

汎用的な情報検索音声対話プラットフォーム

田中 克明 河原 達也 堂下 修司

京都大学 大学院 工学研究科 情報工学教室

〒606-8501 京都市 左京区 吉田本町

e-mail: katsuaki@kuis.kyoto-u.ac.jp

あらまし 汎用的な情報検索音声対話プラットフォームの設計と実装に関して報告する。本プラットフォームを用いることにより、アプリケーション開発者は複雑な文法や語彙の設定を行うことなく、手持ちのデータベースに対して音声対話による情報検索システムを容易に実現できる。また、この対話システムは発話パターンを明示するためユーザは戸惑うことなく発話が行え、内部状態をインタラクティブに提示するためユーザは即座に誤りの訂正が行える。本プラットフォームを適用してホテル検索システムと文献検索システムを実現できた。

キーワード 音声対話, 情報検索, プロトタイプング

Domain Independent Platform of Spoken Dialogue Interface for Information Query

Katsuaki Tanaka Tatsuya Kawahara Shuji Doshita

Department of Information Science

Kyoto University, Kyoto 606-8501, Japan

Abstract We designed and implemented a domain independent platform of spoken dialogue interface for information query. Using this platform, system developers can easily make an information query system for their database with spoken dialogue interface without specifying a complicated grammar and vocabulary for the spoken language module. Because this dialogue system presents acceptable utterance patterns and interactively indicates system's status to users, users can smoothly make utterances or correction in the dialogue. We applied the platform to hotel search system and literature search system.

key words spoken dialogue interface, information query, prototyping

1 はじめに

近年の計算機の普及に伴い、図書館や博物館などの公共施設を中心に各種データベースの整備と検索サービスの公開が広く行われている。このような情報検索システムはキーワードを入力し検索するシステムが多いため音声入力インタフェースの利用が有効であると考えられるが、実際にはキーボードを用いたものがほとんどである。

これには様々な原因が考えられるが、その一つに音声入力インタフェースはコストが高いということを挙げることができる。音声入力インタフェースの導入には文法の記述や適切な語彙の設定等、専門の知識と経験が必要とする多くの作業が伴う。統計的手法を用いればコーパスから必要な文法と語彙を自動生成できるが、タスクに応じたデータを大量に集める必要があり、容易であるとは言えない。また特定のシステムのために記述された文法・語彙や、特定のシステムのために収集されたコーパスはポータビリティが低い。そのため同じような音声対話情報検索システムでも、タスクが異なるだけで文法や語彙を始めから構築しなおさなければならない。

そこで本研究では、データベースからの情報をもとに文法と語彙の設定を半自動的に汎用的な情報検索音声対話プラットフォームを設計・作成する。本プラットフォームはタスクを情報検索に限定し、GUIを採用して発話パターンをユーザに明示することでユーザの発話を誘導するため、限定された発話・対話パターンで音声対話を実現できる。そのため本プラットフォームは文法と語彙の設定を半自動的に、ドメインに依存しないので、様々な種類のデータベースに対して汎用的に用いることができる。

マルチモーダルインタフェース全般を対象としたプロトタイプシステムの研究 [1] はすでに行われているが、本研究ではタスクを情報検索に限定することにより、文法・語彙を直接記述しなくても良いなどの利点を備えたより容易な音声対話インタフェースの実現を目指す。

2 汎用的情報検索音声対話プラットフォームの概要

本プラットフォームは様々なタスクのデータベースに対して、容易に音声対話による情報検索サービスを実現することを目的とする。アプリケーション開発者はデータベースを用意しいくつかの語彙の登録・選

択を行うだけで、文法を記述することなくタスクの設定が行える。また提供される音声対話情報検索システムは、マルチモーダルな応答による認識結果の提示や発話パターンの可視化などの特徴を備えている。

