

MMI 記述言語 XISL によるオンラインショッピングシステムの開発

小林剛典* 中村有作* 桂田浩一* 山田博文** 小林聡*** 新田恒雄*

*豊橋技術科学大学 大学院工学研究科 知識情報工学専攻

**豊橋技術科学大学 マルチメディアセンター

***豊橋技術科学大学 情報処理センター

〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘1 - 1

Email: kobayashi@vox.tutkie.tut.ac.jp

あらまし：本報告では、マルチモーダル対話 (MMI) 記述言語 XISL を用いて開発したオンラインショッピング (OLS) システムについて報告する。本システムは、ショッピングで用いるドキュメント (対話シナリオ (XISL), 顧客・商品データ (XML), 表示スタイル (XSL)) を格納するドキュメントサーバ, 対話進行を管理する対話制御部, 入出力端末を管理するフロントエンドの3つのモジュールから構成される。OLS システムは PC をフロントエンドとし、音声入力, ポインティング入力, 合成音声出力, 擬人化エージェント出力などを利用してショッピング対話を進める。一連の対話を XISL に記述することで、入出力の逐次的, 並列的, 択一的な制御, ページの遷移, 割り込みによるヘルプページの表示などの機能を実現することができた。

キーワード：マルチモーダル対話, XISL, オンラインショッピングシステム, XML

Development of an on-line shopping system by using multi-modal interaction description language XISL

Takanori KOBAYASHI*, Yusaku NAKAMURA*, Kouichi KATSURADA*
Hirobumi YAMADA **, Satoshi KOBAYASHI *** and Tsuneo NITTA*

*Graduate School of Technology, Toyohashi Univ. of Technology

** Multimedia Center, Toyohashi Univ. of Technology

*** Computer Center, Toyohashi Univ. of Technology

1-1 Hibarigaoka, Tempaku-cho, Toyohashi 441-8580, JAPAN

Email: kobayashi@vox.tutkie.tut.ac.jp

Abstract : This paper presents an on-line shopping (OLS) system developed by using multi-modal interaction description language XISL. The OLS system is composed of 3 modules: a front-end module that manages user's inputs and system's outputs, a dialogue-control module that controls dialogue flows, and a document-server module that holds dialogue scenarios (XISL), customers / items data (XML), and display style (XSL). The front-end of the OLS system is implemented on a PC terminal, and it allows users to interact by combining modalities such as speech recognition, speech synthesis, display and pointing devices. Using XISL for describing dialogue scenarios, we realized transition control to the other pages, interruption of dialogue to call help pages, and various types of usage of sequential, parallel and alternative inputs / outputs.

Key words : Multimodal Interaction, XISL, on-line shopping system, XML

1. はじめに

近年のマルチメディア技術の発展に伴い、マルチモーダル対話 (MMI) に関する研究が盛んに行なわれるようになった。W3C では MMI のワーキンググループ [1] が結成され、MMI 記述言語の標準化が進められている。また、SALT フォーラム [2] では、HTML の画面出力に音声インタフェースを付加するためのタグセット SALT が提案され、よりインタラクティブな Web サービスが実現されつつある。

こうした中、我々の研究グループでも MMI を記述する言語 XISL (Extensible Interaction-Sheet Language) [3][4]、およびその実行システムのアーキテクチャ (MMI アーキテクチャ) [5] を提案してきた。XISL は XML ベースの MMI 記述言語であり、対話シナリオの可読性が高く、入出力の制御や割り込み対話の記述が容易であるといった特徴を持つ。したがって XISL の利用により、システム作成者は MMI による Web サービスを容易に提供できる。

本稿では、XISL および MMI アーキテクチャを用いて開発したオンラインショッピング (OLS) システム [6] について報告する。OLS システムはドキュメントサーバ、フロントエンド、対話制御部の 3 つのモジュールから構成される。ドキュメントサーバには、OLS 用の対話シナリオ (XISL)、商品・顧客データ (XML)、表示スタイル (XSL) が個別に格納され、それぞれが再利用し易い構成になっている。フロントエンドは入出力インタフェースのモジュールであり、音声入出力や擬人化エージェントによる出力などを実行する。対話制御部ではショッピング対話の進行や MMI 入出力の統合を行なう。以上に述べたシステム上にある XISL にショッピング対話を記述することで、MM 入出力による商品注文の対話や、割り込み機能を利用したヘルプ対話を実現する。

以下では、まず始めに XISL を概観し、今回作成した OLS システムのアーキテクチャおよびシステムの特徴を述べる。その後、OLS システムで利用できる入出力モダリティについて述べ、OLS の対話の流れと XISL による対話制御の方法について説明する。

