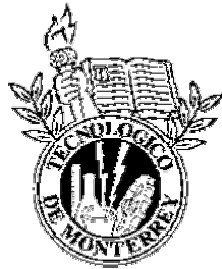


INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY



**TECNOLÓGICO
DE MONTERREY®**

Un Sistema Conversacional para Despliegue de Información Basado en Procesos de Decisión de Markov

Jaqueline Hernández García

**Sometido al Programa de Graduados en Informática y
Computación en cumplimiento parcial con los
requerimientos para obtener el grado de:**

Maestra en Ciencias de la Computación

Asesores:

Dr. Luis Enrique Sucar Succar

Dr. Oscar Arturo Mayora Ibarra

Cuernavaca, Morelos. Noviembre del 2005

Un Sistema Conversacional para Despliegue de Información Basado en Procesos de Decisión de Markov

Presentada por:

Jaqueline Hernández García

Aprobada por:

Dr. Luis Enrique Sucar Succar

Profesor del Departamento de Computación,
ITESM Campus Cuernavaca
Asesor de Tesis

Dr. Oscar Arturo Mayora Ibarra

Profesor del Departamento de Computación,
ITESM Campus Cuernavaca
Co-asesor

Dr. Eduardo Morales Manzanares

Profesor del Departamento de Computación,
ITESM Campus Cuernavaca
Sinodal

Dr. Luis Villaseñor Pineda

Laboratorio de Tecnologías de Lenguaje,
INAOE
Sinodal

Resumen

El trabajo que se presenta en esta tesis se ubica en el contexto de la Cátedra de Robótica Móvil del ITESM Campus Cuernavaca. Dentro de los proyectos de la cátedra, se contempla el desarrollo de un robot anfitrión que proporcione información a los visitantes sobre las diferentes áreas del Campus.

Para facilitar la interacción con el robot se desarrolló el Sistema Conversacional para Despliegue de Información (SCDI), el cual opera como una interfaz mediante la síntesis y reconocimiento de voz, y actúa como el medio de interacción entre los visitantes y el robot anfitrión. El sistema propuesto, permite a los visitantes solicitar información sobre sitios específicos del Campus y visualizar dicha información a manera de imágenes y videos en una pantalla proyectada por un cañón. Su desarrollo se basó en *Voice eXtensible Markup Language* (VoiceXML), uso de *Active Server Pages* (ASP) y el *Internet Information Services* (IIS) de Microsoft. El diseño del SCDI utiliza Procesos de Decisión de Markov (MDPs) para optimizar la ejecución de los diálogos y permitir acceder a la información en la secuencia más corta posible.

Se realizó un análisis de usabilidad comparativo entre la primera versión del sistema estructurada con menús jerárquicos y la versión del sistema optimizado utilizando MDPs. Las pruebas realizadas con usuarios demostraron que la aplicación de MDPs mejoró el desempeño del SCDI en comparación con la versión inicial. La interacción probó ser más eficiente y rápida al utilizar MDPs, según reflejaron los resultados del cuestionario aplicado a los usuarios.

Contenido

Capítulo I. Introducción	9
1.1. Antecedentes	9
1.2. Motivación.....	10
1.3. Planteamiento del Problema.....	10
1.4. Objetivos.....	11
1.4.1 Objetivo general	11
1.4.2 Objetivos específicos	11
1.5. Alcances y Limitaciones.....	11
1.6. Descripción General del Sistema.....	12
1.7. Organización del Documento.....	13
Capítulo II. Interacción Humano - Computadora	15
2.1 Introducción	15
2.2 Uso de Diálogos	16
2.2.1 Diálogo de bajo nivel	16
2.2.2 Diálogo no verbal	16
2.2.3 Diálogo mediante lenguaje natural	17
2.3 Sistemas Basados en Voz.....	18
2.3.1 Sistemas para el desarrollo de aplicaciones de voz.....	18
2.3.2 Aplicaciones de voz.....	19
2.3.3 Kioscos informativos	20
2.4 Robots Guías y Asistentes.....	23
2.4.1 Percepción humana de los robots sociales	29
2.4.2 Actitudes hacia robots	30
2.4.3 Efectos de la apariencia y el diálogo	30
2.5 Sistemas para Despliegue de Información	31
2.5.1 Museos virtuales	32
2.5.2 Museos interactivos y PDAs.....	34
2.6 Resumen del Capítulo	38
Capítulo III. Procesos de Decisión de Markov.....	39
3.1 Definición	39
3.1.1 Métodos de solución	40
3.2 Aplicación de MDPs en Diálogos.....	43
3.3 Resumen del Capítulo	48

Capítulo IV. Descripción del Sistema Conversacional para Despliegue de Información.....	49
4.1 Sistema Conversacional para Despliegue de Información.....	49
4.1.1 Diseño del sistema	49
4.1.2 Implementación	51
4.2 Inclusión de un Avatar	55
4.2.1 Comunicación entre aplicaciones.....	55
4.2.2 Jessica: Un rostro animado.....	56
4.2.3 Ajustes en el rostro animado para las enunciaciones del SCDI	57
4.2.4 Resultados del estudio de usabilidad	57
4.3 Resumen del Capítulo	59
Capítulo V. Aplicación de MDPs al Sistema Conversacional para Despliegue de Información.....	60
5.1 Aplicando MDPs	60
5.1.1 Modelado del MDP de Edificio principal	61
5.1.2 Modelado del MDP de Aulas	64
5.1.3 Modelado del MDP de Oficinas	65
5.1.4 Modelado del MDP de Áreas deportivas	65
5.2 Implementación de MDPs.....	66
5.3 Resumen del Capítulo	66
Capítulo VI. Experimentos y Resultados	68
6.1. Proceso de Experimentación	68
6.2. Resultados	70
6.3 Observaciones	84
6.4. Análisis	85
6.5. Resumen del Capítulo	87
Capítulo VII. Conclusiones y Trabajo Futuro.....	88
7.1. Conclusiones	88
7.2. Trabajo Futuro	90
Referencias	92
Anexo A. Modelo de los MDPs.....	100
Anexo B. Diálogos del SCDI.....	105
Anexo C. Diseño e Implementación del SCDI	111