本プラットフォームの概要を図1に示す。本プラットフォームはタスク規定ファイル自動生成システムと汎用的音声対話情報検索システムの2つのシステムで構成される。手持ちのデータベースに対して音声対話情報検索システムの作成を行うアプリケーション開発者と、検索システムを利用するユーザの2種類の使用者が存在する。

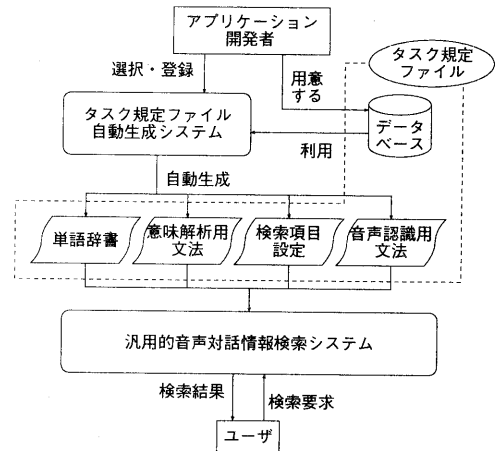


図1: 汎用的音声対話情報検索プラットフォームの概要

タスク規定ファイル自動生成システムは、アプリケーション開発者が手持ちのデータベースに対する音声対話情報検索システムを作成する際に使用するシステムである。このシステムを利用して検索システムの仕様を指定すると、システムが単語辞書、音声認識用文法、意味解析用文法、検索項目設定の4つを自動生成する。なお、ここではこの4つにデータベースを加えた5つのファイルをタスク規定ファイルとよぶ。

汎用的音声対話情報検索システムはタスク規定ファイルに従い、特定のデータベースに対する音声対話情報検索システムとして動作する、汎用的な情報検索システムである。このシステムの使用者であるユーザはすべての入力を音声により行う。曖昧な発話がなされた場合はシステムがユーザに問い返し、対話的に解決する。

3 汎用的音声対話情報検索システム

本システムは音声によるユーザの指示で情報検索を行う。本システム自身は汎用的であり、特定のタスクやデータベースに依存しない。定められた形式に沿ったタスク規定ファイルを与えることで、特定のデータベースに対する音声対話情報検索システムとして動作する。概要を図2に示す。

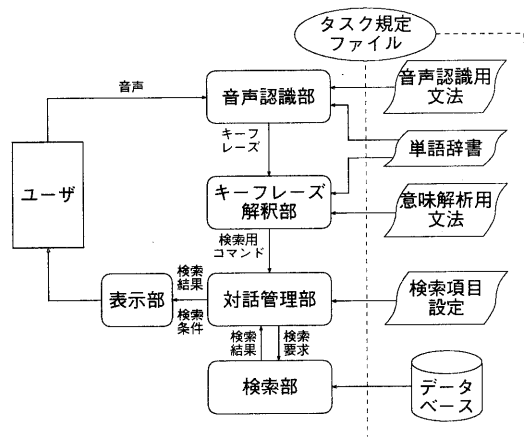


図 2: 汎用的情報検索システムの概要

本システムの外観の例を図3に示す。検索条件はすべてAND条件に限定している。各検索項目は特定のデータベース項目に対応しており、対応するデータベース項目がその検索条件を満たす場合に検索結果として表示される。検索条件はユーザの発話に応じて埋められていき、検索は1発話毎に行われるが、候補が多すぎる場合はヒットした件数のみを表示し、ユーザにさらに候補を絞り込むことを促す。検索結果が一定件数以下の場合にはヒットしたものすべてが表示される。また本システムでは、ユーザの発話の意図はその発話に含まれる個々のキーフレーズの意味を組み合わせることで構成できると仮定し、ユーザの発話からキーフレーズを抽出してから各キーフレーズ毎に意味解析を行う。システムが受理できるキーフレーズには検索条件を追加するフレーズの他に、UNDOフレーズや検索条件の個別削除・全削除を行うフレーズも含まれており、検索条件の全削除フレーズを発話すれば情報検索を最初からやり直すことができる。

従来の音声対話システムはどのような発話でも受理できるという訳ではないにも関わらず、ユーザが何を発話して良いのか分からないという問題点があっ

ホテル検索システム

検索条件

所在が

ホテルタイプが

(ビジネス、カプセル、旅館から選択)

シングル料金が 以下

ツイン料金が 以下

付属施設に がある

検索結果

名称 **ホテル

所在地 奈良県 奈良市

電話番号0742-***-****

.....

