

ハワイ・ボイス・ブラウザ  
— マルチモーダル情報サービスシステム

有田正剛 西村健士 島津秀雄  
日本電気(株)情報メディア研究所

関係データベースとハイパーメディアは、情報検索の様式が根本的に異なる。一方は集合ベースの検索、もう一方はインスタンスベースの探索である。そのため、従来、別々に研究開発されて来た。しかしながら、人間が情報を探す場合には、上に述べた両者の検索様式が混在している。我々の目的は、関係データベースとハイパーメディアを組み合わせ、集合ベースの検索とインスタンスベースの探索を同時に可能とする情報サービスシステムを構築することである。データモデルに関しては、拡張 ER モデルによって大局的な情報を表現し、関係モデルによって局所的な情報を表現する。また、両者の検索様式の違いを吸収するために、音声対話を中心としたマルチモーダルなインタフェースを採用し、特に、新しい対話モデルを採用する。ユーザのアクションのタイプを識別し、その結果に応じて適切なデータモデルを使用する。

Hawaii Voice Browser  
— Multimodal Information Service System

Seigo Arita Kenshi Nisimura Hideo Shimazu  
Information Technology Research Laboratories  
NEC Corporation

Nowadays there are, mainly, two types of information service systems(ISS), relational database system and hypermedia system. They has been researched and developed separately, because the manners to find informations are quite different. One is set-based retrieval and the other is instance-based navigation. But it seems that people have the flexibility in finding informations to use those two different manners simultaneously. In this paper, we aim to integrate those two types of ISSs and to supply users with cognitively more natural ISS. As to data model, we use an enhanced ER model to express global informations and use a relational data model to express local informations. To absorb the gap between the manners in finding informations, we employ a multimodal interface with a novel dialogue model. The dialogue model recognizes user's action types and choose a suitable data model.

## 1 はじめに

情報サービスシステムには、関係データベースによる情報検索とハイパーメディアによる情報探索の二つの流れがある。関係データベースによる情報検索においては、ユーザは欲するデータの条件を追加して、検索対象を段階的に絞り込んで行く。一方、ハイパーメディアによる情報探索においては、リンクとノードから構成された情報空間を渡り歩いて行くことで、欲するデータを探す。一言でいえば、一方は集合ベースの検索、もう一方はインスタンスベースの探索である。このように、関係データベースとハイパーメディアとは、情報検索の様式が根本的に異なるため、従来、別々に研究開発されて来た。

しかしながら、人間が情報を探す場合には、上に述べた両者の検索様式が混在している。例えば、ツアー情報の検索では、価格や出発日に関しては、条件を満たすツアーデータの絞り込みが行なわれ、同時に目的地周辺の観光スポットの探索は、情報空間の渡り歩きである。

我々の目的は、関係データベースとハイパーメディアを組み合せ、集合ベースの検索とインスタンスベースの探索を同時に可能とする情報サービスシステムを構築することである。ハイパーメディアによって情報空間を探索し、何か気に入ったデータがあればその詳細を関係データベースに問い合わせるシステムである。この実現には、両者の検索様式の違いを吸収するインタフェースの開発が鍵となる。我々は、音声対話を中心としたマルチモーダルなインタフェース [1][2] を採用し、特に、新しい対話モデルを採用することで、両者のギャップを埋めることを試みる。

試作システム「ハワイ・ボイス・ブラウザ」では、具体例として、ハワイツアー情報を取り上げた。ハワイ・ボイス・ブラウザでは、以下のようにして、ハワイの名所やホテルを探索しながら、ツアー情報の検索を行う。

### 1. ナビゲート

ユーザは、リンクを辿ったり、音声言語によって直接言及したりして、ハワイの名所やホテル等をあちらこちらと見て回る。

### 2. リトリバル

気に入ったものがあれば、その詳細を音声言語で問い合わせたり、類似したデータを検索したりする。

### 3. リポート

詳細を確認した結果、それが気に入れば、その旨システムに伝える。

リポートされた情報が十分集まれば、システムが主体的にユーザの希望に沿ったツアーを提案する。ハワイ・ボイス・ブラウザは、ハイパーメディアと関係データベースを組み合せ、その検索様式の違いを新しい対話モデルによって吸収することで、人間にとってより自然な情報検索を実現している。