Índice de Figuras

Figura 1.1 Diagrama de bloques de SCDI con MDPs.	13
Figura 2.1 Interfaz gráfica de MASK.	21
Figura 2.2 El kiosco MATCH.	22
Figura 2.3 Interfaz gráfica de MASK.	22
Figura 2.4 Robots de entretenimiento usados en el Museo para la Comunicación de Berlín.	23
Figura 2.5 RHINO interactuando con el público.	24
Figura 2.6 El robot Minerva.	25
Figura 2.7 Estados de ánimo de Minerva.	26
Figura 2.8 El robot Sage.	27
Figura 2.9 El robot CERO.	28
Figura 2.10 Personaje montado en CERO.	28
Figura 2.11 La teoría del Uncanny valley.	31
Figura 2.12 Exposición sobre la guerra civil norteamericana.	32
Figura 2.13 Museum system.	33
Figura 2.14 Virtual museum Canada.	33
Figura 2.15 El KORE browser.	34
Figura 2.16 Aplicación para el museo de mármol.	35
Figura 2.17 Pantalla del mapa general.	36
Figura 2.18 Pantalla de MINIE desplegando una imagen de la biblioteca.	37
Figura 3.1 El robot Flo.	44
Figura 3.2 Grafo simplificado del MDP básico utilizado por el manejador de diálogos en Flo.	44
Figura 3.3 Pearl interactuando con los ancianos en el asilo.	45
Figura 3.4 HOMER.	46
Figura 3.5 Cabeza de HOMER.	47
Figura 4.1. Funcionalidad de SCDI.	50
Figura 4.2 Estructura del Sistema Conversacional para Despliegue de Información.	52
Figura 4.3 Diagrama de bloques de SCDI.	53
Figura 4.4. Funcionamiento de SCDI y Jessica.	56
Figura 4.5 El rostro animado Jessica.	58
Figura 5.1 Estructura de los MDPs en SCDI.	61
Figura 5.2 Matriz de transición para la acción 1 del menú edificio principal.	63
Figura 5.3 Política obtenida para el MDP de "edificio principal".	67
Figura 6.1 Respuesta a la pregunta uno sobre la efectividad del sistema.	71
Figura 6.2 Respuesta a la pregunta dos sobre la efectividad del sistema.	71
Figura 6.3 Respuesta a la pregunta tres sobre la efectividad del sistema.	72
Figura 6.4 Respuesta a la pregunta cuatro sobre la efectividad del sistema.	72
Figura 6.5 Respuesta a la pregunta cinco sobre la efectividad del sistema.	73
Figura 6.6 Respuesta a la pregunta seis sobre la eficiencia del sistema.	73
Figura 6.7 Respuesta a la pregunta siete sobre la eficiencia del sistema.	74
Figura 6.8 Respuesta a la pregunta ocho sobre la eficiencia del sistema.	74

Figura 6.9 Respuesta a la pregunta nueve sobre la satisfacción del usuario.....	75
Figura 6.10 Respuesta a la pregunta diez sobre la satisfacción del usuario.....	76
Figura 6.11 Respuesta a la pregunta once sobre la satisfacción del usuario.....	76
Figura 6.12 Usuario interactuando con SCDI.....	86
Figura 6.13 Ambiente de realización de pruebas.	87

Índice de Tablas

Tabla 5.1 Estados para el submenú “edificio principal”	62
Tabla 5.2 Acciones para el submenú “edificio principal”.	62
Tabla 6.1 Tiempos de interacción para la tarea 1 del grupo de usuarios 1.	77
Tabla 6.2 Tiempos de interacción para la tarea 2 del grupo de usuarios 1.	78
Tabla 6.3 Tiempos de interacción para la tarea 1 del grupo de usuarios 2.	78
Tabla 6.4 Tiempos de interacción para la tarea 2 del grupo de usuarios 2.	78

Capítulo I. Introducción

1.1. Antecedentes

El campo de la interacción humano-computadora es relativamente nuevo en comparación con otras áreas de estudio dentro de la computación. Su objetivo ha sido el de mejorar la comunicación entre humanos y máquinas, realizando notables avances en el diseño e implementación de interfaces que hacen que dicha comunicación sea eficiente.

Una de las técnicas de uso común en la interacción humano-computadora es la interacción mediante voz. Existe una gran variedad de aplicaciones que hacen uso del reconocimiento de voz, que en su mayoría se han usado para comandar y controlar interfaces de usuario gráficas [Strong 1993] o como un sustituto del teclado y ratón en aplicaciones de dictado [Danis 1994]. La incursión del reconocimiento de voz en interfaces promete brindar una interacción más natural y simple. Los sistemas basados en voz han tenido diversas aplicaciones, entre ellas se encuentran los kioscos de servicio, cuya característica principal es la de proporcionar información sobre un lugar específico. Usualmente son sistemas multimodales, en donde se utilizan diferentes tecnologías que se complementan para ofrecer una interacción más completa, como el uso de *touchscreens*, plumas, síntesis y reconocimiento de voz, y despliegue de información multimedia. Su uso puede encontrarse en estaciones de trenes, de autobuses o aeropuertos para la venta y reservación de boletos, o para registro de pasajeros; o como centro informativo en museos, centros turísticos y edificios [Johnston 2005].

El presente proyecto se plantea dentro del contexto de la Cátedra de Robótica Móvil del ITESM Campus Cuernavaca. La Cátedra es un grupo de investigación dedicado a enfrentar diversos problemas de la robótica móvil, entre ellos navegación, exploración, localización, arquitectura, aprendizaje, laboratorios virtuales e interacción humano-robot. Ésta última línea de investigación es donde se ubica específicamente el tema de la presente tesis, cuyo objetivo es el de proporcionar una interfaz mediante el uso de voz para un robot guía.

En el ITESM Campus Cuernavaca se desarrolló una tesis abordando el tema de los sistemas guía. El sistema "Interfaz móvil para visualización de información en ambientes interactivos mediante PDAs" [Alarcón 2004], permite realizar un paseo guiado por el Campus. La idea general es que al contar con un PDA, el usuario puede acceder a un mapa interactivo del Campus y consultar información sobre sus distintas áreas. Esta información consiste de imágenes y videos respectivos a cada área.

El trabajo realizado por Alarcón sirve como punto de partida para esta tesis, ya que se trasladó la idea de consultar información del Campus en una PDA a un sistema de interacción por voz montado en un robot. En lo posterior se describirá como se relacionan ambos trabajos.