2. MMI 記述言語 XISL

XISL は、XML ベースの対話記述言語であり、一

```
<?xml version="1.0" encoding="Shift-JIS"?>
<!DOCTYPE xisl SYSTEM "xisl.dtd">
<xisl version="1.0">
  <head>... </head>
  <body>
    <dialog id="Top_Page">
      <exchange>
        <operation target="goods_list.xml">
          <input type="pointing" event="_l_click"
            match="//goods [goods='@name']/button[1]" />
        </operation>
        <action>
          <output type="speech" event="play">
            <![CDATA[
              <param name="speech_text">
                <value expr="goods">が選択されました。
              </param>
            ]]>
          </output>
        </action>
      </exchange>
    </dialog>
  </body>
</xisl>
```

図1 XISL の記述例

組の対話シナリオを表す dialog の集合から構成される。OLS システムでは、ユーザ ID の認証対話、商品閲覧対話などが一つの dialog に記述される。特定の dialog では、入力を常時待ち受けることも可能であり、この機能を用いて OLS システムのヘルプ対話を記述している。

各 dialog は、対話の最小単位である exchange の集合を内部に含み、各 exchange は、operation と action をそれぞれ一つずつ持つ。operation は入力記述部で、例えば OLS システムでは音声やポインティングの入力待ち受けが内部の input に記述される。一方、action はアクション記述部で、OLS システムでは合成音声や擬人化エージェントによる出力を内部の output に記述することができる。operation や action では、入力 (出力) の逐次的、並列的、択一的な受け付け (実行) を記述することができ、OLS システムではポインティングと音声を組み合わせた入力などを実現している。

図1に XISL の記述例を示す。この例では、ウィンドウ (ブラウザ) に表示されている商品のボタン (goods_list.xml に記述されたボタン要素) がクリックされたときに、システムが選択された商品名を音声で出力する対話記述を示している。

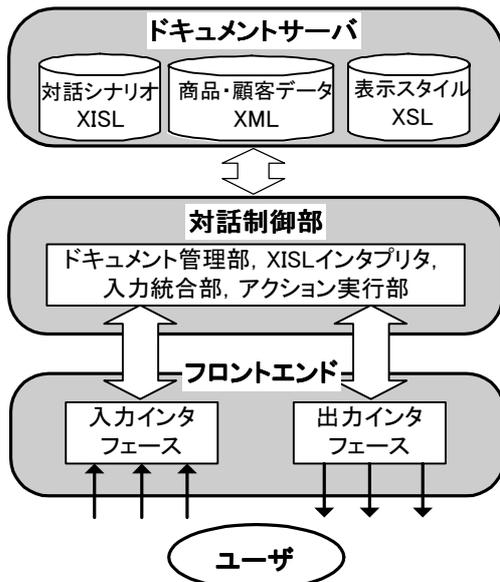


図2 OLSシステムのアーキテクチャ

3. OLSシステム

図2に、今回実装したOLSシステムのアーキテクチャを示す。システムは、フロントエンド、対話制御部、及びWWW上のドキュメントサーバから構成される。

フロントエンドは、ユーザからの入力を受け付ける入力インタフェースと、ユーザへの出力を行なう出力インタフェースからなる。なお、今回実装したシステムではフロントエンドとして、kiosk端末を模擬したPCを用いた。

対話制御部は、XISLインタプリタ、ドキュメント管理部、入力統合部、アクション実行部からなり、OLSの対話進行を管理する。

WWW上のドキュメントサーバには、OLSシステムのドキュメント（対話シナリオ（XISL）、商品・顧客データ（XML）、表示スタイル（XSL））が保持される。三つのドキュメント群をそれぞれ独立に取り扱えることで、商品在庫数や顧客情報の管理がし易く、データの保守性が向上するといった利点を持つ。また、対話シナリオや表示スタイルを独立に変更できることは、異なるユーザインタフェースを持つ端末からのシームレスなサービスを可能にする[5]。

4. OLSシステムのフロントエンド

今回実装したOLSシステムはPC端末上で動作するため、PCで一般的に用いられる入出力モダリティが利用できる。これに加えて本システムで

は、音声入出力や擬人化エージェントによる出力を可能にし、MMIによる操作を実現した。以下、フロントエンドの実装について述べる。

4.1 入力モダリティ

本システムでは入力モダリティとして、音声、ポインティング、キーボードを採用した。各モダリティの実装方法を以下に示す。なお、XISLにおける各モダリティの記述方法は表1を参照されたい。