2節では、ハワイ・ボイス・ブラウザの動作とシステム構成を見る。3節では、ハワイ・ボイス・ブラウザのデータモ

デルを見る。とくに、ハイパーメディアと関係データベースがどのように組み合わせられているかを見る。4、5節では、ハワイ・ボイス・ブラウザの対話モデルを説明する。

## 2 ハワイ・ボイス・ブラウザの動作とシステム構成

ハワイ・ボイス・ブラウザは、関係データベースとハイパーメディアを用い、それらに統一したマルチモーダルなインタフェースを提供する。これにより、集合ベースの検索とインスタンスベースの探索とを組み合わせた情報サービスを提供することができる。図1にハワイ・ボイス・ブラウザの動作例を示す。

ユーザが「海辺でゴルフ楽しみたい。」と発話すると、ハワイ・ボイス・ブラウザは、ワイレアゴルフクラブを示している。ここで、ハワイ・ボイス・ブラウザは、海辺にあるゴルフ場の中から任意の一つのゴルフ場を選択し、応答した。それに対して、ユーザは「オアフ島がいい。」と応答している。システムがたまたま選んだワイレアゴルフクラブはマウイ島にあるが、ユーザはこれが気に入らなかったわけである。そこで、システムは、類似データとして、オアフ島にある海辺のゴルフ場であるリンクスを示している。次に、ユーザはオプションツアーのボタンをクリックした。システムはリンクスゴルフ場の利用できるゴルフプランとしてオアフ島ゴルフプランを提示する。さらに、ユーザが「いくら」と発話すると、システムは関係データベースを検索し、「110ドルからです。」と合成音声で応答している。最後にユーザは「気に入った。」と発話し、オアフ島ゴルフプランを登録した。この情報は後にメインツアーの検索に用いられる。

このように、ハワイ・ボイス・ブラウザでは、通常のハイパーメディアのようにリンクを辿って関連情報を検索することができるほか、リンクを辿ることなしに、音声言語によって直接に適切なノードに飛ぶことができる。また、ディスプレイ上に表示しきれない情報を音声言語によって問い合わせたり、現在のノードに対応するデータと類似したデータを検索することができる。

図2はハワイ・ボイス・ブラウザのシステム構成である。端末で入力された音声は音声認識装置により単語列に変換されて、またポインティング情報はポイントされたボタンの識別子に変換されてホスト部に送られる。音声認識方式は不特定話者雑音単語認識であり、音声認識装置 DS-1000 によって行う [3]。ユーザは単語ごとに 0.3 秒程度のポーズをとる必要がある。

ホスト部では、対話管理モジュールが発話解釈モジュール、データ管理モジュール、応答生成モジュールを管理し、ユーザに対して対話的に情報提供する。対話管理部は、ユーザの入力が発話のときには、発話解釈モジュールを使用し、発話の言及しているハイパーメディア上のノードや関係データベース中の属性を求める。ユーザの入力がポインティングのときには、直ちにハイパーメディア上のノードが得られる。次に、



データ管理モジュールを使用し、求めるデータを得る。得られたデータの端末部での表現形式は、応答生成モジュールが予め用意されたテンプレートを使用して生成する。表現形式には、画像の表示位置や、ディスプレイ上に表示されるテキスト情報とその表示位置、そして音声合成の対象となるテキスト情報などが含まれる。音声合成は、NECが開発した、NEC版 Windows Sound System と同じ音声合成方式を用いている [4]。

また、対話管理モジュールには、ユーザが履歴したノードの履歴が保存されている。発話解釈モジュールはこれを参照することで、ユーザの発話を過去の発話と関連づけて解釈する。

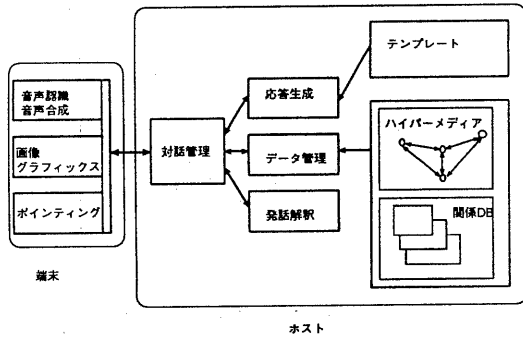


図2: ハワイ・ボイス・ブラウザのシステム構成

### 3 データモデル

ハワイ・ボイス・ブラウザでは、ハワイツアーに関わる情報をハイパーメディアと関係データベースによって表現する。ハワイツアーに関連するオブジェクトの概略やその間の関係をハイパーメディアで表現し、個々のオブジェクトの詳細は関係データベース中のテーブルあるいはビューとして表現する。