1.2. Motivación

El habla es una herramienta que el ser humano aprende a temprana edad, y el medio usado diariamente para la comunicación. Hablar es una habilidad innata propia de los humanos y el poder tener aplicaciones controladas por voz resulta por lo tanto más cómodo y natural. Hablar es una alternativa que libera la atención durante tareas donde se requiere la ocupación de la vista, de las manos, o que requieren cierta movilidad. Se dice que la interacción por voz es la nueva tendencia dentro del área de desarrollo de interfaces, con el potencial para cambiar fundamentalmente la forma en que se interactúa con computadoras [Lai 2000]. Otro punto a favor de los sistemas basados en voz es que ofrecen ventajas para la interacción con usuarios visualmente discapacitados o con limitaciones motoras. Por otro lado es útil para usuarios comunes al evitar la molesta tarea de identificar letras muy pequeñas en los monitores.

El hecho de incluir el lenguaje natural como mecanismo de entrada hace que las computadoras aparenten ser más humanas. Esto hace que la interacción sea más atractiva y satisfactoria para los usuarios, una característica deseable para cualquier sistema que pretenda atraer a un público diverso.

1.3. Planteamiento del Problema

La Cátedra de Robótica Móvil está desarrollando actualmente un proyecto para implantar un robot que actúe como anfitrión en el ITESM Campus Cuernavaca. Una de las necesidades fundamentales de este proyecto es la inclusión de un mecanismo de interacción que brinde un acceso fácil y amigable para el visitante. La interacción por voz es una alternativa atractiva para basar el sistema, ya que es la vía natural para la comunicación entre humanos. El diseño e implementación de un sistema basado en voz debe contemplar una interacción sencilla e intuitiva para el usuario. Se plantea que el sistema robótico presente información verbal o en forma de imágenes y videos sobre las distintas áreas del Campus a solicitud de un usuario. Así también, se pretende que los diálogos se den de una manera natural utilizando técnicas conversacionales de iniciativa mixta.

Un aspecto importante es el desempeño óptimo de los diálogos a ejecutar. Esto significa que el usuario debe obtener la información deseada en el menor tiempo posible lo que implica que se ejecute una secuencia de diálogos mínima.

Para ello se considera aplicar Procesos de Decisión de Markov (MDPs) [Puterman 1994] que permitan tomar la secuencia más adecuada de entre las opciones existentes, de acuerdo a la solicitud del usuario registrada. De igual manera es primordial brindar al usuario una interfaz visual a la cual dirigirse cuando está intentando comunicarse con el robot, por lo que como complemento al sistema de diálogos se considera incluir un rostro animado en 3D.

1.4. Objetivos

1.4.1 Objetivo general

La presente tesis tiene como objetivo general diseñar e implementar un sistema de diálogos óptimo mediante el uso de MDPs, que permitan interactuar con un robot anfitrión con la finalidad de desplegar información multimedia al usuario.

1.4.2 Objetivos específicos

1. Utilizar el enfoque de Diseño Centrado en el Usuario para la formulación de los diálogos.
2. Diseñar los diálogos mediante un enfoque de iniciativa mixta.
3. Implementar el sistema de diálogos bajo el lenguaje de marcado VoiceXML.
4. Optimizar los diálogos haciendo uso de MDPs.
5. Hacer uso de MDPs jerárquicos.
6. Realizar pruebas del sistema con distintos usuarios con el fin de verificar el buen desempeño del sistema.

1.5. Alcances y Limitaciones

Para el reconocimiento y síntesis de voz se utiliza el *Voice eXtensible Markup Language* VXML [VXML Manual 2004], un software gratuito, por lo tanto las gramáticas creadas estarán restringidas a diálogos de iniciativa mixta u orientados por el sistema.

Las necesidades del proyecto demandan que el sistema deba ser interactivo de manera que el robot actúe como anfitrión. Es decir, el visitante podrá interactuar de forma verbal con el robot con el fin de que éste le presente la información requerida eficientemente. Esto significa que el sistema debe responder de manera acertada y rápida en la medida de lo posible para brindar una interacción satisfactoria al visitante. Por lo cual el desempeño eficiente del

sistema implica que la selección de los diálogos sea óptima. Para el caso de este sistema, esto se logra utilizando MDPs, donde los estados corresponden a las solicitudes realizadas por el usuario, las acciones a las respuestas del sistema, y la recompensa esta en función de la meta, o sea, la información solicitada.

Una de las características distintivas del sistema es la presentación de información visual por parte del robot, como se describió anteriormente. Para ello se usa tecnología basada en HTML y ASPs que tiene la finalidad de desplegar videos e imágenes. Dichos videos e imágenes se proyectarán en una pantalla mediante un cañón. Por otro lado, una de las limitaciones más restrictivas del sistema es su sensibilidad al ruido, por lo que se contempla que el usuario use un micrófono de diadema para garantizar su cercanía con el dispositivo y permitir un mejor reconocimiento de los comandos que enuncie. La restricción del ruido también condujo a que los comandos orales proporcionados al sistema consistan de frases cortas para permitir un mejor reconocimiento de los mismos.

El sistema es dependiente de la plataforma Windows XP, ya que como se explicará posteriormente, se hace uso de *Internet Information Services* (IIS) de Microsoft que opera bajo dicha plataforma, ya que el uso de un servidor *web* era necesario para poder comunicar los diferentes módulos dentro del sistema, así como con el proyector de imágenes y videos.

1.6. Descripción General del Sistema

El SCDI es un sistema que presenta información multimedia al usuario mediante la interacción por voz. Su funcionamiento consiste en solicitar información accediendo a las opciones que el sistema vaya presentando. Su estructura está basada en el uso de MDPs, que permite un acceso a la información más rápido y versátil. El sistema está organizado en dos módulos principales, el módulo de interacción que se encuentra sobre el robot y el módulo de despliegue que se encuentra en un servidor externo conectado a un cañón para proyectar la información. El módulo de interacción consiste de una parte desarrollada en VXML que llevará a cabo la síntesis y el reconocimiento de voz, y la parte de MDPs que proporciona la política óptima para cada estado dado un conjunto de acciones que el sistema puede realizar. En la figura 1.1 se muestra el diagrama de bloques del sistema conversacional, el proceso comienza en la parte de MDP que genera un archivo de texto conteniendo la política óptima. A su vez la parte de VXML accede al archivo de texto mediante un ASP para verificar que acción ejecutar de acuerdo a su estado actual y el estado destino. El mismo ASP devolverá la instrucción al bloque VXML para ejecutar cierta acción, si dicha acción implica desplegar información, el bloque VXML mandará la orden a otro ASP para ubicar la información solicitada y finalmente desplegarla en una página HTML.