- 音声

音声入力を処理する認識エンジンにはIBM® ViaVoice™8 Proを使用した。本システムでは認識エンジンが出力した書き下し文から、XISLで指定した文法中のキーワードをスポッティングすることでユーザの発話音声を特定している[7]。文法はW3Cで提唱されているSRGS（Speech Recognition Grammar Specification）[8]に基づいて記述している。
- ポインティング

マウスまたはタッチパネルを入力デバイスとし、左（右）クリック、左（右）ダブルクリック、ドラッグ&ドロップなどのイベントを受け付ける。
- キーボード

キーが押された時に発生するイベントを待ち受け、押されたキーの内容を取得する。本システムで取得可能なキーは、英文字（52文字）、数字（10文字）、記号（32文字）、制御文字（3文字）である。

4.2 出力モダリティ

出力モダリティとしては、音声合成、ウィンドウ、動画、オーディオ、擬人化エージェントを採用した。各モダリティの実装方法を以下に示す。XISLの記述方法については表2を参照されたい。

- 音声合成

音声出力を行なう合成エンジンには東芝LaLaVoice™2001を使用した。XISL記述により、出力の音量、声質、数字の読み方、読み上げ速度などを指定できる。
- 動画 / オーディオ

動画及びオーディオ出力には、Microsoft® MediaPlayer、RealNetworks™ RealPlayer™を使用した。使用するプレイヤーの選択は、XISLで出力指定されたファイルの

表1 XISL による入力モダリティの記述

入力モダリティ	type 属性	操作	event 属性	match 属性	return 属性
音声	"speech"	認識完了	"recognize"	文法規則のパス	1,2,...:文法規則と同名の変数の集合
ポインティング	"pointing"	クリック	左:"l_click", 右:"r_click"	入力対象となるXML オブジェクトのパス	1:XML オブジェクトのパス 2: イベントが発生したx座標 3: イベントが発生したy座標
		ダブルクリック	左:"ld_click" 右:"rd_click"		
			
キーボード	"keyboard"	キー入力	"press"	対象となるコンテンツ のルートノード	1:入力されたキー

表2 XISL による出力モダリティの記述

出力モダリティ	type 属性	操作	event 属性	CDATA セクション内で記述する<param>要素	
				<param>の name 属性	<param>の内容
音声	"speech"	音声出力	"play"	"speech-text"	読み上げテキスト
				volume	音量
			
擬人化エージェント	"agent"	バルーン出力	"balloon"	"agent-name"	操作対象のエージェント
				"balloon-text"	表示テキスト
		スピーチ出力	"speak"	"agent-name"	操作対象のエージェント
				"speech-text"	読み上げテキスト
...	

拡張子によって判断する。

- ウィンドウ
ウィンドウ(ブラウザ)出力には Microsoft® Internet Explorer6 の ActiveX コントロールを使用した。ウィンドウには、XSL により HTML に変換された XML コンテンツを表示する。また、通常のウィンドウの他にモジュールウィンドウも出力できる。
- 擬人化エージェント
擬人化エージェントのモジュールには Microsoft® Agent2.0 を使用した。擬人化エージェントは、スピーチ(合成音声)、バルーン(テキスト表示)、アクション(アニメーション)を用いた出力を行なうことができる。

5. OLS システムにおける MMI

5.1 対話の流れ

OLS システムでは、図3に示す流れに従って対話が進む。システムを起動した後、まず始めにユーザの ID 認証を行なう。認証が終了すると、商品閲覧の対話へ進む。この対話からは、詳細説明、購入回数決定、ショッピングカートの更新の各対話へと遷移することができる。購入する商品と個数が決定すると、個人情報記入の対話へと進む。

既に ID 認証に成功しているユーザの個人情報は、動的にフォームに入力される。個人情報記入後、ID を登録していないユーザは ID の新規登録を行なう。また、システム起動から終了までの間、割り込み対話であるヘルプを常に呼び出せる。以上の対話が終了するとシステムは終了する。