#### 3.1 ハイパーメディアによる表現

ハイパーメディアのデータモデルとしては拡張 E-R モデルを採用する [5]。拡張 E-R モデルは、Entity(実体)、Relationship(関連)、Attribute(属性) からなる E-R モデルに加えて、is-a、is-part-of の関係を考慮するモデルである。後に述べるように、is-a、is-part-of の 2 つの関係をを用いて発話の解釈を行なう。

図3にハワイ・ボイス・ブラウザの、拡張 E-R モデルで表現したハワイツアー情報を示す。各ノードには、タイトル、説明画像、音声出力用メッセージ、キーワード、そしてデータベース参照という5つの属性がある。タイトル、説明画像、

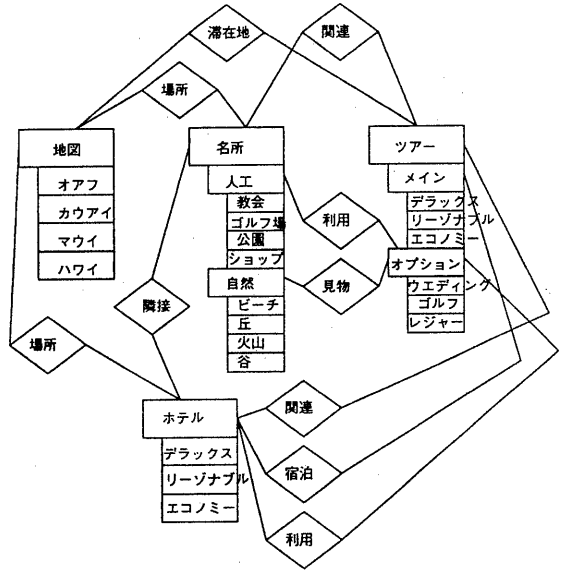


図3: 拡張 E-R モデルによるハワイツアー情報

音声出力用メッセージはノードの表現に用いる。それらのレイアウトの詳細はテンプレートとして用意しておく(図2)。キーワード属性は、値として文字列をもち、後に述べるように、発話の解釈に用いる。データベース参照は、関係データベース中の、対応するテーブル ID とレコード ID を値にもつ。データベース参照 ID を通じてハイパーメディアと関係データベースを関連づける。関係データベースはオブジェクトの詳細情報の問い合わせなどに用いる。

#### 3.2 関係データベースによる表現

関係データベースには、各オブジェクトの詳細情報が格納されている。例えば、図3のメインツアーノードはツアーテーブルに対応し、メインツアーのインスタンスに対応するノードはツアーテーブルのレコードに対応する。ツアーテーブルには、ツアーの値段、出発日、航空会社、日数などのツアーの詳細を表現する属性がある。

#### 4 ハイパーメディアへの音声言語による問い合わせ

ここではハイパーメディアへの音声言語による問い合わせの解釈を説明する。音声言語によって、リンクを辿ることなしに、直接に適切なノードに飛ぶことが目的である。したがって、ここでいうハイパーメディアへの音声言語による問い合

わけの解釈とは、問い合わせが指示するノードの集合を求めることである。

ハイパーメディアへの問い合わせの解釈は以下のように行う。

1. 問い合わせが言及しているノードを全て求める。  
問い合わせが言及しているノードとは、そのキーワード属性の値となっている単語が問い合わせに含まれているノードである。
2. 上で求めたノード  $N_1, \dots, N_m$  のセンタリング・ノード  $CN$  を全て求める。  
ノード  $N_1, \dots, N_m$  のセンタリング・ノード  $CN$  とは、次式をみたすノードである。

$$CN (is-a)^*(is-part-of)?(is-a)^* Ni (\forall i)$$

ここで、 $(is-a)^*$  は  $is-a$  の 0 回以上の繰り返し、 $(is-part-of)?$  は  $is-part-of$  の 0 回または 1 回の適用を示す。

