makecdata

実行例：makecdata

備考：データファイル作成後、類似検索システム実行ディレクトリ (~ mima/work/SBDR/{DEMO, DEMO2}/) にコピーする。尚、新聞データファイルの格納ディレクトリ、検索対象文データのファイル名を変更する場合は、ソースファイル「~ mima/work/SBDR/tools/ReadASAHI.c」の以下の部分を変更し、再コンパイルする。

```
#define ASAHI_FILE_PATH "/home/as31/mima/work/ASAHI/"  
#define ASAHI_CORPUS_TEXT_FILE_NAME "corpus.asahi.text"
```

A.3 用例データのコーパスからの抽出

類似検索のインデックスとして用いるデータは、EDR 日本語コーパスより抽出した「XのY」、「XでY」、「X{が,は,に,を}Yする」の言語表現に対応する用例である。以下に抽出方法を説明する。

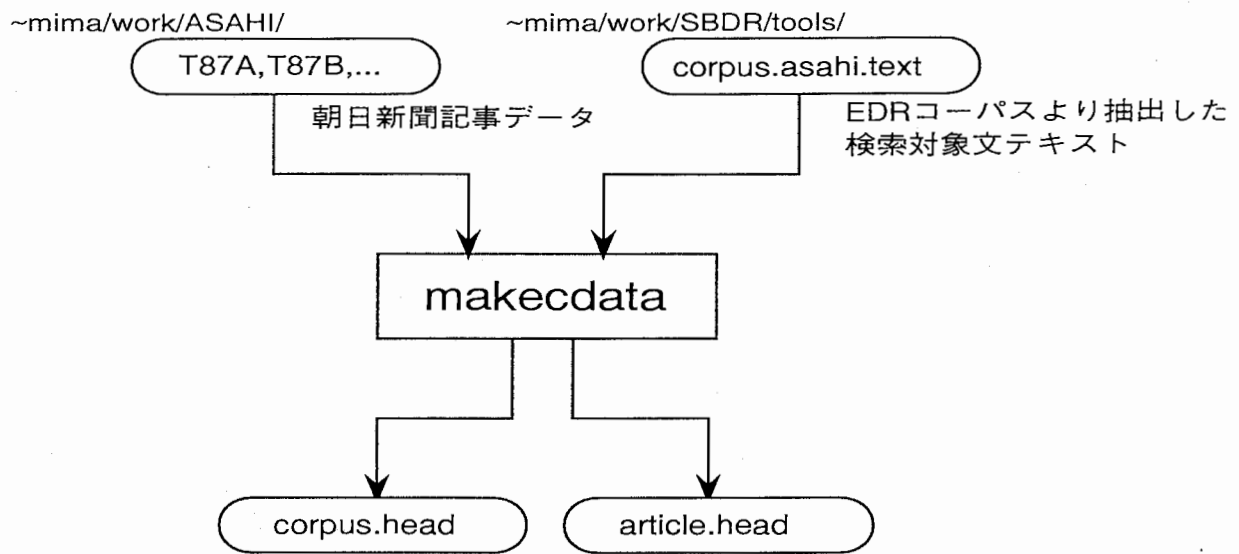


図 A.1: corpus.head, article.head の作成

抽出方法：「getexamples」を実行する（図 A.2）。

実行ディレクトリ：~ mima/work/SBDR/tools/

実行ファイル名：getexamples

書式：getexamples [抽出する言語表現の種類]

(a) 抽出する言語表現の種類

- * no ... “「XのY」”の言語表現に対応する用例を抽出する。
- * de ... “「XでY」”の言語表現に対応する用例を抽出する。
- * ga ... “「XがYする」”の言語表現に対応する用例を抽出する。
- * ha ... “「XはYする」”の言語表現に対応する用例を抽出する。
- * ni ... “「XにYする」”の言語表現に対応する用例を抽出する。
- * wo ... “「XをYする」”の言語表現に対応する用例を抽出する。