.....

図 3: 検索システム外観

た。このためマイクを前にして戸惑うユーザが多く、結局受理できる発話のパターンを示したマニュアルを用意し、事前にユーザに見せておく必要があった。また認識誤りの発生によりユーザと計算機との間に誤解が生じ、対話の破綻を招く事態がしばしば発生した。

本システムでは図3のような外観を持つインタフェースを用いることで、これらの問題の解決を図る。まず、画面に発話パターンが明示されているためユーザは戸惑うことなく発話でき、システムにとってはユーザの発話を誘導できる。そのため、あらゆる発話を受理できるような複雑な文法を書かなくとも、ユーザの発話の認識が可能である。本プラットフォームは汎用性を高めるためにタスクの知識を用いないので、完全な文法を用意することが出来ないが、発話パターンの明示によりその短所を補う。ただし、発話パターンからの多少の変動には対応できるよう、文法を用意する。

次に、図3の空欄はユーザの発話に応じてインタラクティブに埋められる。ユーザは認識結果と現在の検索条件を容易に把握できるため、システムが認識誤りを起こしても把握でき、対処することができる[2]。致命的な認識誤りが発生した場合は、UNDOフレーズにより一つ前の発話の状態に戻ることが可能であり、また一部に誤りが生じた場合はその部分だけ言い直すことで誤りから回復することができる。

次に本システムの各構成要素について説明する。

3.1 音声認識部

音声認識部は入力された音声単語列に変換し、キーフレーズを抽出する。

本システムでは音声認識パーザに当研究室で開発された Julian [3] を使用し、オートマトンパーザによる解析を行う。Julian は、探索手法に単語間の接続に関する制約である単語対文法をヒューリスティックとして用いた A* 探索を採用している [4]。ここでは音素モデルに、1 状態あたり 16 連続混合分布を持つ monophone モデル [5] を用いている。

3.2 キーフレーズ解釈部

キーフレーズ解釈部ではキーフレーズの意味解釈を行い、認識されたキーワードを検索用コマンドの表現に変換する。

キーフレーズの意味解釈は単語辞書と意味解析用文法を用いて行い、フレーズの種類を特定しキーワードを抽出する。フレーズが検索条件の追加フレーズであれば、対応する検索項目にキーワードの解釈を行った後、検索条件として登録する。検索条件の個別削除フレーズであれば、対応する検索項目の検索条件をクリアする。全削除フレーズであればすべての検索条件を削除し、UNDO フレーズであれば検索条件を 1 つ前の発話の状態に戻す。

抽出されたキーワードには「所在が奈良県のホテルを」というフレーズの「奈良県」のようにそのまま検索キーとなるキーワードもあるが、「搭乗月日が今日の」というフレーズの「今日」のように、そのままでは検索キーとして用いることのできないキーワードもある。「今日」が何日を表すのかは静的に記述できず、実行時に解決する必要がある。

そこでシステムにいくつかの動的な解釈方法を用意しておき、「今日」「昨日」「明日」「今年」等のキーワードにどの解釈方法を用いるかを指定しておく。動的な解釈に対応したキーワードは「日付」「年」などの概念毎に準備しており、アプリケーション開発者がタスク規定ファイル自動生成システムで指定することにより容易に語彙に組み込むことができる。