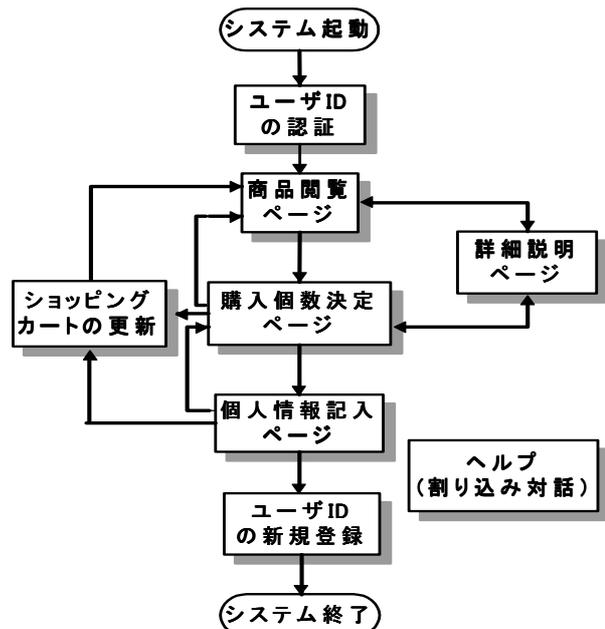


図3 OLS における対話の流れ



図4 商品閲覧ページ

5.2 XISL による対話の制御

ショッピング対話の起点となる商品閲覧ページの表示画面と、商品閲覧の対話をそれぞれ図4、図5に示す。また図6には商品閲覧ページ用のXMLコンテンツの一部を示す。

ユーザは、図4に示す商品閲覧ページ上で購入したい商品を選択する。図4-には購入できる商品が表示されており、この中から購入したい商品をポインティングまたは音声で選ぶことができる。ポインティング入力、XMLコンテンツの要素に対する入力とし、`<input>`の`match`属性で入力対象の要素を示すパスを指定する。音声による入力では、購入したい商品とその個数を同時に指定することができる。音声入力の対象となる語彙は、`<input>`の`match`属性で指定した文法ファイル(図7)の中に記述されている。

図4-はショッピングカートであり、ユーザの現在の購入状況を表示している。ショッピングカートの修正、または削除を選択すると、購入状況を変更できる。図4-は、図4-に表示されている商品を変更するためのカテゴリリストである。希望するカテゴリ名を選択することで、表示商品を変更できる。ポインティング入力または音声入力により商品が選択されると、購入個数決定ページへと遷移する。

以上に示した対話はすべてXISLにより制御される。XISLによる対話記述を図5に示す。

図5-(a),(b)を含む`<begin>`は、`<dialog>`が読み込まれて最初に行なわれる動作である。図5-(a)は、商品閲覧ページを開くという出力命令であり、図5-(b)は、擬人化エージェントにバルーンとスピーチによる出力を行なわせる命令

```

<?xml version="1.0" encoding="Shift_JIS" standalone="no"?>
<!DOCTYPE xisl SYSTEM "/xisl01.dtd">
<xisl ver="1.0" application="/root.xisl">
  <body>
    <dialog id="Top_Page" comb="all" repeat="">
      <begin>
        <- Window を開く ->
        <output type="window" event="open">
          <![CDATA[
            <param name="window-name">OLS_Ver1</param>
            <param name="uri">/top.xml</param>
          ]]>
        </output>
        <- エージェント出力: speech, balloon ->
        <output type="agent" event="speak-ballon">
          <![CDATA[
            <param name="agent-name">Wizard</param>
            <param name="speak-ballon-text">
              いらっしやいませ。ごゆっくり閲覧して下さい。
            </param>
          ]]>
        </output>
      </begin>
      <- 購入商品の選択 ->
      <exchange>
        <operation target="/top.xml" comb="all">
          <input type="pointing" event="l_click"
            match="/goods[goods=@name]/button[1]" />
          <input type="speech" event="recognize"
            match="/grammar.xml#buy_goods"
            return="goods" />
          <par_input>
            <input type="pointing" event="l_click"
              match="/goods[goods=@name]/picture" />
            <input type="speech" event="recognize"
              match="/grammar.xml#num_buy"
              return="number" />
          </par_input>
          <operation>
            <action>
              <- 音声合成出力 ->
              <output type="speech" event="play">
                <![CDATA[
                  <param name="speech-text">
                    <value expr="goods"/>の購入ページへ移ります。
                </param>
              ]]>
            </output>
            <- /Purchase.xisl へ遷移 ->
            <goto next="/Purchase.xisl" arg="$goods, $number" />
          </action>
        </exchange>
        ...
      </dialog>
    </body>
  </xisl>

```

図5 XISL による対話制御の記述

である。これらの動作が終了した状態を示したものが、図4の画面である。

図5-(c),(d),(e),(f)は図4-に表示されている画面に対する操作と、それに対するシステムの応答を示している。図5-(c)は各商品の

```

<category type="coat">
  <goods name="ロングコート" id="0001">
    <object>
      <picture id="main">long_coat.jpg</picture> ----- ( )
      <button id="buy">購入</button> ----- ( )
      <button id="explain">詳細説明</button>
      ...
    </object>
    <price value="25000">¥25000</price>
  </goods>
  ...
</category>