El uso de MDPs ha sido una técnica poco abordada para el manejo de diálogos, como puede observarse en la revisión de la literatura dentro del capítulo tres. En este trabajo se demuestra que la aplicación de MDPs a interfaces por voz puede brindar un desempeño más provechoso de las mismas y en consecuencia brindar una experiencia de interacción más agradable para el usuario.

Para comprobar esto, se realizó un estudio de usabilidad en donde se comparó la primera versión del sistema conversacional con estructura jerárquica de menús y la versión usando MDPs. Los resultados mostraron que la segunda versión del sistema tuvo un funcionamiento superior al primer sistema, proporcionando un uso más eficiente. Los usuarios denotaron tener una interacción más satisfactoria y rápida, como se observó en los cuestionarios aplicados.

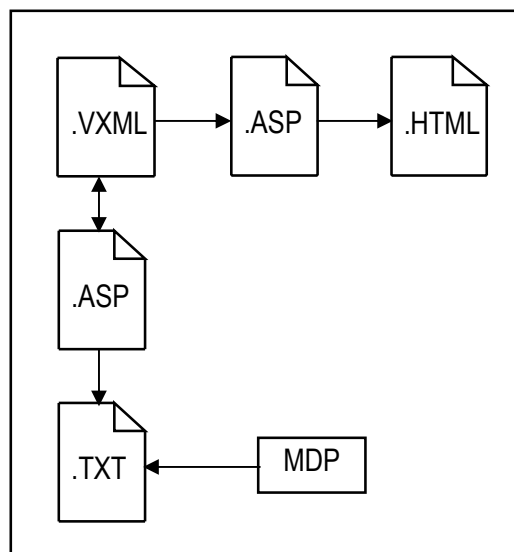


Figura 1.1 Diagrama de bloques de SCDI con MDPs.

1.7. Organización del Documento

La presente tesis esta organizada de la siguiente manera: En el capítulo siguiente se presenta una introducción al concepto de interacción humano-computadora y específicamente el trabajo relacionado con sistemas basados en voz, robots guías y asistentes, y sistemas de despliegue de información.

En seguida, el capítulo tres aborda la temática de los Procesos de Decisión de Markov (MDPs), sus métodos de solución y el trabajo realizado con MDPs y aplicaciones por voz.

El capítulo cuatro describe el Sistema Conversacional para Despliegue de Información (SCDI). Esto es, su diseño, implementación, una descripción general de su funcionamiento, y la inclusión de un avatar.

El capítulo cinco explica la metodología utilizada en la aplicación de los MDPs al SCDI. Se describe el modelado de los MDPs y su implementación dentro del SCDI.

El capítulo seis incluye los experimentos realizados con el sistema conversacional y los resultados derivados de los mismos.

Finalmente el capítulo siete presenta las conclusiones del proyecto y el trabajo futuro.

Capítulo II. Interacción Humano - Computadora

En este capítulo se describirá el concepto de la interacción humano-computadora, el área de la interacción humano-robot dentro de ese contexto y el uso del reconocimiento y síntesis de voz para el desarrollo de interfaces. También se abordará el tema de las actitudes de los humanos con respecto a los robots.

2.1 Introducción

La interacción humano-computadora (IHC) es una disciplina que concierne el diseño, evaluación e implementación de sistemas computacionales interactivos para el uso de humanos y el estudio de fenómenos mayores que los rodean [SIGCHI 2005]. Surgió como un campo de estudio a partir de las áreas de gráficos computacionales, sistemas operativos, factores humanos, ergonomía, ingeniería industrial, psicología cognitiva y la parte de sistemas de las ciencias computacionales.

El campo de la IHC, ha hecho grandes avances hacia el entendimiento y la mejora de nuestra interacción con tecnologías basadas en computadoras. Hasta hace poco tiempo, los investigadores de IHC habían hecho poco trabajo con robots. Sin embargo, los recientes avances en tecnología computacional, inteligencia artificial, simulación y entendimiento del habla, y controles remotos, han conducido a descubrimientos en tecnología robótica que ofrecen implicaciones significativas para la comunidad de IHC. Los robots móviles autónomos pueden identificar y rastrear la posición de un usuario, responder a preguntas habladas, mostrar información textual o espacial, y navegar por comandos mientras evitan obstáculos. Los robots pronto podrán asistir en un rango de tareas que son indeseables, inseguras, confusas, mal pagadas, o aburridas para las personas [Kiesler 2004].

Uno de los objetivos para la interacción entre las personas y los robots, se basa en conversaciones acerca de tareas que una persona y un robot pueden hacer juntos. Esta meta no solo requiere conocimiento lingüístico acerca de la operación de conversación, y conocimiento del mundo real de como realizar las tareas en conjunto, sino que el robot también debe interpretar y producir comportamientos que transmitan la intención de mantener la interacción o de terminarla [Sidner 2003].

2.2 Uso de Diálogos

El argumento más fuerte para hacer uso de la voz como modalidad de entrada es que representa la manera más natural para comunicación de los humanos. Es nuestro medio primario de comunicación, lo aprendemos a muy temprana edad y utilizamos diariamente en nuestras vidas.

El diálogo es un proceso conjunto de comunicación que involucra compartir información, como datos, símbolos o contexto entre dos o más partes [Landsdale 1994]. Los humanos emplean una variedad de señales sociales para-lingüísticas (expresiones faciales, gestos, etc.) para regular el flujo del diálogo [Cassell 1999]. Dichas señales han probado ser útiles para controlar un diálogo entre un humano y un robot [Breazeal 2003].

El diálogo, sin importar su forma, es significativo sólo si se aterriza, por ejemplo, cuando los símbolos utilizados por cada parte describen conceptos comunes. Si los símbolos difieren, primero debe darse el intercambio de información o el aprendizaje antes de que pueda proceder la comunicación. Aunque la comunicación entre humanos y robots puede ocurrir de diversas formas, se considera que existen tres tipos principales de diálogos [Fong 2003a]: de bajo nivel (pre-lingüístico), no verbal y lenguaje natural.