3.3 対話管理部

対話管理部では曖昧性の解消と省略補完を行い、検索を実行する。

本システムで発生する曖昧性は、次の 2 種類が考えられる。

- 条件追加フレーズと検索項目の対応が一意に定まらない
- キーワードに同音異義語が含まれる

以下で、この 2 つの曖昧性の解消方法を述べた後、省略の補完について述べる。なお、この 2 種類の曖昧性が同時に発生する場合も考えられるが、まずフレーズと検索項目の対応を定め、その後に同音異義語の区別をつける。

3.3.1 フレーズと検索項目の対応の曖昧性

本システムではユーザの発話はフレーズ毎に処理され、各フレーズは検索項目と 1 対 1 の対応が取られる。この時、一意に検索項目との対応が取れずに曖昧性が発生する場合がある。例えばホテル検索においてユーザが「1 万円以下の」というフレーズを発話したとすると、このフレーズには「シングル料金が [] 円以下」「ツイン料金が [] 円以下」の 2 つの検索項目が対応し曖昧性が発生する。この場合システムは曖昧性を含むフレーズがどの検索項目に対応するのかをユーザに問い返す。上記のホテル検索の例であればシステムは

10,000 円はシングル料金ですか、ツイン料金ですか。

それとも、この検索条件を取り消しますか。

とユーザに問いかける。この場合のユーザの返答は

- シングル料金です。
- ツイン料金です。
- 取り消します。

のいずれかであると予想できる。

3.3.2 同音異義語の区別

同一の検索項目のキーワードに同音異義語が含まれる場合、曖昧性が発生する。例えば文献検索において「著者が『かわはら』の」というフレーズをユーザが発話し、著者のキーワードに「河原」と「川原」が存在した場合、この発話からは区別がつかない。この場合キーワードを音声で読み上げるだけでは候補の区別がつかないので、候補に番号を付与し、ユーザには番号で選択してもらう。先ほどの文献検索の例であればシステムは

“かわはら”は
 1 番 河原 ですか、
 2 番 川原 ですか。
 両方で検索しますか
 それとも、この検索条件を取り消します
 か。

とユーザに問いかける。この方法は同音異表記語の場合は区別がつくが、例えばホテル検索でユーザが「所在が」おおみやちょう”の」と発話した場合に三重県大宮町と京都府大宮町の区別をつけたい時のような、同表記異義語の区別をつけることはできない。そこで、前後の情報を用いて曖昧性を解消することを考える。このホテル検索の例であればシステムは

“おおみやちょう”は三重県大宮町ですか、
 京都府大宮町ですか。
 両方で検索しますか。
 それとも、この検索条件を取り消します
 か。

とユーザに問いかける。

3.3.3 省略の補完

キーワードで省略が行われている場合は、default 値を用いて省略の補完を行う。例えば「16 日の」というフレーズには何月の 16 日なのかという情報が含まれていないので、今月の 16 日と解釈する。ただ、過去に 12 月に関する検索が行われていれば 12 月を default 値に設定し、12 月 16 日と解釈する。

3.4 検索部

対話管理部から受け取った検索条件を用いて、データベースに対する検索を行う。検索はユーザの 1 発話毎に実行され、検索条件を満たす件数が多すぎる場合はヒットした件数のみを表示する。

