

文献検索をタスクとした音声対話システムの応答生成

桐山 伸也 広瀬 啓吉
東京大学 工学部

学術文献検索を音声によって行なうシステムを構築した。システムの応答生成にあたって、適切な量の情報を応答文に含めること、対話の各時点で適度に確認や状況説明を交えること、TTSではなくCTSによって応答生成を行なうことの3点に主眼を置いて研究を進めた。システムのとる対話戦略として、システム主導寄り・ユーザ主導寄りといった複数の動作モードを設定するとともに、文献の著者名や掲載誌名等の情報を質問応答の形で引き出す機能において、情報の省略・補完の方式が異なる複数の応答形式を用意した上で、実際にシステムを稼働させて評価実験を行なった。

Speech Reply Generation of a Spoken Dialogue System for Literature Retrieval

Shinya KIRIYAMA Keikichi HIROSE
Faculty of Engineering, University of Tokyo

We propose a spoken dialogue system for literature retrieval, and discuss of speech reply generation. The system considers three points: appropriate amount of information included in a sentence, appropriate guidance from the system to users, and inappropriateness of using text-to-speech conversion devices for reply speech generation. In order to generate a wide variety of output sentences, they were converted from concept representations. The generated sentences were given as sequences of phone and prosodic labels so that they can be directly fed to the speech synthesizer. The system evaluated by comparing system-oriented and user-oriented dialogue modes, and also by determining appropriate ellipsis levels of output speech. The results indicate the version at the middle of two extremes was the best for the both cases, though they differ depending on factors such as whether the users having good experience with dialogue systems, or whether the system having a display.

1 はじめに

学術文献検索を音声によって行なうシステムを構築した。ユーザとの円滑な対話を進めていくにあたり、適度に確認や状況説明を交えながら、システムが対話の各時点で適切な応答をすることに主眼を置き、過去に提案されたシステム [1] を参考にしつつ研究を進めた。

応答文のシステム内部表現として文概念レベル・韻律句レベル・単語レベルの3段階の表現形式を用いることにより、概念からの応答生成を試みた。

システムの基本的な対話戦略と、文献情報を

質問応答形式で引き出す際の応答形式それぞれに複数のモードを用意し、これらの評価を実験を通して行なった。

2 システムの概要

2.1 システムの機能

本システムは、検索語によって文献検索を行なう機能、検索結果を表示し文献の概要を読み上げる機能、文献情報に関する質問に答える機能という3つの機能を持つ。

表 1: 使用文献データ一覧

学会誌名	論文数
日本音響学会誌	218
電子情報通信学会論文誌 D-I, D-II	640
情報処理学会論文誌	318
学術情報センター紀要	15
合計	1191

検索機能

タイトル・キーワード・アブストラクト・著者名・発行年・掲載誌名の6カテゴリに属する単語が検索語として使用できる。また、これら6カテゴリにまたがって検索する単語の属するカテゴリを独立に設け、合計7つのカテゴリによって検索を行なう。各カテゴリの内部及びカテゴリ間で、AND 検索と OR 検索を組み合わせて文献検索を行なうことができる。

結果出力機能

文献のタイトルだけを表示する簡略表示機能、著者名・掲載誌・概要をも表示する詳細表示機能に加え、TTS システムによって文献のアブストラクトを読み上げる機能を実装している。

質問応答機能

検索結果に対して、文献のタイトル・著者名・発行年・掲載誌名と、アブストラクトの有無の各項目についての質問に音声で応答する機能である。

2.2 文献データベース

本システムでは、学術情報センター電子図書館サービス [2] における文献データの一部を使用した。表 1 に使用データ一覧を示す。

2.3 システムの構成

本システムは、音声認識部・対話管理部・検索部・音声合成部の4つのモジュールからなる。図 1 に本システムの構成を示す。

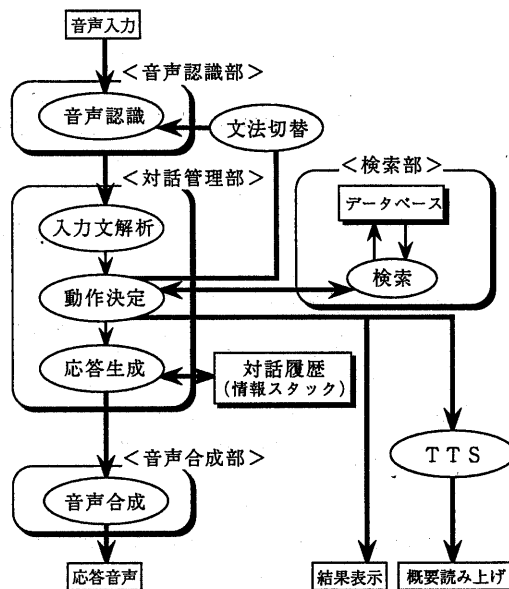


図 1: システムの構成

音声認識部

HTK[3]により作成した音声認識器を用いた。音韻モデルについては、情報処理振興事業協会 (IPA) の独創的情報技術育成事業の研究成果物である「日本語ディクテーション基本ソフトウェア 97 年度版」に含まれる triphone モデルを利用した [4]。