2.2.1 Diálogo de bajo nivel

El diálogo de bajo nivel puede verse como un lenguaje que refleja un estado previo a un lenguaje completo. Billiard y Dautenhahn [Billiard 1997]; [Billiard 1998]; [Billiard 1999] describen una serie de experimentos en donde se educó a un robot autónomo con un proto-lenguaje sintético. El aprendizaje del lenguaje resulta de múltiples asociaciones espacio-temporales a través del espacio de estados sensor-actuador del robot.

Steels ha examinado la hipótesis de que la comunicación se ajusta a un proceso de aprendizaje social y que el significado es inicialmente dependiente del contexto [Steels 2001a]; [Steels 2001b]. En sus experimentos, un perro robot aprende palabras simples que describen la presencia de objetos (pelota, roja, etc.), su comportamiento (caminar, sentarse) y las partes de su cuerpo (pierna, cabeza).

2.2.2 Diálogo no verbal

El diálogo no verbal puede definirse como un conjunto o sistema de símbolos que no requiere de vocabulario y que permite la comunicación. Existen varias maneras de lenguaje no verbal incluyendo el posicionamiento del cuerpo, gesticulaciones y acciones físicas. Debido a que muchos robots tienen la

capacidad rudimentaria de reconocer y producir voz, el diálogo no verbal es una alternativa útil. Nicolescu y Mataric [Nicolescu 2001] describen un robot que pide ayuda a los humanos comunicando sus necesidades e intenciones a través de sus acciones.

Las convenciones sociales o normas también pueden ser expresadas a través de diálogo no verbal. Una de tales convenciones es la proxemia, el uso social del espacio [Hall 1966]. Las normas de proxemia incluyen conocer como formarse en una fila, como caminar en los pasillos, etc. Respetando estas convenciones espaciales puede involucrar la consideración de numerosos factores administrativos, culturales, etc. [Nakauchi 2000].

2.2.3 Diálogo mediante lenguaje natural

El término “lenguaje natural” se usa para distinguir los lenguajes hablados por los humanos para el propósito de comunicación, de aquellos construidos como los lenguajes de programación en computación. El diálogo mediante lenguaje natural se determina por factores que van desde las capacidades físicas y perceptivas de los participantes, hasta las características sociales y culturales de la situación. Hasta que punto las interfaces humano-robot deben basarse en lenguaje natural sigue siendo un tema de debate [Severinson 2003].

Severinson, Green y Hüttenrauch [Severinson 2003] discuten que tan explícita debe ser la retroalimentación para que los usuarios interactúen con robots de servicio. Su enfoque es proveer lenguaje natural. Fong, Thorpe y Baur [Fong 2001] [Fong 2003b] describen como un diálogo de alto nivel puede permitir que un humano asista a un robot. En su sistema, el diálogo se limita a asuntos de movilidad como navegación, evadir obstáculos, etc., haciendo énfasis en actos de pregunta-respuesta.

El habla es un método altamente efectivo para comunicar emociones. Los parámetros principales que rigen el contenido emocional del habla son el volumen, intensidad y prosodia. En [Murray 1993] se afirma que los efectos vocales causados por emociones particulares son consistentes entre interlocutores con algunas diferencias menores.

La calidad de la síntesis de voz es significativamente más pobre que la síntesis de expresiones faciales y lenguaje corporal [Bartneck 2002]. A pesar de esto, se ha probado que es posible generar habla emocional. En [Cahn 1999] se describe un sistema para mapear calidad emocional, como tristeza, a las especificaciones de un sintetizador de voz, incluyendo articulación, intensidad y calidad de voz.

El habla emocional se ha utilizado en pocos sistemas robóticos hasta la fecha. Breazeal describe el diseño del sistema de vocalización Kismet. Se generan enunciaciones expresivas al ensamblar cadenas de fonemas con acentos intensos [Breazeal 2002].

2.3 Sistemas Basados en Voz

En esta sección se describirán algunas de las herramientas para el desarrollo de aplicaciones de voz y algunos sistemas que hacen uso de voz como medio principal de interacción.

2.3.1 Sistemas para el desarrollo de aplicaciones de voz

El Hidden Markov Model Toolkit (HTK) [HTK 2005] es una herramienta portable para desarrollar y manipular Modelos Ocultos de Markov (HMM). Principalmente se usa para investigación en reconocimiento de voz, a pesar de que se ha usado en numerosas aplicaciones incluyendo investigación en síntesis de voz, reconocimiento de caracteres y secuenciación de DNA. HTK consiste de un conjunto de módulos, librerías y herramientas disponibles en C. Las herramientas proveen facilidades sofisticadas para el análisis de voz, entrenamiento de HMM, análisis de pruebas y resultados.

Sphinx-4 es una herramienta desarrollada en la universidad de Carnegie Mellon [CMUSphinx 2005]. Es un sistema de reconocimiento de voz de estado del arte, escrito completamente en Java. Creado en conjunto con Sun Microsystems, Mitsubishi, Hewlett Packard, UCSC y el MIT. Es un sistema muy flexible capaz de ejecutar diferentes tipos de tareas de reconocimiento.

CSLU Toolkit [Sutton 1998] fue desarrollado por el centro para la comprensión de lenguaje hablado en Oregon (CSLU por sus siglas en inglés). Integra reconocimiento de voz, síntesis de voz, animación facial y lectura por voz junto con otras herramientas para crear interfaces de usuario basadas en voz. El *toolkit* es gratuito y fue diseñado para soportar investigación básica, desarrollo y actividades educativas relacionadas con sistemas de lenguaje hablado e interfaces humano-computadora.

ViaVoice de IBM [ViaVoice 2005] proporciona un entorno informático multimodal. Soporta vocabularios especializados como médico y legal. Permite dictado, edición y corrección de textos y navegación con *Internet Explorer* mediante voz.

VoiceXML [VXML Manual 2004], es un lenguaje de marcado para desarrollo de aplicaciones de voz bajo entornos web. Soporta tanto reconocimiento como síntesis de voz y vocabulario para diversos idiomas. Está diseñado para crear diálogos que contengan síntesis de voz, audio digitalizado, reconocimiento de voz o marcado de teclas, grabación de voz, telefonía y conversaciones de iniciativa mixta.

Se eligió VoiceXML como lenguaje para desarrollar el Sistema Conversacional para Despliegue de Información, ya que permite utilizar infraestructura, herramientas y técnicas *web* muy usuales. Es la única especificación de un lenguaje basado en XML para desarrollo de aplicaciones de voz que recibió la Recomendación Final del W3C (*World Wide Web Consortium*). También permite diseño de aplicaciones distribuidas y ofrece portabilidad para las aplicaciones. Algunos desarrolladores encuentran que el desarrollo bajo VoiceXML es tres veces más rápido que algunos ambientes tradicionales de IVR (*Interactive Voice Response*) [VoiceXML Forum 2005].