実行例：getexamples no

備考：抽出された用例データは標準出力 (stdout) に出力される。用例データは、類似検索処理部を再構築する際にロードする。尚、EDR辞書の格納ディレクトリを変更する場合は、以下の環境変数のディレクトリ部分を変更する。

```
CDDICT=/data/as31/mima/EDR/dict/EDcd2.1
CODICT=/data/as31/mima/EDR/dios/EDco1.0
WDDICT=/data/as31/mima/EDR/dict/EDwd2.1
WDIDX=/data/as31/mima/EDR/dios/EDci1.0
```

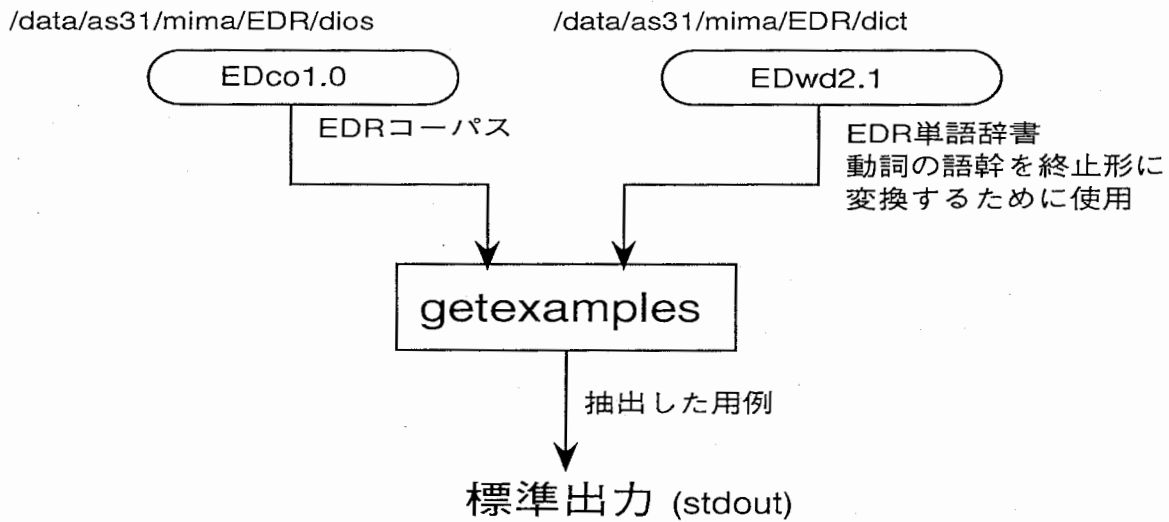


図 A.2: 用例データのコーパスからの抽出

付録 B

類似検索システムの操作例

B.1 入力文のサンプルと検索例

以下に、入力文のサンプルを掲載する。検索パターン入力ウィンドウより入力を行なう。

政府が提唱する & 空港の場所

- 文パターンから「～が～する」を選択する。
- “政府”，” 提唱” を入力する。
- 検索ボタンを押す。
- B.1 が現れる。
- 文パターンから「～の～」を選択する。
- “空港”，” 場所” を入力する。
- 「政府が提唱する」との検索方式を選択する。
- 検索ボタンを押す。
- B.2が現れ、検索結果としてB.3（検索方式が'コーパス /AND' の場合）の記事が得られる。