3. 上で求めたセンタリング・ノードの集合を解釈結果とする。

例えば、「海辺の教会で結婚式したい。」という発話のセンタリング・ノードとして、図 4 に示すように、カルバリ教会がある。

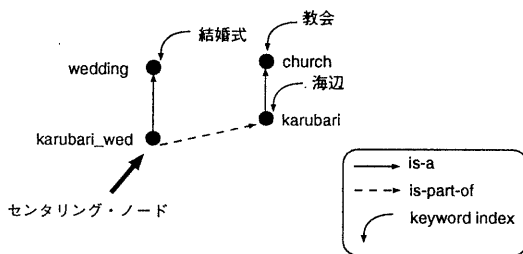


図 4: センタリング・ノード

## 5 典型例ベースの対話

関係データベースを対象とした集合ベースの検索とハイパーメディアを対象としたインスタンスベースの探索とを融合するために典型例ベースの対話モデルを導入する。典型例ベースの対話モデルは、関係データベースの検索における対話とハイパーメディアの探索における対話を融合したものともみることができる(図 5)。

典型例ベースの対話モデルでは、システムとユーザとの対話は、センターとその近傍であるリージョンを基に行われる。リージョンとは、ユーザの発話内容に合致する全てのノードの集合であり、センターとは、リージョンから選択した一つ

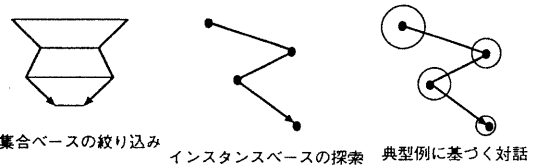


図 5: 関係データベースにおける対話、ハイパーメディアにおける対話および典型例ベースの対話モデルの比較

のノードで、ハイパーメディア上のカレントノードに一致する。ユーザは音声言語によって、リージョンを絞り込んだり、センターに対する修正要求を行うことでリージョン内を探索する。もちろん、リンクを辿ることで直接にセンターを変更することもできる。

### 5.1 発話タイプ

ユーザの発話タイプとして以下の 5 つを用意する。

**設定** 「海辺でゴルフ 楽しみたい。」など、リージョンの設定に相当する発話。

**確認** 「いくら？」など、センターの詳細の確認に相当する発話。

**修正** 「オアフ島がいい。」など、センターの修正に相当する発話。

**却下** 「いやだ。」など、センターの選び直しの指示に相当する発話。

**了承** 「OK。」など、センターの承認に相当する発話。システムは、センターを登録する。

図 1 の発話とそのタイプは以下ようになる。

1. 「海辺でゴルフ 楽しみたい。」：設定
2. 「オアフ島がいい」：修正
3. 「ゴルフプラン」ボタンのクリック：設定
4. 「いくら」：確認
5. 「気に入った。」：了承

システムは発話タイプに応じて、関係データベースとハイパーメディアを使い分け、センターとリージョンを操作する。

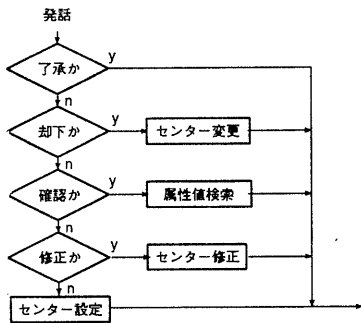


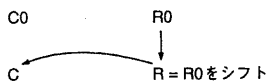
図 6: 発話タイプの判定とシステムの応答

## 5.2 発話タイプの判定とシステムの応答

発話タイプの判定とシステムの応答について述べる。図 6 にそのフローチャートを示す。対話の状態はセンター  $C$  とリージョン  $R$  で表される。ハイパーメディア上の各ノードは、データベース参照属性によって、関係データベース中のテーブルまたはレコードと対応づいていることに注意する (第 3 節参照)。

発話が了承を表すキーワード一語からなり、センターが既に設定されているときは、発話のタイプは了承と見なされ、センターが登録される (図 1 の 6)。これは後にメインツアのの検索に用いられる

