

エージェントを利用した映像検索のための ユーザーインターフェイス

住吉英樹 望月祐一 金淵培 柴田正啓 井上誠喜

NHK 放送技術研究所

〒157 東京都世田谷区砧 1-10-11

あらし

我々は、放送局が持つ大量の映像を利用した、新しいコンテンツサービスとして教育分野への応用などを検討している。映像検索システムを中心とする教育システムでは、従来の単純なキーワードによる検索ツールだけでなく、その映像に関する各種の付加情報を効果的に提示できるユーザーインターフェイスが必要になる。

今回、学習者が問い合わせや質問、検索といった作業を自然言語による対話とCGキャラクタによるプレゼンテーションを行いながら、映像や付随する各種の知識の習得を目的とするユーザーインターフェイスシステムを試作した。本報告では、試作したシステムで使用する、対話形ユーザーインターフェイスの概要と自然言語対話処理について述べる。

キーワード エージェント、映像検索、ユーザーインターフェイス、自然言語対話

A User Interface for Video Retrieval System by Using Agent Technique.

H. Sumiyoshi Y. Motiduki Y. B. Kim M. Shibata S. Inoue

NHK Science and Technical Research Laboratories

1-10-11 Kinuta, Setagaya-Ku, Tokyo, 157

Abstract Our investigations include a study of applications of the system in the educational field, as a new contents service using the huge volume of video images available at broadcast stations. In an educational system centered around an video search system, it is necessary to incorporate not only traditional search tools based simply on key words, but also a user interface that can effectively furnish a variety of additional information related to the video images. We have developed a prototype user interface system, the goal of which is to allow learners to use mainly natural language dialog to conduct such operations as inquiries, questions, and searches, and in doing so gain knowledge of videos and various related fields. In this paper, we will provide an outline of the dialog-based user interface and discuss natural language dialog processing.

key words Agent, Video retrieval, User interface, Natural language intaraction

1. はじめに

放送局、特にNHKでは過去の貴重な映像も含め、膨大な量の映像を蓄積しており、これらの映像をデータベース化し、再利用するためのアーカイブセンターの構築も計画されている。

一方で、2000年末から開始されるデジタル放送のデータチャンネルを用いれば、映像とともにそれを説明する文章情報や、画像、映像情報を付加して放送することができる。また、これらの情報を一括して蓄積するホームサーバーなどにより、再視聴も容易になる。さらに、将来的には、端末側に蓄積された情報だけでなく、ネットワークを通じて放送局側にある情報や映像なども検索できるようになるであろう。

我々は、このような技術を背景にした新しいコンテンツサービスを検討しており、その一形態として自発的な調査、学習を支援する映像データベースを用いた学習アプリケーションについて試作を進めることにした。

ここで用いる映像データベースには、科学や歴史を中心とする映像情報の蓄積を行い、利用者が興味のある問題、対象物について自分の力で調査し、さらに調査結果を利用して映像作品に仕上げたり、発表したりする。このような作業過程を通じて学習を行うことを想定している。このときの映像データベースを利用方法を考えた場合、従来のような単純なキーワードインデックスによる検索ツールを提供するだけでは、あまり効果的に学習が進むとは思えない。また、検索時には映像だけでなく、その映像に関する各種の情報を提示することで学習効果を高められるのではないかと考える。

今回、上記の目的に適用できるユーザーインターフェイスとして、学習者が興味を持った対象に関する問い合わせや質問、映像検索といった作業を自然言語対話により行いながら、映像内容や付随する各種の知識を習得することを目的とするユーザーインターフェイスシステムを試作した。

本報告では、試作したシステムで使用する、自然言語対話処理と対話形ユーザーインターフェイスの概要について述べる。

2. 自然言語対話処理

2. 1 特徴

今回我々が構築した検索用ユーザーインターフェイスは、利用者との対話を行う、自然言語対話システムを中心技術として利用している。

コンピュータと人間との対話を目的とした自然言語処理システムに関しては、ELIZAなど古くから、多くの研究が行われてきた[1]。

しかし、その多くは構文解析、意味解析を行うもので、処理が複雑であるとともに、構文や意味の解析処理が書き言葉を中心としたモデルであるために、非文法的なあるいは、断片的な発話への対応が課題とされている[2]。

本システムの自然言語処理システムは、ELIZAのパターンマッチング手法を拡張したものである。会話文中の語の出現パターンをいくつかのメタキャラクターとともに指定し、入力会話文とのマッチングにより会話処理を進めるものである。また、バッファメモリーを使用した変数による条件制御が可能でより高度な会話パターンが構成できる。

これにより、話し言葉の特徴である非文法的な文章や、語の欠落、断片的な語による会話などがある程度許容でき、情報入力時、特に音声認識による質問の入力を行う際の非文法的な入力文に対する頑強性と早いレスポンスを特徴とする。