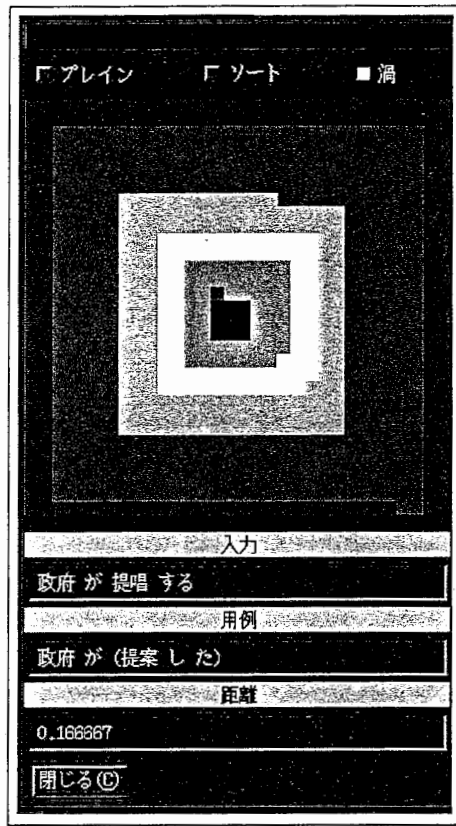


図 B.1: 政府が提唱する

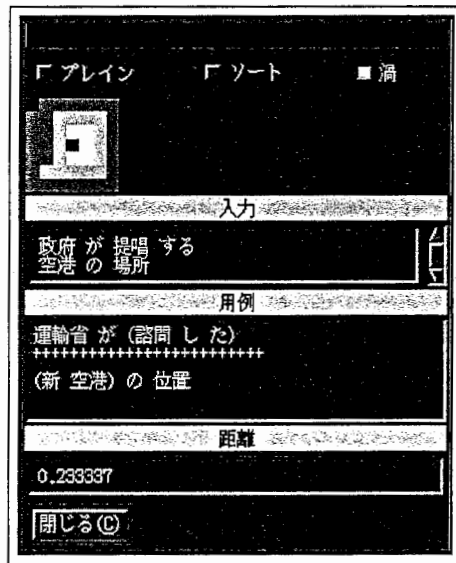


図 B.2: 政府が提唱する, 空港の場所

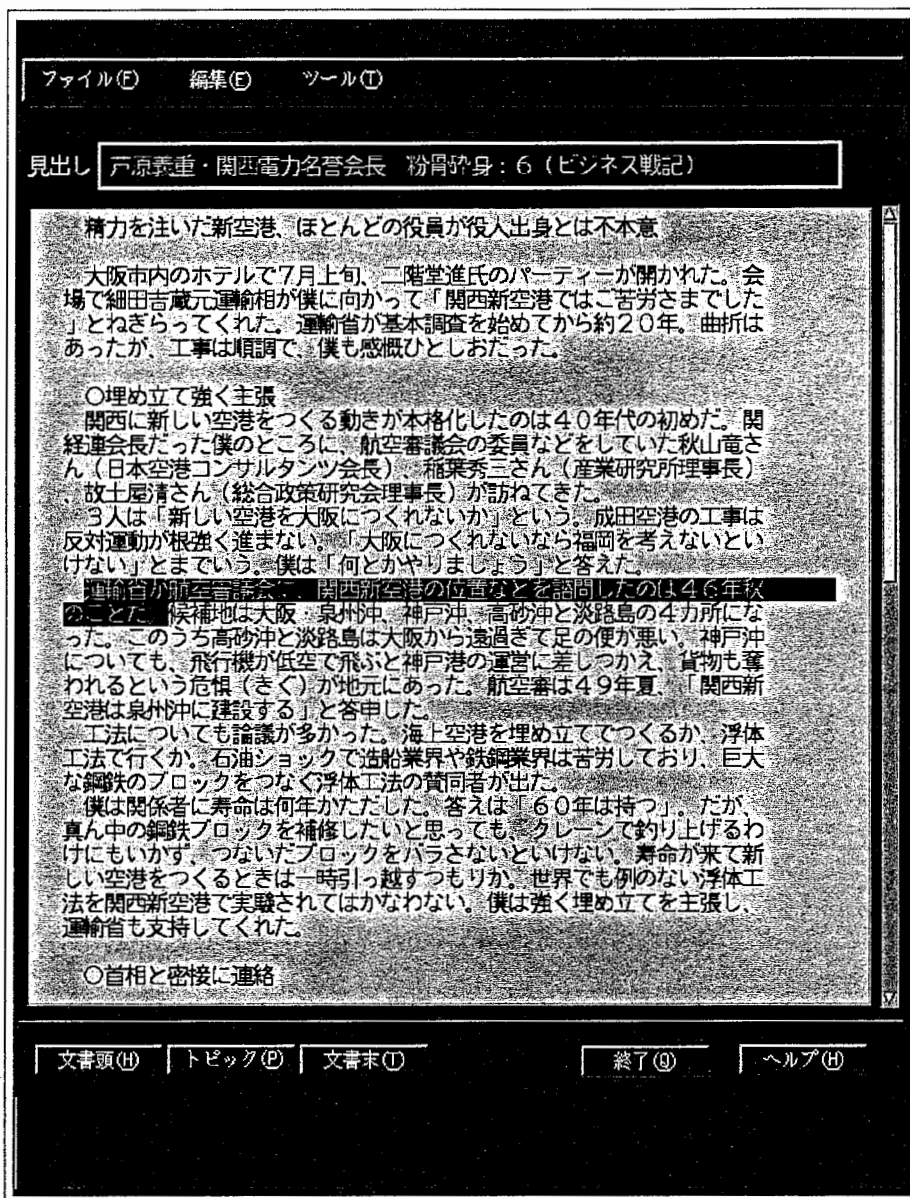


図 B.3: 検索された記事例（政府が提唱する、空港の場所）

政府が提唱する & 売上税に反対する

- 文パターンから「～が～する」を選択する.
- “政府”, ” 提唱” を入力する.
- 検索ボタンを押す.
- B.1 が現れる.
- 文パターンから「～に～する」を選択する.

- “売上税”，” 反対” を入力する.
- 「政府が提唱する」との検索方式を選択する.