3.5 表示部

各検索項目毎に発話パターンと現在の検索条件を表示し、検索が実行される毎に検索結果を表示する。

4 タスク規定ファイル自動生成システム

本システムは既存のデータベースに対して音声対話情報検索を実現するために、汎用的音声対話情報検

索システム用のタスク規定ファイルを生成するシステムである。図 4 にシステムの概要を示す。

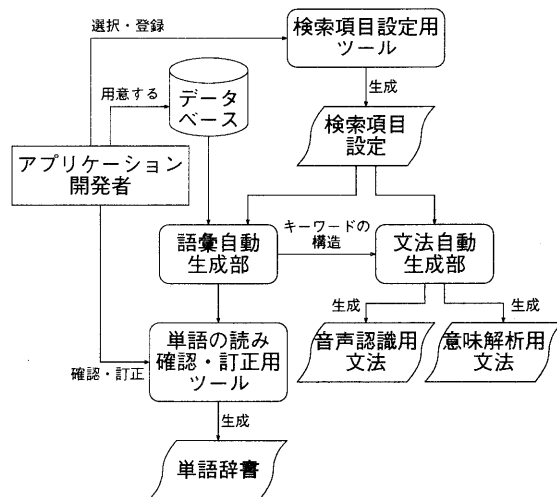


図 4: タスク規定ファイル自動生成システムの概要

本システムは検索項目設定用ツールと語彙の読み確認・訂正用ツールの 2 つのツール、そして語彙自動生成部と文法自動生成部の 2 つのモジュールから成る。アプリケーション開発者はまず検索項目設定用ツールを用いて、情報検索システムの詳細を決定し、検索項目設定ファイルを作成する。これに従って文法自動生成部が文法を、語彙自動生成部が語彙を自動生成する。この際にデータベースから自動抽出された単語には読みが自動的に付与されるが誤りが予想されるので、単語の読み確認・訂正ツールを用いて、読みの訂正を行う。

以下では、タスク規定ファイルと 2 つのツール及び 2 つのモジュールについて説明する。

4.1 タスク規定ファイル

タスク規定ファイルとは、データベース、単語辞書、音声認識用文法、意味解析用文法、検索項目設定の 5 つである。このうち、データベースはアプリケーション開発者が用意する必要があるが、その他はタスク規定ファイル自動生成システムが作成する。以下で、この 5 つのファイルについて説明する。

データベース 検索の対象となるファイル。項目名と項目値のペアの集合であり、個々のデータ(ホテル検索なら 1 つのホテルのデータ、文献検索

なら1つの文献のデータ)は空行によって区切られている。表1にホテル検索用のデータベースの例を示す。

単語辞書 音声認識と意味解析に用いる単語辞書。語彙の一部はデータベースから自動抽出され、読みが付与された後、単語辞書に登録される。

音声認識用文法 正規文法で記述された、音声認識に用いる文法。検索項目に従い、文法自動生成部が作成する。

意味解析用文法 意味解析に用いる文法。フレーズの種類を特定するために用いられ、音声認識用文法のタグつきサブセットである。

検索項目設定 各検索項目の表示名称や検索タイプなどを記述したファイル。

表 1: データベースの例

| | |
|----------|-------------|
| 名称 | *** ホテル |
| タイプ | ビジネス |
| 所在 | 三重県 津市 |
| 交通 | 近鉄津駅から徒歩で1分 |
| シングル料金下限 | 5000 |
| シングル料金上限 | 6600 |
| 立地 | 駅周辺 |
| 付帯施設 | レストラン |
| 付帯施設 | カフェ |
| 周辺レジャー | ゴルフ |
| 周辺レジャー | 海水浴 |
| 名称 | *** 旅館 |

4.2 検索項目設定用ツール

アプリケーション開発者は検索項目設定用ツールを用いて検索システムの仕様を指定する。

本システムで指定する項目にはシステム名称、データベースのファイル名、そして検索対象名がある。検索対象名は例えばホテル検索であれば「ホテル、宿、旅館」などであり、文献検索では「文献、資料、論文」などである。これらには自動的に読みが付与され、語彙に登録される。

その他に検索項目毎に、その検索項目が対応するデータベース項目名、検索タイプ、キーワードの選択方法、表示する発話パターンとその別名を指定する。以下でこの4つについて詳しく説明する。

検索項目が対応するデータベース項目名称 この検索項目が対応するデータベース項目を指定する。

検索タイプ 検索タイプは「含む、含む(追加型)、以上、以下」の4つから選択する。

キーワードの選択方法 キーワードの選択方法はデータベースから自動抽出する方法と、一般的な概念から選択する方法がある。データベースから自動抽出する場合は対応するデータベースの項目値を抽出し、単語に分解して語彙に登録する。一般的な概念には「金額」「月日」「時間」「年」が用意されている。