2.3.2 Aplicaciones de voz

Actualmente el uso de voz como medio de interacción está teniendo un auge en el desarrollo de aplicaciones, sin embargo hace varias décadas que la voz se utiliza en diferentes tipos de sistemas. A continuación se describen algunas aplicaciones que usan entradas de voz como medio primordial de interacción.

Hyperphone [Muller 1990] es un navegador de documentos basado en voz. Utiliza manipulaciones de voz y un vocabulario precompilado para realizar secuencias de navegación independientes del orador. El usuario puede rastrear información previa al navegar hacia atrás a través de las ligas históricas almacenadas. El sistema incluye interacción mediante interrupción para intervenir en una lista de objetos en la opción deseada.

El sistema Chatter [Ly 1993] usaba el teléfono como una interfaz para aplicaciones de mensajes y *rolodex* (archivo giratorio con tarjetas removibles, usualmente usado para almacenar nombres, teléfonos y direcciones). Mejoró el sistema PhoneShell con reconocimiento de voz independiente del usuario. Una contribución mayor de este desarrollo fue el uso de razonamiento basado en memoria para predecir las preferencias del usuario al navegar e incorporar diálogos que solicitaran dichas preferencias.

VoiceNotes [Stifelman 1993] tuvo el propósito de crear una herramienta "*hand held*" que permitiera grabar pensamientos e ideas espontáneas para recuperarlas posteriormente. El usuario podía definir temas generales para almacenar dinámicamente y borrar la información concerniente en distintas categorías. La navegación dentro de las categorías era secuencial en una jerarquía de un solo nivel.

Edwards, Mynatt y Rodriguez [Edwards 1993] propusieron el sistema Mercator para proveer acceso a X-Windows para usuarios visualmente discapacitados. Desarrollaron un modelo fuera de pantalla para crear maneras alternativas de presentar información. Se definió un agente lector de pantalla para traducir la presentación gráfica en una presentación no visual y para proveer diversos mecanismos de entrada que fueran más apropiados con la nueva interfaz. Se combinó interacción táctil y auditiva para hacer la interfaz accesible a los usuarios.

El sistema Pegasus desarrollado por el MIT [Zue 1994], presenta un ambiente gráfico que acepta entradas de voz para interactuar con un sistema de reservación de vuelos en línea. El diseño de los diálogos fue preparado de tal forma que condujera al usuario a través de una conversación dirigida a metas. Las secuencias de diálogos se construyeron con una máquina de estados finitos que alcanzaba metas después de no más de aproximadamente 30 estados de diálogo.

Yankelovich, Levow y Marx [Yankelovich 1994] desarrollaron SpeechActs que provee acceso a aplicaciones de oficina como mensajes de texto o voz, calendario, *rolodex* e información específica como estado del tiempo y cambio de divisas. Usa un enfoque de comandos y control con selección dinámica de vocabulario que cambia de acuerdo al diálogo actual.

CU Communicator [CU Communicator 2005] es un sistema de diálogos para acceder a disponibilidad, reservaciones y presupuestos de aerolíneas, hoteles y renta de autos en vivo. Los usuarios pueden marcar al sistema utilizando un teléfono y lenguaje natural para interactuar con el agente de viajes automatizado. El agente accede al internet por la información de vuelos, hoteles y renta de autos. Utiliza parseo robusto y estrategias de control de diálogos para ser lo más flexible posible a la varianza de usuarios. Para poder desarrollar fácilmente los sistemas, adoptan una representación declarativa extensa, donde el dominio de información específica se provee en archivos externos

2.3.3 Kioscos informativos

El kiosco automatizado Multimodal Multimedia Service Kiosk (MASK) [Lamel 2002] originalmente surgió como un servicio informativo en una estación de trenes, pero que tiene el potencial para usarse en sectores como la industria y comercio incluyendo tiendas, oficinas, bancos o centros turísticos. El kiosco tiene un *touchscreen* para entradas táctiles, bocinas y dos micrófonos ubicados arriba y debajo de la pantalla. Es capaz de proveer los horarios y precios de trenes y simulación de la venta de boletos. Los componentes principales para el entendimiento de habla son el reconocedor de voz, el componente de lenguaje natural y un componente de obtención de información que incluye acceso a base de datos y generación de respuestas. La figura 2.1 muestra la interfaz del kiosco.

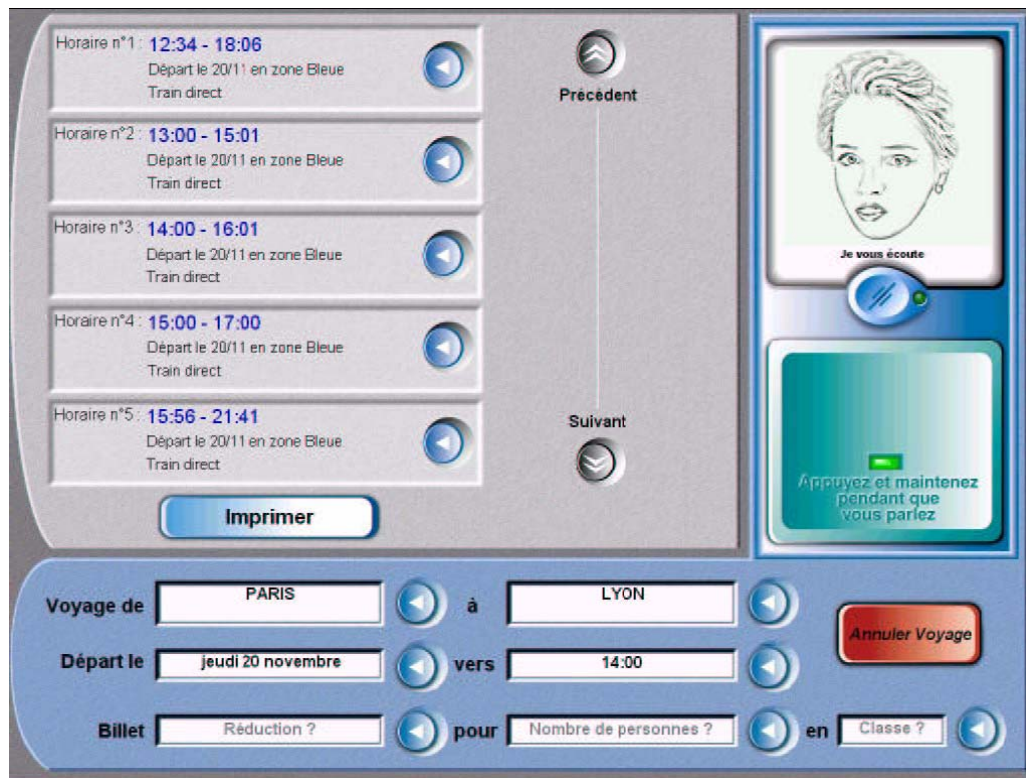


Figura 2.1 Interfaz gráfica de MASK.