- 検索ボタンを押す.
- B.4 が現れ，検索結果として B.5（検索方式が’コーパス /AND’ の場合）の記事が得られる.

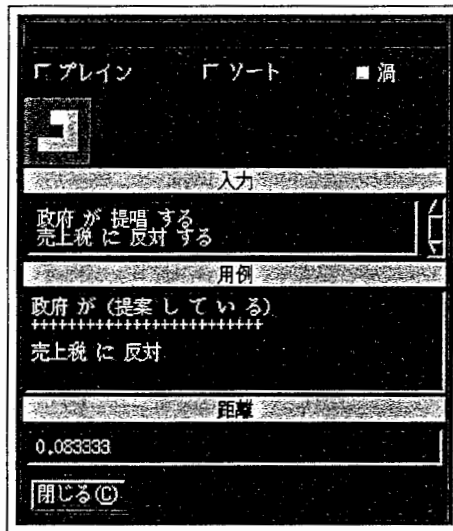


図 B.4: 政府が提唱する, 売上税に反対する

マスコミの問題 & 問題が生じる

- 文パターンから「～の～」を選択する.
- “マスコミ”，” 問題” を入力する.
- 検索ボタンを押す.
- B.6 が現れる.
- 文パターンから「～が～」を選択する.
- “問題”，” 生じる” を入力する.
- 「日本のマスコミ」との検索方式を選択する.
- 検索ボタンを押す.
- B.7 が現れ，検索結果として B.8（検索方式が’コーパス /AND’ の場合）の記事が得られる.