また、出現パターンが簡単なテンプレートにより記述できるので、会話データの記述も容易である。

2. 2 自然言語処理プログラム NLI

本システムで使用した自然言語対話処理プログラムは、FACTS (FIPA実証プロジェクト) で実験を行ったエージェントテレビ[3] (Audio-Visual Entertainment and Broadcasting:AVEB)で使用する対話用自然言語処理プログラム Natural Language Interface (以下NLI)を日本語化したシステムである。

このNLIの本来の目的は、エージェントTVにおいて、ユーザーとの対話を通して、ユーザーのプロファイル (名前や嗜好など) の獲得とシステムの制御を行うものである。今回は、このNLIを応用して映像検索対象および問い合わせ情報の抽出に利用した。

本NLIでの言語処理は、一般の対話処理システムと異なり、形態素解析レベルにとどめており、

抽出された形態素群と設定したパターンとのマッチングを取り、マッチした場合に決められた処理(返事など)を行う。

語の表面的な分析により検索語を抽出するアプローチは、高野[4]も行なっており、自然言語による利用者からの入力検索条件や要求を的確に表現していない場合に適するものと考えられる。

形態素レベルに分割した単語とそれらの出現パターンを抽象的に表現するメタキャラクターを含んだ文字列でテンプレートを定義し、利用者からの問い合わせ文とテンプレートが一致するかどうか比較し、一致した場合に記述されている動作を行う。このテンプレートを複数集めたものをテンプレート辞書ファイルと呼び、対話処理の知識情報に相当する。また、対話のテーマなどにより複数のテンプレート辞書を切り替えて使うことができる。さらに同義語辞書、単語辞書、変数定義ファイルなどを用いて対話処理を制御している。

以下に対話処理を行うためにNLIが用いる各ファイルについて説明する。

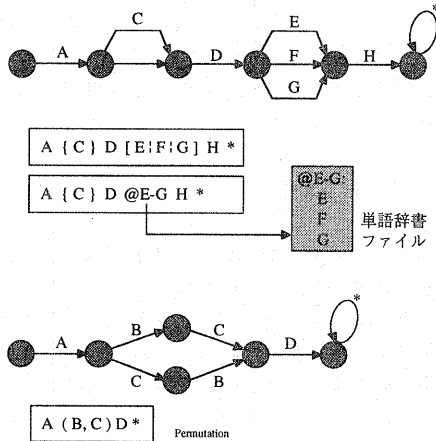


図1 テンプレートパターンの構成

1表 メタキャラクター

メタキャラクター	動作
*	0 ~ 無限個の単語に対応 (より多く)
%	0 ~ 無限個の単語に対応 (より少なく)
+	1個の単語に対応
!	一致してはいけない単語
{ }	{ }内で指定された単語, または存在しない
[]	[]内で指定された単語 (複数の場合は:で区切る)
()	()内の単語はどのような順序で出現してもよい

{}, [], ()のネストは不可

(1) テンプレート辞書ファイル

テンプレート内で用いられる規則要素には以下のものがある。下記の要素を利用して、図1に示すような構成が可能である。また、処理中に複数のテンプレート辞書ファイルを切り替えて使用できる。

(a) パターン

1表のメタキャラクターを組み合わせてテンプレートパターンを構成する。このテンプレート内で使用される語も形態素解析され、問い合わせ文の形態素解析後のパターンと比較が行われる。

(b) 変数

バッファメモリーの使用による変数。コンディションとアクションで使用できる。

var

(c) コンディション

条件式の左辺と右辺を比較、真であればアクションを実行。条件式には等号==と不等号!=使用できる。and, or を使用して複数の条件が記述できる。

c.abc==value

(d) アクション

左辺の変数に右辺の値を代入する。

a.abc = 値

特殊なアクション

goto 図2のように複数のテンプレート辞書ファイル間を移動可能。

control 他のプログラムからNLI処理系が呼びだされている場合、controlに代入された文字列をプログラムに送信する。

(e) 重要度

テンプレートパターンの重要度を数値で記述する。数値が大きいほど重要度が高く、同じテンプレ

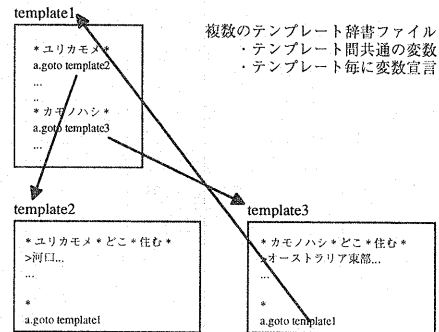


図2 gotoによるテンプレートの切り替え

レートでも優先的に一致する。記述されていなければ0として扱う。また、複数のパターンがマッチした場合、パターン長が長いほうが優先される。

w.7

(f) 返答文 (含む制御)