La aplicación MATCHKiosk [Johnston 2005], es una guía multimodal interactiva de una ciudad que ayuda a localizar puntos de interés en una ciudad, como restaurantes o estaciones de metro. Tiene un *touchscreen* que soporta entradas tanto de tacto como de pluma y contiene una impresora. El gabinete (figura 2.2) contiene bocinas y un micrófono montado arriba de la pantalla. Su interfaz gráfica (figura 2.3) tiene tres componentes principales: en la derecha un panel con un mapa dinámico, un botón para hablar y una ventana para retroalimentación en el reconocimiento de voz. Las funciones básicas del sistema incluyen permitir al usuario localizar restaurantes y otros puntos de interés basado en atributos como precio, ubicación y tipo de comida, para solicitar información como dirección, y reseñas, y proveer información sobre las estaciones de metro cercanas. También se incluyen comandos para acercamientos y alejamientos sobre el mapa.



Figura 2.2 El kiosco MATCH.



Figura 2.3 Interfaz gráfica de MASK.

2.4 Robots Guías y Asistentes

Un ámbito interesante en donde se han aplicado los sistemas basados en voz, son los sistemas de museo. Estos sistemas se caracterizan por proporcionar al visitante, un tour guiado del museo, además de presentarle información complementaria sobre las obras expuestas en él. Un ejemplo de ello son los robots de entretenimiento usados en el *Museo para la Comunicación* de Berlín [Graf 2000]. La peculiaridad de estos robots radica en que además de guiar al visitante a través de las instalaciones, éste puede solicitar al robot información más detallada sobre el museo en general. Dicha información consiste en brindar datos adicionales en forma verbal o en forma de imágenes o videos presentados en una pantalla colocada sobre el mismo robot. En la figura 2.4 se muestra a los robots del museo.



Figura 2.4 Robots de entretenimiento usados en el Museo para la Comunicación de Berlín.

En [Burgard 1998] se describen las funciones desempeñadas por el robot guía denominado RHINO, que fue implantado en el *Museo Alemán* de Bonn durante un período de seis días. Dicho robot tiene las tareas principales de interactuar con los visitantes al museo y con usuarios a través de la red, a la vez que navega por el museo entre los grupos de personas y evitando obstáculos. La interacción con los visitantes se da de manera multimedial, ya que integra texto, gráficos, grabaciones y sonido. El robot cuenta con un tablero consistente de cuatro botones de distintos colores y una pantalla. El usuario puede elegir entre una explicación breve pregrabada o un tour mediante alguno de los botones que se encuentran sobre el tablero del robot. Cuando el robot se aproxima a alguna de las exhibiciones del museo, despliega un aviso en su pantalla y reproduce una pequeña explicación verbal pregrabada. Los usuarios pueden solicitar información más detallada o dejar que el robot proceda. Una vez que termine un tour, el robot regresa a una posición de inicio predefinida en el área de entrada a la espera de nuevos visitantes. En la figura 2.5 se ilustra a RHINO en el museo.

RHINO también posee la habilidad de reaccionar físicamente a la gente, ya que utiliza movimientos de cuerpo y cabeza, y sonidos para expresar intención o insatisfacción. Esto es, la cámara usada como cabeza indica la dirección en la que se desea dirigir. Así también utiliza un sonido como claxon para indicar insatisfacción cuando la gente u otro obstáculo no previsto obstruyen su paso.

Otra característica importante de la interfaz de usuario de RHINO, es su interfaz vía web. La página web permite que los usuarios observen la operación del robot. En ella se muestran tres imágenes, la primera presenta una imagen proveniente de la cámara de RHINO; la segunda imagen muestra el mapa del ambiente y las acciones actuales de RHINO; y finalmente una imagen que muestra la perspectiva desde una de las cámaras colocadas en las paredes del museo. Cuando RHINO está realizando la explicación de alguna de las exhibiciones, se despliega información relacionada en la página que incluye ligas a material más detallado.



Figura 2.5 RHINO interactuando con el público.

Minerva [Thrun 1999] es otro robot guía que fue exhibido en operación durante dos semanas en el *Museo Nacional Smithsonian de Historia Americana*. Fue desarrollado por el mismo equipo que trabajó con RHINO. Al igual que RHINO, las tareas de Minerva principalmente consisten en navegar en ambientes dinámicos e interactuar con los visitantes al museo y usuarios en la web. Sin embargo, presenta mejoras en comparación con RHINO, puesto que posee un repertorio de

interacción más rico que radica en “estados emocionales” que comunican sus intenciones y también posee una interfaz web mejorada. Cuando da tours, Minerva utiliza su cara, dirección de la cabeza y voz para interactuar con la gente. Para modelar sus “estados emocionales” se empleó un autómata estocástico de estados finitos, para comunicar sus intenciones a los visitantes. Sus estados de ánimo van desde contenta a enojada, dependiendo de la persistencia de la gente para obstruir su paso. La figura 2.6 muestra a Minerva.

Cuando estaba contenta, Minerva sonreía y le pedía amablemente a la gente que se retirara del camino; cuando estaba enojada, fruncía la cara y su voz sonaba enojada. La mayoría de los visitantes no tuvo dificultad en identificar los estados emocionales e intenciones del robot. Para atraer a la gente Minerva usó un enfoque basado en memoria de aprendizaje por refuerzo. Recibía una recompensa de acuerdo a la cercanía de las personas; sin embargo, acercarse demasiado conducía a una penalización, ya que se violaba su espacio. El comportamiento de Minerva dependía de la densidad de gente. Las acciones posibles incluían diferentes estrategias, como mover la cabeza, mostrar diferentes expresiones (contenta, triste, enojada) y diferentes anuncios verbales (“Ven aquí”, “¿te gustan los robots?”). El aprendizaje ocurría durante el período de un minuto en fases de “agrupamiento” que sucedían entre tours. La figura 2.7 ilustra los estados de ánimo de Minerva.