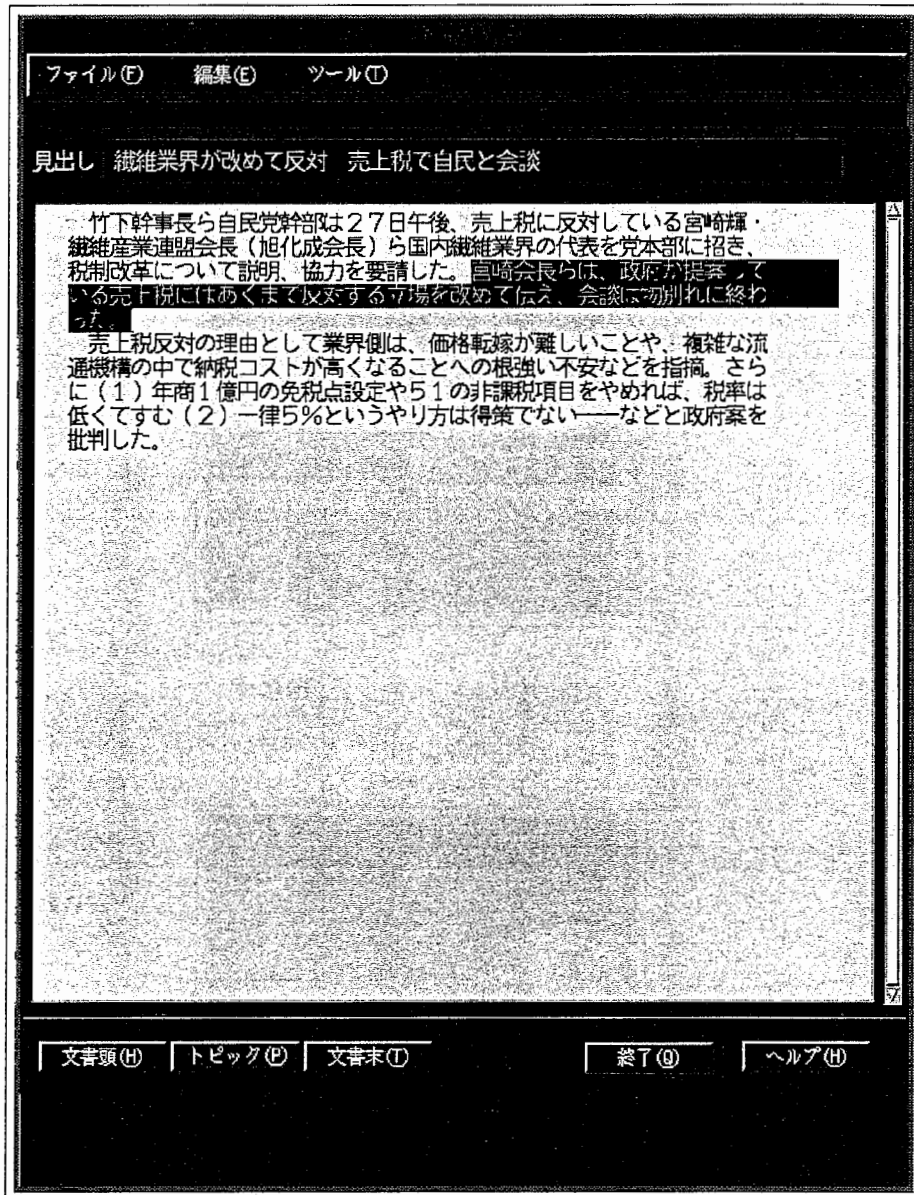


図 B.5: 検索された記事例（政府が提唱する、売上税に反対する）

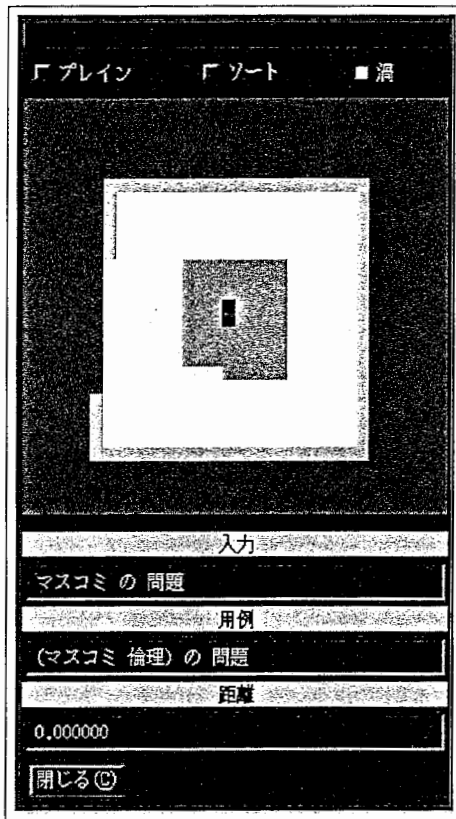


図 B.6: マスコミの問題