言語モデルについては、本タスクに適応した単語辞書並びに文法記述ファイルを作成して用いた。単語辞書については、対話データと文献データからそれぞれタスク依存語と検索語を抽出し、形態素解析ソフト ChaSen[5]によって読みを付加して作成した。文法については、1) 検索語入力用、2) コマンド入力用、3) 番号入力用、4) yes/no 入力用、の4つの文法を用意し対話の各時点で適宜選択して用いるようにした。

対話管理部

認識結果を解析して、システムの動作を決定し、応答を生成する。入力がコマンドならそれを実行する。詳しくは3節で述べる。

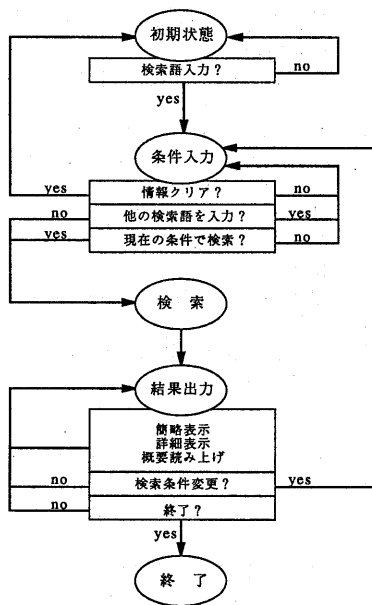


図 2: システムの状態遷移

検索部

入力検索語をカテゴリに分類して検索文字列を生成し、これに合致する文献を検索する。

音声合成部

対話管理部が生成した単音・韻律記号列を入力とし、ターミナルアナログ方式の音声合成器によって応答音声を出力する。

3 対話管理

3.1 対話管理方式

本システムでは、状態遷移表に基づく対話管理方式を採用した。図 2 に本システムの状態遷移図を示す。

ここで、条件入力状態から検索状態に移行するタイミングや、検索直後の検索結果の出力方法などは、一意に定まるものではなく、幾通りか方法が考えられる。そこで本システムでは、これらを制御する方式(以後、基本動作モードと呼ぶ)を複数用意し、選択的に用いられるようにした。

3.2 処理の流れ

この基本動作モードに基づく対話管理モジュールの処理の流れは以下のようになる。

1. 認識結果から検索語やコマンド等の情報を抽出して、入力概念を得る。
2. 現在のシステムの状態と入力情報、基本動作モードから状態フラグを決定する。状態遷移表を参照してシステムの動作を決定し、応答文概念情報を得る。
3. 決定されたシステムの動作に基づき、結果表示・概要読み上げといったコマンドを実行する。
4. 応答文の概念表現を入力として、応答生成プロセスを起動する。
5. 次状態において想定される入力を決定して音声認識部の認識文法を切替え、音声認識器を入力待機状態にして状態遷移する。

3.3 基本動作モード

本システムで設定した基本動作モードは、ユーザ主導寄り・システム主導寄り・両者の中間という 3 つである。

ユーザ主導

システム側の発話は、通常、初期状態と終了状態における挨拶と該当件数の通知だけである。状態遷移表に記述のない入力に際しては再入力を促す。基本的にユーザの入力によってのみ動作するモードである。

システム主導

ユーザの入力に対して極力確認を行ってから実際の動作に移る。検索結果やシステムの状態に応じて、その時点でふさわしいと判断されるコマンドをユーザに勧めるといった動作をする。基本的にユーザはシステムからの質問に答える形になる。

両者の中間

検索語入力から検索への移行時には、その旨をユーザに確認する。検索後、結果に対するコマンドを実行した際には、その旨をユーザに通知する。システムとユーザがいわば対等の状態で対話が進むモードである。