```

図6 XML コンテンツ (top.xml) の一部

```

$buy_goods = $goods [を] $buy; ----- ( )
$num_buy_02 = (($goods) [(これ) [を] $number $buy; ----- ( )
$number = (いちまい | いっこ | ひとつ){1} | ( にこ | ふたつ ){2} | ( さんこ | みっつ ){3} | ... ;
$buy = ( かい たい | こうにゆう | ください ) 購入 ;
$goods = ( $coat | $pants | $polo_shirt | $skritwear | ... );
$coat = ノーマルコート{0001} | ロングコート{0002} ... ;
...

```

図7 指定する文法ファイル (grammar.xml)

購入ボタンに対するポインティング入力の待ち受けを示している。図5 - (c) に示したポインティング入力の対象となるボタン要素は、図6 - () に示したXML コンテンツの要素に対応している。

図5 - (d) では、図7 - () に記述した語彙の音声入力を待ち受けている。図5 - (e) は、<par_input>で囲まれており、図4 - の各商品の画像要素 (図6 - ()) に対するポインティング入力と図7 - () 記述された語彙の音声入力とを並列的に待ち受けていることを示している。また、<operation>の属性 comb には属性値 alt が指定されており、図5 - (c), (d), (e) の入力は択一的で、どれか一つの入力があると、図5 - (f) が実行される。図5 - (f) は合成音声出力を示している。

図5 で示した対話の動作例を図8 に示す。まず始めに商品閲覧ページが開かれ、擬人化エージェントが「いらっしゃいませ...」と発話し、バルーンを表示する。ユーザが図4 - の中央に表示されているロングコートの画像をクリックしながら、「これを1つください」と発話した場合、図5 - (e) の入力が満たされ、合成音声「ロングコートの購入ページに移ります」が出力され、次の対話へと遷移する。

```

S: 商品閲覧ページ(top.xml)を開く
S: 擬人化エージェントによる出力 (音声+バルーン)
「いらっしゃいませ。ごゆっくり閲覧してください。」
U: 画像 (ロングコート) をクリック +
「これを一つください」と発話
S: 音声合成エンジンによる出力
「ロングコートの購入ページへ移ります。」
S: 次の対話(Purchase.xml)へ遷移

```

図8 対話の動作例

6. まとめ

MMI 記述言語 XISL を用いた OLS システムを開発した。本システムは、対話シナリオ、商品・顧客データ、表示スタイルを分割管理しており、それぞれのドキュメントを再利用し易いという特徴を持つ。また XISL を利用することで、ショッピング対話の記述、MM 入出力の制御およびヘルプ対話の割り込みなどを実現した。今回のシステム開発により、MMI 記述言語 XISL の優れた対話記述力を確認できた。

今後は PC 以外のフロントエンドに対して、同様の OLS システムを構築すると共に、マルチタスク、マルチドメインへの対応を検討したい。

本研究は電気通信普及財団の助成を受けて行われた。ここに深謝いたします。

参考文献

[1] <http://www.w3.org/2002/mmi/>
 [2] <http://www.saltforum.org/>
 [3] 小林聡, 中村有作, 桂田浩一, 山田博文, 新田恒雄: “マルチモーダル対話記述言語 XISL の提案”, 情報処理学会研究報告 2001-SLP-37, pp.43-48 (2001).
 [4] <http://www.vox.tutkie.tut.ac.jp/XISL/XISL.html>
 [5] 桂田浩一, 大谷佳彦, 中村有作, 小林聡, 山田博文, 新田恒雄: “多様な端末からのアクセスを可能にする MMI アーキテクチャ”, 情報処理学会研究報告書 2002-SLP-40, pp.51-56 (2002)
 [6] 小林剛典, 足立裕秋, 桂田浩一, 小林聡, 山田博文, 新田恒雄: “MMI 記述言語 XISL を用いたオンラインショッピングシステムの試作”, 情報処理学会第 64 回全国大会講演論文集 (分冊 4), pp.107-108 (2002)
 [7] 新田恒雄, 浅見弘道, 伊勢路真吾, 福田隆, 桂田浩一: “汎用 LVCSR を用いた対話音声の認識”, 情報処理学会研究報告 2002-SLP-41, pp.69-74 (2002).
 [8] <http://www.w3.org/TR/speech-grammar/>