表示する発話パターンとその別名 画面上に表示する発話パターンを指定する。発話パターンは「データベース項目名称 + キーワード + 条件」であり、データベース項目名称と条件を指定することで、表示する発話パターンを決定する。また発話パターンには別名を登録することができる。この別名は画面上には表示されないが、ユーザの発話のゆらぎに対処するためのものである。

4.3 語彙自動生成部

検索項目設定用ツールにより指定された検索システムの仕様に従ってデータベースからキーワードを自動抽出し、読みを付与して語彙を設定する。また単語に自動的に読みを付与するが、この際に長母音化に関して「ou」を「o:」に変える等の読みの修正を行う。

システムにより作成される語彙は、大きく4つに分類できる。

1. 助詞などの一般的な単語
2. 検索項目設定用ツールで登録された単語
3. データベースから自動抽出されたキーワード
4. 一般的な概念の選択により、システムが登録したキーワード

助詞などの一般的な単語はシステム開発者の仕様に関わらず、登録される。また、検索対象名と表示発話パターン及びその別名として検索項目設定用ツールで登録された単語は読みを自動付与した後、語彙に登録される。

データベースからキーワードを自動抽出する場合、単語が分かち書きされているなら接続条件を保存し、それを反映した正規文法を記述する。例えば、京都府の後に続く単語は京都市、宇治市、舞鶴市…であり、京都市の後には左京区、右京区、中京区…が続く

といった構造を，データベースの地名データからから自動的に構築する．データベースの項目値が分かち書きされていない場合は形態素解析ツールを利用し，名詞を抜き出してキーワードとして利用する．本システムでは形態素解析ツールに JUMAN version 3.5[6] を用いている．分かち書きの有無に関わらず，読みは自動的に付与される．

キーワードの選択方法に一般的な概念からの選択が指定されている場合，選択された概念とその細かさ，上限，下限に従って単語を生成し，語彙に登録する．

4.4 文法自動生成部

検索システムの仕様に従って文法を自動生成する．文法は「いくつかのフレーズ + 文末表現」という構造になっており，フレーズ内文法は「データベース項目名称 + キーワード + 条件」が基本的構造である．フレーズ内文法は対応する検索項目の仕様によって，1つのフレーズに複数のキーワードを並べることが許される文法や，キーワードの構造を反映した文法となる．

4.5 語彙の読み確認・訂正ツール

アプリケーション開発者によって登録された語彙やデータベースから自動抽出された語彙には，形態素解析ツールを用いて自動的に読みが付与されるが，完全な読みの付与は期待できないので，人手による読みの確認・訂正が必要である．また，同じ表記の語彙に別の読みを付与することで，別名を登録することが可能である．例えば，「電子情報通信学会」に「でんしじょうほうつうしんがっかい」という読みとは別に「しんがくかい」という読みを付与することができる．

5 実装

本プラットフォームを Sun ワークステーション上で Perl/Tk を用いて実装した．また，関西のホテルデータベースと文献データベースを用い，ホテル検索システムと文献検索システムを実現した．図5にホテル検索システムの外観を，図6に文献検索システムの外観を示し，表2に両データベースと生成された語彙・文法のサイズを示す．また図7と図8に汎用的音声対話情報検索システムの画面写真とタスク規定ファ

イル自動生成システムの検索項目設定用ツールの画面写真を示す．

| ホテル検索システム | | | |
|-----------|---|---|-----|
| 所在が | [|] | |
| ホテルタイプが | [|] | |
| シングル料金が | [|] | 円以下 |
| ツイン料金が | [|] | 円以下 |
| 付帯施設に | [|] | がある |

図 5: ホテル検索システムの外観

| 文献検索システム | | | |
|----------|---|---|--------|
| タイトルに | [|] | を含む |
| 著者に | [|] | を含む |
| 掲載紙が | [|] | |
| 掲載年度が | [|] | 年より新しい |
| 掲載年度が | [|] | 年より古い |