3.4 質問応答時の応答形式

質問文は、「3番の掲載誌名は何ですか」というように、表示されている文献一覧の番号と質問項目を入力する形式をとる。

質問文から番号と質問項目を抽出する際、質問文中にこれらの情報がない場合には対話履歴を参照して情報を補完する。対話履歴を参照しても情報が得られなかった場合には、不足している情報をユーザに問い返す。

本システムでは、質問に対する応答の方法として以下の3通りを用意しており、これらを選択的に用いることができる。

1. 質問に対する答のみを回答する
2. 質問文中に含まれなかった要素を補って答える
3. 全ての要素を補って答える

3.5 対話例

図3に、基本動作モードが両者の中間、質問応答形式が質問文中に含まれなかった要素を補って答える、というモードにおける対話例を示す。

4 応答生成

4.1 システム内部表現

本システムにおける応答文のシステム内部表現としては、文概念・韻律句コード列・単語コード列の3段階の表現を用いることとした。

これら3表現間の変換に際して、韻律情報(=品詞・アクセントの型と位置・モーラ数)と単語の見出し(=単語そのもの)を記述した

- S-1: どのような論文をお探ですか?
U-1: 音声認識
S-2: 現在の条件で検索しますか?
U-2: はい
S-3: では検索します。
検索した結果24件が該当します。
もう少し条件を絞りますか?
U-3: はい
S-4: 条件を入力して下さい。
U-4: HMM
S-5: では現在の条件で検索しますか?
U-5: はい
S-6: 検索した結果以下の8件が該当します。
どれかについてアブストラクトを表示しますか?
U-6: はい
S-7: 何番のアブストラクトを表示しますか?
U-7: 1番
S-8: 1番のアブストラクトを表示します。
U-8: タイトルを表示して下さい
S-9: 以下にタイトルを表示します。
U-9: 2番の著者名は何ですか
S-10: 池田思朗です。
U-10: 3番は誰ですか
S-11: 著者は中川聖一 鶴見豊です。
U-11: 雑誌名は何ですか
S-12: 3番は電子情報通信学会論文誌D2です。
U-12: もう結構です
S-13: 終了しますか?
U-13: はい
S-14: 御利用ありがとうございました。

図3: 対話例(両者の中間・省略要素を補完)

応答生成用単語辞書と、応答文に現れ得る韻律句を単語コード列ないし単語分類コード列としてコード化した韻律句辞書、そして状態遷移表が定める応答文概念を韻律句コード列として記述した文概念辞書の3つの辞書を用意した。ここで、単語分類コードとはキーワード・著者名・該当件数といった実行時に変動する単語を記述するためのコードである。

表2に本システムで用いている文概念の一覧を示す。

4.2 処理過程

本システムにおける応答生成の処理過程は以下のようになる。

表 2: 文概念表現一覧

文概念	文概念の意味
01	定型文
02	検索語入力要求
03	実行コマンド通知
04	実行コマンド確認
05	該当件数通知
06	指示・通達
07	質問応答

1. “検索語入力要求”, “該当件数の通知”, と いった文概念情報を入力として受けとる.
2. 各入力文概念に応じて情報スタックを参照し, 応答に含める付加情報を得て, 文概念辞書を検索する文概念コードを生成する.
3. 文概念辞書に従って韻律句コード列を得る.
4. 各韻律句コードに対して, 韻律句辞書から単語(分類)コード列を得, 情報スタックを参照して単語分類コードを単語コードに変換する.
5. 単語辞書から単語コードを単語の見出しに変換して表層文を得る.
6. 単語辞書から各単語の韻律情報を読み込む.
7. 文概念に応じて対応する韻律句に焦点情報を付加する.
8. 韻律句単位で文全体を走査し, 対話音声用の韻律規則 [6] に則って, フレーズ指令を然るべき位置に立てる.
9. フレーズ指令と焦点情報からアクセント指令を決定する.
10. 韻律句毎の音韻記号列をつなぎ合わせ, 文全体の音韻記号列を生成する.