パターンが一致したときに返される文字列。システムが答える文章となる。また、後述の情報提示部に対する映像提示やCGキャラクターの動作制御コマンドを含めることができる。

> 回答文 <playVideo video1 100 200 400 200>

(2) 単語辞書ファイル

テンプレート内で多くの語を並列的に記述するのを防ぐため、語群を@WordFileNameで示す単語辞書ファイル内に下記のように記述することができる。

ファイル名: fishName

アユ

サケ

マス

テンプレート内では、その語群を代表するものとして@WordFileNameで指定する。

@fishName 食べられる*

(3) 同義語 (シノニム) 辞書ファイル

同じ意味を持つ複数の語の表記を統一するためのファイルで以下の構造を持つ。

アユ←あゆ:鮎:香魚

同義語処理は解析の最初に行われ、コロンで区切られた語のリストの2番目以降と同じものがある場合は、リスト先頭の語に置き換えられ言語処理が行われる

(4) 変数定義ファイル

テンプレート内で使用する変数を宣言するファイルである。

3. 検索用ユーザーインターフェイス

前記の自然言語処理プログラムを用いて、映像検索学習用のユーザーインターフェイスシステムを試作した。本システムの基本的な機能は、利用者から入力された文章を処理し、マッチしたテンプレートに記述されるアクション、具体的には、映像検索や情報の提示を行うものである。

例えば、「ユリカモメってどんな鳥?」という問い合わせ要求が入力された場合、「ユリカモメ」を検索キーワードとして映像と情報を検索し、その結果を利用者にWebコンテンツとして提示する。キーワードを抽出した後何を行うかは、自然言語処理部の返答文に定義できる。

実験システムとしては、図3に示すようにWebブラウザ、自然言語処理プログラム、Web Server、CGIプログラム、MS-Agentなどを組み合わせた構成とした。想定したコンテンツ群(映像、静止画など)は、ローカルWebサーバーに蓄積し、利用者からの要求に応じて蓄積された映像や情報の表示を行う。

3. 1 情報提示部

検索結果の動画、静止画、文字を表示する情報提示部にはWeb Browserを利用した。

基本的な表示データ(HTML,CGI)は、我々が研究を進めている番組制作システムDTPP(Desk-Top Program Production)[5]で使用する、番組情報(番組の構造や映像のキャプション、アナウンスコメン

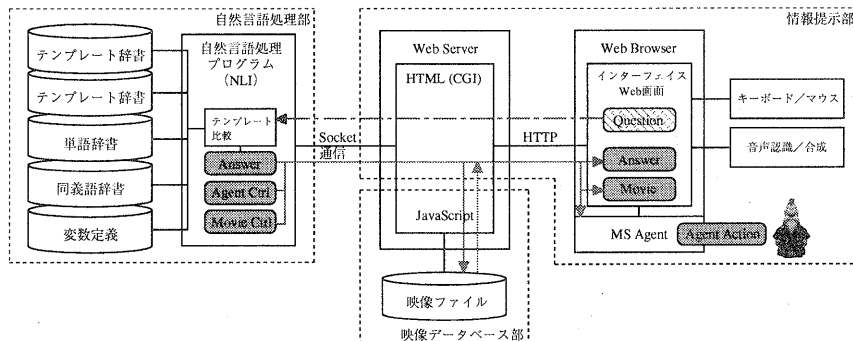


図3 システム構成図

トが構造化されて記述されている)を変換して利用している。DTPPでは番組を構造的に設計し、編集することができる。この番組情報をXML化し、図4のようなページを表示するHTMLデータに変換している。また、編集時に映像に付加されたキャプション情報を映像の説明文とするとともに、検索用データとして利用した。

情報提示部は、利用者から入力された問い合わせ文を自然言語処理部に送り、その結果の返答文を画面に表示するとともに、MS Agentと音声合成ソフトを制御して、返答文の発話、映像、画像情報のポインティングなどにより情報の提示を効果的に行っている。

MS Agentの制御は、CGIスクリプト中に特定のパターンのエージェント動作プログラムを組み込み、それを呼び出すコマンドを自然言語処理システムのテンプレート辞書ファイルに記述しておくことで、MS Agentが説明をすると同時に映像を指し示したりする。

3. 2 映像データベース部

映像表示データベースの構成方法は、さまざまな方法が考えられたが、今回は実験システムであり、検索条件の抽出と付属する情報の提示用ユーザーインターフェイスの実験が主目的であること

から、JavaScriptを利用したWebPageの表示制御機能を利用した。

Webページに埋め込まれた(画面に表示する)映像へのキャプションを検索用インデックスデータとし、自然言語処理部で検索キーワードとして抽出された語を含むかどうか比較し、含むもののみを検索結果として画面上に表示させる方法をとった。