発話が却下を表すキーワード一語からなり、センター  $C_0$  が既に設定されているときは、発話のタイプは却下と見なされ、リージョン  $R_0$  をシフト (第一要素を取り出し最後の要素として付加する) した  $R$  を新たなリージョン  $R$  とし、リージョン  $R$  の第一要素を新たなセンター  $C$  とする (図 7)。ユーザーが条件を追加せずにリージョン内のノードを順に見て回る場合がこのケースに相当する。



C:  $R$  の第一要素

図 7: センター変更

センター  $C_0$  が既に設定されていて、発話がセンターに対応するテーブルの属性名を表す語のみを含むとき、発話のタイプは確認と見なされ、センター (に対応するレコード) の対象属性の値を返す。センターの詳細の問い合わせである (図 1 の 5)。

発話がセンターに対応するテーブルの属性名と属性値を表す語のみを含むとき、発話のタイプは修正と見なされる。発話をセンターに対応するテーブルに対する検索命令として解釈し、データベースを検索し、結果  $T$  を得る [1][2]。例えば、センターがワイレアゴルフクラブで、発話が「オフ島がいい」のとき (図 1 の 2)、対応するゴルフ場テーブル中から島属性の値がオフである全てのレコードを検索し、それらに対応する全てのノードが  $T$  となる。新たなリージョン  $R$  を  $R = R_0 \cap T$  とし、 $R$  中の要素でセンター  $C_0$  と、修正対象の属性を除いて、最も類似した要素を新たなセンター  $C$  とする (図 8)。これにより、ユーザには島が変化する他はなるべく違うのないゴルフ場が示される。



C:  $R$  の要素のうち  $C_0$  に最も類似した要素

図 8: センター修正

以上でないとき、発話のタイプは設定と見なされる。発話は、ハイパーメディアに対する問い合わせとして解釈され、そのセンタリング・ノード (第 4 節参照) を全て求め、新たなリージョン  $R$  とし、 $R$  に含まれる任意の要素をセンター  $C$  とする (図 9)。ユーザーが新たに検索を始めた場合や検索対象を変更した場合 (図 1 の 1、3) がこのケースに相当する。



C:  $R$  に含まれる任意の要素

図 9: センター設定

## 6 おわりに

本論文では、関係データベースとハイパーメディアを組み合わせた情報サービスシステム「ハワイ・ボイス・ブラウザ」の説明を行なった。ハワイ・ボイス・ブラウザでは、音声言語やマウスポインティングによって、ハイパーメディア空間を探索し、何か気に入ったデータがあればその詳細を関係データベースに問い合わせることができるのを見た。

特に、典型例に基づいた対話を実現した。これにより、ハイパーメディアの探索とデータベースの検索とが自然に融合することができ、入力手段としての音声言語の表現力を生か

すことができた。人間にとってより自然な情報検索に一步近付いたと言える。

今後の課題としては、対話のシナリオを用意し、ユーザが何を言ったらいいのかわからなくなった時に、適切な誘導をすることがある。また、音声認識を単語単位から文節単位にグレードアップすることも検討中である。

## 謝辞

(株)JTB ハワイ殿には、ハワイの画像データやツアーデータを提供して頂きました。ここにお礼申し上げます。

## 参考文献

- [1] Arita,S., Shimazu,H., Takashima,Y.: "Simple + Robust = Pragmatic: A Natural Language Query Processing Model for Card-type Databases", Proc. of the 13th Annual Conference of the Cognitive Science Society, 1992.
- [2] 有田, 西村, 島津: "マルチモーダル自然言語インタフェース MM-SIMPLA", 第7回人工知能学会全国大会予稿集, pp.585-588, 1993.
- [3] 古賀, 吉田, 塚田, 渡辺, 岡村: "不特定話者連続音声認識用ハードウェアの開発", 電子情報通信学会研究会, SP91-116, pp.65-72, 1992.
- [4] 岩田ほか; "パソコン向けソフトウェア日本語テキスト音声合成", 日本音響学会講演論文集, 2-8-13, 1993-10.
- [5] Batini,C., Ceri,S., and Navathe,S.B. "Conceptual Database Design An Entity-Relationship Approach," The Benjamin/Cummings Publishing Company, 1992