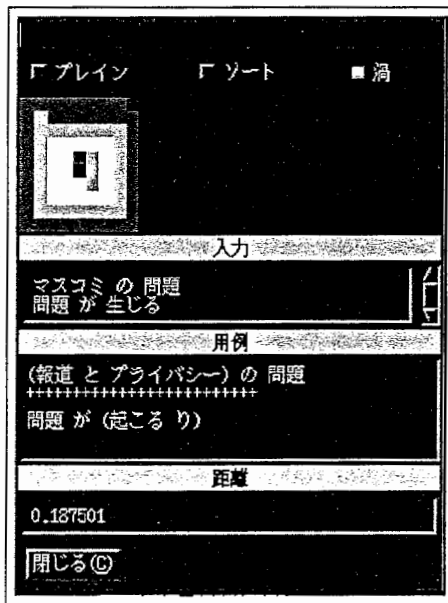


図 B.7: マスコミの問題, 問題が生じる

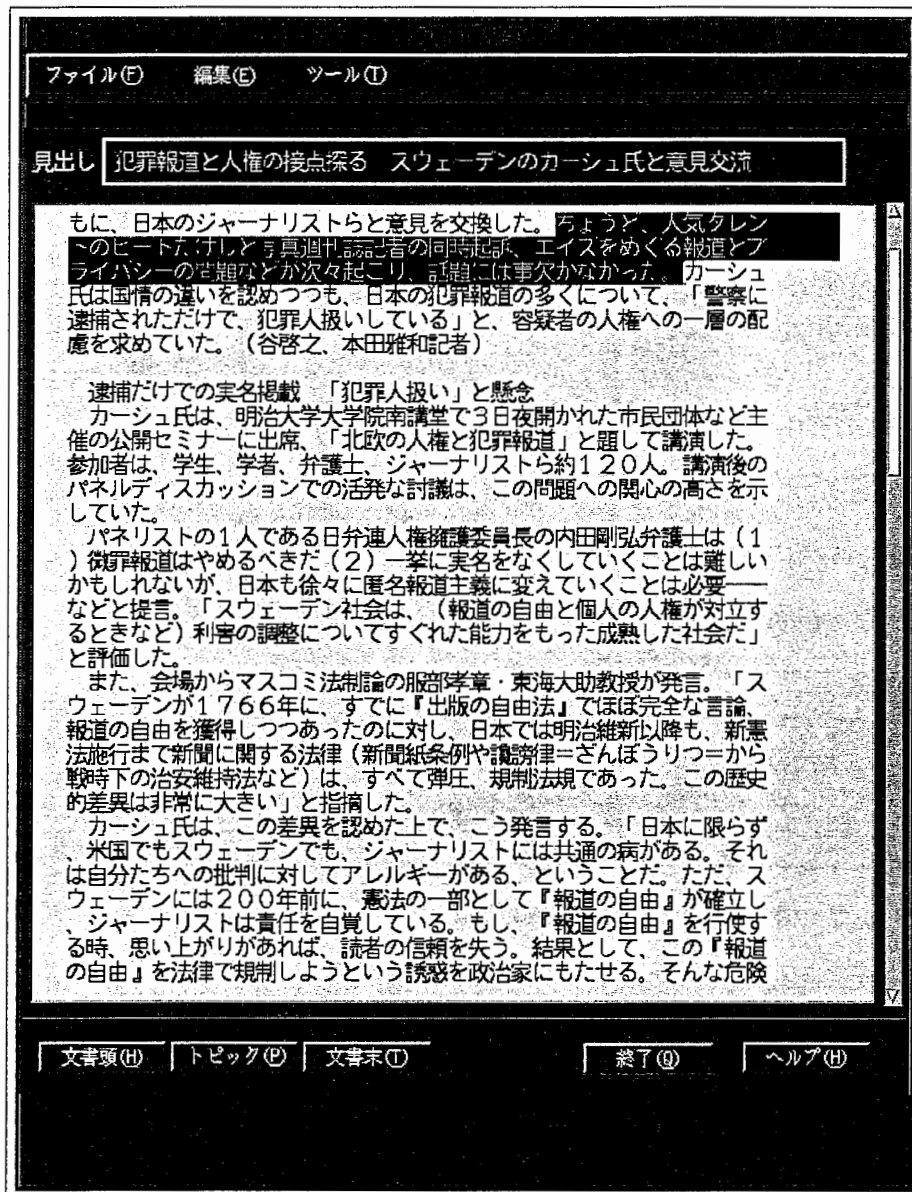


図 B.8: 検索された記事例 (マスコミの問題, 問題が生じる)