4. 実験

実験的な対話型の映像検索アプリケーションを作成し、検索、提示実験を行った。実験に使用する辞書ファイル(テンプレート群)は会話を想定し、手入力したものである。

想定した会話は、興味のある映像を検索し、映像およびそれに関連する情報があれば、その対象に関する情報を対話的に提示する。具体的には川の生き物に関する映像の対話的な検索とその映像に関連する各種情報を説明的に提示するもので、対象が動物であれば、どこに住んでいるのか、何を食べるのか、などの情報である。

例えば、ユリカモメの映像を検索する場合。

検索の前後に検索対象となるユリカモメに関する情報を対話的に提示する。

想定する問い合わせは、以下のようなものであ



図4 ユーザーインターフェイス画面

(解説データは日立デジタル平凡社 DVD版世界大百科辞典より)

る。

- ・ユリカモメってどんな鳥？
- ・ユリカモメは、どこに住んでいるの？
- ・ユリカモメの映像が見たい。

提示する内容は、想定した問い合わせに対応する、以下のような情報を用意し、一度入力されたユリカモメという主題語を省略した、生息地や食物などの問い合わせにも応じられる実験用テンプレート群を作成した。

- ・名前：正式名
- ・科目：何科、何目
- ・特徴：形、色、大きさ
- ・生息地：地域、場所、環境
- ・食物：なにを

4. 1 実験用テンプレート

マッチしたテンプレートの想定する話題。例えば、「ユリカモメはどんな鳥」という質問を想定したテンプレート

*ユリカモメ*鳥*

に入力文がマッチした場合、この質問に答えると同時に、バッファメモリ subject に「ユリカモメ」を記憶させ、現在の話題を保持する。

連続して行われた質問が「どこに住んでいるのか」というような場合には、対象を話題となっている語であると想定し、話題語の変数である \$\$subject のチェックを行い、以下のテンプレートをマッチさせる

*どこ*住む*

c.\$\$subject==ユリカモメ

逆に、「どこに住んでいる」という入力があったときには、変数 question に「どこに住む」と設定し、「なにが？」という質問を利用者に返す。これに対する返事として、「ユリカモメ」を想定して、次のテンプレートを用意し、回答させる。

ユリカモメ

c.\$\$question == どこに住む

>ゆりかもめは〇〇に住んでいます。

このように話題を質問の主題と仮定し、会話文で起きやすい、主題の欠落に対応させる。また、語順の入れ替わりにも複数のテンプレートを登録することで対応する。他に話題が移った時には subject, question をクリアする。

4. 2 会話データ生成に向けた検討

会話データとなるテンプレート辞書は構成としては簡易なものであるが、まとまった会話をさせるためには、かなりの記述量になり、また、システム動作に関する知識も必要である。

この問題を解決する手段として、提示する情報を分類して所定の形式で入力することで、この情報に関する想定会話に対応できるテンプレート群を生成する実験を行った。

動物に関する映像データベースのような情報であれば、対象が持つ情報、問い合わせ文ともに限定できるので、これを利用して主語の省略や語順の変更などに対応できるテンプレート群を必要な情報を入力するだけで生成する。

現時点では、特定分野でしか使用できないが、用途の限られた映像データベースの情報を補足する対話処理には有効と考える。

今後は制作時に収集された各種の情報、映像、画像などから効率的に提示するデータの生成方法を検討する。

5. まとめ

実験システムにより、対話的な検索および情報提示を実現する基本的なユーザーインターフェイスが構築できた。現状のシステムは実験的なもので、映像データベースとの結合や会話システムの改良、提示方法の組み立てなどに課題が残る。特に、対話モデルが1問1答を基本とした構成であり、より複雑な会話を行うための会話モデルの構築や、検索履歴を用いて情報の提示方法を変更するなど、エージェントTVのフロントエンドを担えるように改良を進めていく予定である。

参考文献

- [1]Stuart C.Shapiro ほか、大須賀節雄監訳、「人工知能大辞典」、丸善、(1991)
- [2]熊本、「自然言語対話システムにおける協調応答の生成」、人工知能学会誌、Vol.14, No.1, pp.3-10, (1999)
- [3]村崎ほか、「FACTS プロジェクトにおけるエージェントテレビの研究開発」、電子情報通信学会技術報告、AI2000, (2000)
- [4]高野ほか、「自然言語を用いた対話形式による文書検索における辞典情報の利用」、電子情報通信学会技術報告、NLC2000-7, pp49-54, (2000)
- [5]住吉ほか、「階層化番組制作手法における番組情報構造」、電子情報通信学会総合大会、SMD-2, (1998)