図 4 に上述の処理を施す例を示す. これはユーザからの「著者名は何ですか」との質問

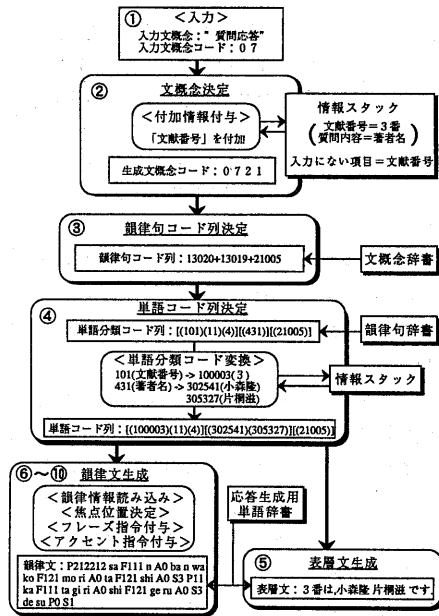


図 4: 応答生成の処理過程

表 3: 評価実験における対話モードの対応表

対話	基本動作	質問応答
1 回目	システム主導	全ての要素を補完
2 回目	両者の中間	省略要素を補完
3 回目	ユーザ主導	回答情報のみ

に対し, 文献番号を補って「3番は, 小森隆 片桐滋 です」という応答文を生成する過程を示したものである.

5 評価及び今後の課題

5.1 評価実験

構築したシステムを実際に稼働させて評価実験を行なった.

3通りの基本動作モードと, 質問応答機能における3通りの応答モードについて各々のモードに順位をつけてもらうと同時にシステムの使用感について感想を述べてもらった. 1人の被験者に表3のモードで3回の対話を試行してもらった.

評価結果を表4に示す.

表 4: 評価実験結果

評価項目	基本動作	質問応答
1位	両者の中間	省略補完
2位	システム主導	完全補完
3位	ユーザ主導	回答のみ

まず基本動作モードの結果は、1) 被験者は全員、本システムを初めて使用したために、ユーザ主導よりはシステム主導に近いほうが良いと判断した。2) システム主導モードが入力の際に確認をするのがややくどく、検索したりコマンドを実行するのに時間がかかる。3) ユーザ主導モードはシステムが基本的に応答を行なわないために入力した発話が理解されたかどうか不安で、結果出力の後、次にどうしたらよいかわからない。といったことに起因すると考えられる。

質問応答形式については、1) 回答情報だけを答えるモードでは入力した番号が正しく認識されているのか確認できない。2) 全てを補完するモードと入力にないモードでは、その違いが分かりにくく、また全ての要素を答えても応答がくどいとは感じられない。といったことが考察される。

5.2 今後の課題

まず、音声認識の頑健性が上げられる。今回の実験での認識性能は、全体を通して文理解率が71.1%、単語認識率が68.8%であった。結果を分析したところ、文献番号とキーワードの認識誤りが多くあったことが分かった。

今回のシステムでは認識誤りから対話が破綻することは回避できていたが、ディスプレイなしの環境下や、助詞等からの意味抽出などより知的な処理を行なっていく場合には、さらなる対処法が必要となる。

もう一つの課題はシステムの機能拡張である。評価実験に際して、現システムの機能の乏しさを訴える意見が少なからずあったことから、今後はより知的な処理を行なえるシステムに拡張していくことが望まれる。

例えば、文献のアブストラクトの要約を説明したり、ユーザに応じてシステムが内部的に基本動作モードを切替えたりといった機能が考えられる。

6 まとめ

本稿では、応答生成に主眼をおいて構築した学術情報検索をタスクとする音声対話システムについて報告した。

応答生成に関して、3段階のシステム内部表現を用いて概念からの応答生成を行ない、対話の流れに即して適量の情報と適切な韻律を持った応答音声の出力を実現した。

構築したシステムを実際に稼働させて、システム主導とユーザ主導の中間的な動作モードが良い、質問応答形式についてはある程度情報を補完した応答が良い、との結果を得た。

評価実験を通して、音声認識部の頑健性とシステムの持つ機能の拡張が課題として現れた。

今後はこうした課題を克服すべく、より知的なシステムへと改良を進めていく。

参考文献

- [1] 鈴木敏克, 小杉康宏, 妹尾徹, 広瀬啓吉: “巡視者支援対話型システムの検討,” 94-SLP-2-5, pp.31-34 (1994.7).
- [2] <http://els.nacsis.ac.jp/>
- [3] S.Young, J.Jansen, J.Odell, D.Ollason, P.Woodland: “The HTK Book,” v2.0, Cambridge University, (1995).
- [4] T. Kawahara et al.: “Japanese Dictation Toolkit – 1997 version –,” J. Acoust. Soc. Jpn. (E), 20-3, pp. 233-239, (1999).
- [5] <http://cl.aistnara.ac.jp/lab/nlt/chasen.html>
- [6] 広瀬啓吉, 川波弘道, 阪田真弓: “音声対話システムの出力音声の韻律的特徴の合成,” 人工知能学会全国大会 (第10回) 論文集, 15-01, pp.399-402 (1996.6).