付録 C

トラブル・シューティング

C.1 システム起動時の異常

- システム起動後、しばらく (30 秒以上) しても、システムのウィンドウが現われない場合の原因と対処

1. 原因, 確認, 対処

(原因) 類似検索システム本体のプロセスが既に起動中である。

(確認) `ps -x` コマンドを実行し、

```
sys_display -p ソケット名1
```

```
例: sys_display -p SBDR-SEQ
```

が2つあれば、既に起動中である。

(対処) 既に起動されているシステムを終了させる。

2. 原因, 確認, 対処

(原因) 本体システムが稼働するマシンでリモートシェルコマンドが動かない。

(確認) 本表示システムのホストマシンから

```
rlogin 本体システムが稼働するマシンのホスト名
```

```
例: rlogin as05
```

を実行し、ログインできるかどうかを調べる

(対処) (a) 類似検索システム本体が稼働するホストにログインできない理由を調査し、回復させる。

(b) 今、起動したプロセスを削除 (kill) したいなら。

```
kill.proc
```

を実行する¹。

(c) (本体システムが稼働するホスト) にログイン出来るようになったら、再度、run.sldr を実行する

¹kill.proc においても、類似検索システム本体が稼働するホストにログインしようとするので、プロンプトは返らない。しかし、少なくとも30秒程度待った後、control+C (コントロールキーを押しながらCキーを押す) により中断することで、GUI表示部が稼働するホスト上における類似検索システム本体プロセスは削除 (kill) されているはずである

索引

- 類似検索システムのプロセスを削除するシェルスクリプトの設定, 16
- 類似検索システムのモジュール構成, 11
- 類似検索システムの稼働環境, 9
- 類似検索システムの稼働環境の設定, 17
- 類似検索システムの実行ディレクトリ, 34
- 類似検索システムの操作例, 39
- 類似検索システムの利用方法, 16
- 類似検索システム起動時のソケット通信プロトコル, 13
- 類似検索システム起動時のソケット通信手順, 11
- 類似検索システム起動用のシェルスクリプトの設定, 16
- 類似検索システム実行ファイル, 34
- 類似検索システム実行ファイル及びツール類についての解説, 34
- 類似検索システム終了時のソケット通信プロトコル, 15
- 類似検索システム終了時のソケット通信手順, 12
- システムの起動方法, 18
- システム起動時の異常, 46
- シソーラス, 1, 31
- ソケット通信のプロトコル, 13
- ソケット通信の手順, 11
- データファイル構築ツール, 35
- トータル用例空間, 27
- トラブル・シューティング, 46
- フレーズレベルの用例空間, 24, 26
- プロトタイプシステムによる検索実験, 29
- ユーザインタフェース, 18
- 意味距離の渦巻配列による視覚化ウィンドウ, 24
- 意味距離計算, 4
- 記事のインデックスとして用いた文（検索対象文）の詳細, 30
- 検索実験に用いたターゲット・データベースの詳細, 29
- 検索速度に対する考察, 31
- 検索対象文, 3
- 検索入力, 21
- 削除リストウィンドウ, 23
- 実験結果に対する考察, 30
- 情報の視覚化, 6
- 新聞データファイル, 35
- 新聞記事ビューアー, 19
- 入力文のサンプルと検索例, 39
- 目的情報, 3
- 用例, 3
- 用例の各言語表現別の内訳, 30
- 用例の各要素に対応する 2 次元意味距離視覚化, 26
- 用例の言語表現, 3

用例データのコーパスからの抽出, 36

用例空間, 7

類似検索, 1

類似検索におけるインデックス, 3

類似検索を用いた情報検索の概要, 3

類似検索システムの構成, 8

類似用例検索時のソケット通信プロトコル,

14

類似用例検索時のソケット通信手順, 12