図 6: 文献検索システムの外観

表 2: ホテル検索システムと文献検索システムの詳細

| 項目 | ホテル検索 | 文献検索 |
|------------|--------|--------|
| データベースのサイズ | 850 KB | 957 KB |
| データの件数 | 2040 | 6155 |
| 語彙数 | 715 | 7966 |
| 文法の行数 | 426 | 4565 |

6 おわりに

本稿では汎用的な情報検索音声対話プラットフォームの設計と実装について報告した．本プラットフォームは音声対話に必要なファイルを自動生成するタスク規定ファイル自動生成システムと，タスクの知識を持たない汎用的音声対話情報検索システムから成り，様々なタスクのデータベースに対して容易に適用が可能である．また発話パターンを画面上に明示するインタフェースを採用しているので，ユーザの発話の誘導が期待できる．認識結果は即座に画面上に反映され，ユーザは容易にシステムの状態を知ることができ，誤りからの復帰が容易である．また，本プラットフォー

ムを用いてホテル検索システムと文献検索システムを実現した。

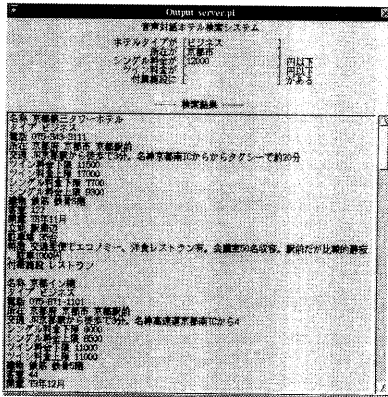


図 7: 汎用的音声対話情報検索システムの画面写真 (ホテル検索システム実行中)

参考文献

- [1] 新田恒雄, 神尾広幸, 雨宮美香, 松浦博, 内山ありさ, 田村正文: マルチモーダル UI とラビッドプロトタイピング, 情報処理学会研究報告, 95-SLP-7-5 (1995).
- [2] 竹林洋一: 音声自由対話システム TOSBURG II - ユーザ中心のマルチモーダルインタフェースの実現に向けて -, 電子情報通信学会論文誌, Vol. J77-DII, No. 8, pp. 1417-1428 (1994).
- [3] 李晃伸, 河原達也, 堂下修司: 文法カテゴリ対制約を用いた A* 探索に基づく大語彙連続音声認識パーザ, 情処学研報, 98-SLP-24-15 (1998).
- [4] 河原達也, 松本真治, 堂下修司: 単語対制約をヒューリスティックとする A* 探索に基づく会話音声認識, 電子情報通信学会論文誌, Vol. J77-DII, No. 1, pp. 1-8 (1994).
- [5] 河原達也, 李晃伸, 小林哲則, 武田一哉, 峯松信明, 伊藤克亘, 伊藤彰則, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏: 日本語ディクテーション基本ソフトウェア (97 年度版) の性能評価, 情処学研報, 98-SLP-21-10 (1998).
- [6] 黒橋禎夫, 長尾真: 日本語形態素解析システム JUMAN version 3.5 (1998).

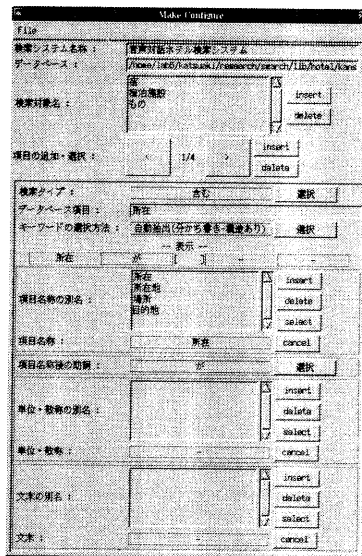


図 8: 検索項目設定用ツールの画面写真