Figura 2.6 El robot Minerva.

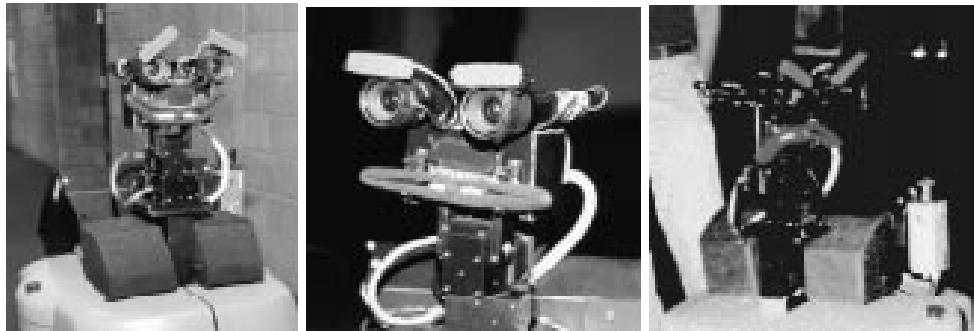


Figura 2.7 Estados de ánimo de Minerva.

Minerva estaba conectada a la web, desde donde los usuarios podían controlarla y ver imágenes grabadas por ella. También podían ver su ubicación en un mapa. Además, usando una cámara con *zoom*, los usuarios podían ver a Minerva y a los visitantes cercanos desde cierta distancia. Durante los horarios de apertura del museo los visitantes controlaban a Minerva seleccionando tours en una *touchscreen* montada en su espalda. Cada tres horas, los usuarios de la web votaban por los tours que deseaban realizar.

Sage [Nourbakhsh 1999] es un robot móvil educativo que se instaló en el *Museo Carnegie de Historia Natural* durante un período de nueve meses. El objetivo fue crear un robot que actuara como educador, captara la atención de la audiencia y proporcionara información educativa efectivamente. La idea general es tener un robot educativo que atraiga a los visitantes y los guíe a las exhibiciones de interés. El robot da la bienvenida a los visitantes y los invita a presionar un botón que se encuentra sobre él para realizar un tour. En cada parada del tour muestra una presentación multimedia que se compone de animación computacional, filmaciones de los paleontólogos, recreaciones, efectos de sonido e imágenes. El video mostrado por Sage cuenta con subtítulos para aquellas personas con discapacidades auditivas. El tour informativo dura 30 minutos en total realizando seis paradas.

Sage también cuenta con distintos estados de ánimo, siendo estos: contento/ocupado, solitario, cansado, frustrado y confundido. Se utilizó un modelo difuso de máquina de estados, en donde distintas acciones externas conducirían la transición de un estado de ánimo a otro. Por ejemplo, si una persona bloquea el camino de Sage, éste tendrá una transición en su estado de ánimo de contento a frustrado, lo cual llevará a que eleve el tono de voz y solicite a la persona que se retire. En contraste, cuando Sage está solo y un visitante llega a bloquear su camino, éste tendrá un comportamiento juguetón e incitador, para invitar a la persona a realizar un tour. La figura 2.8 muestra a Sage.

CERO [Green 2001] es un robot diseñado para auxiliar a personas físicamente discapacitadas en un ambiente de oficina. Su función principal es la de asistir a los usuarios en sus tareas diarias como traer o entregar objetos, y facilitarlas para las personas con impedimentos físicos. El robot no cuenta con una pantalla con la cual interactuar, por lo que se le implantó un pequeño personaje que permite desplegar ciertos gestos. La interfaz por voz utilizada para CERO tiene un enfoque centrado en el usuario y un diseño basado en estados. Las tareas desempeñadas por el sistema son: Ir a, que pone al robot a navegar; entregar, para llevar un objeto mientras navega; y traer, donde se dirige a un usuario y espera a recibir un objeto para llevarlo de regreso. La figura 2.9 muestra una imagen de CERO.



Figura 2.8 El robot Sage.

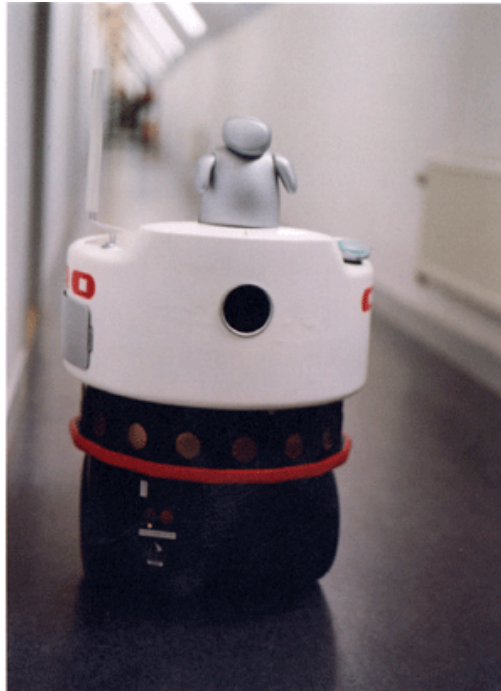


Figura 2.9 El robot CERO.

Se construyó un sistema de diálogos modular. Cuando recibe alguna entrada de audio, lo mapea a un posible comando utilizando una gramática libre de contexto. El resultado del reconocimiento de voz se pasa como una lista de palabras al análisis semántico que produce una representación lógica del comando. El manejador de diálogos intenta mapear la expresión lógica a una acción apropiada. El manejador de diálogos se basa en un conjunto de reglas que deciden que acciones son apropiadas dada una entrada y el estado actual del sistema. La figura 2.10 presenta el personaje montado en CERO.



Figura 2.10 Personaje montado en CERO.

Hygeiorobot es un asistente para hospitales propuesto en [Spiliotopoulos 2001]. Desempeña tareas simples en un hospital como entregar mensajes o medicinas a habitaciones específicas e interactuar con el personal del hospital vía diálogos hablados. No se contemplan entradas multimodales, ya que el robot no cuenta con dispositivos apuntadores o teclado. Se utilizó un enfoque basado en estados para los diálogos con el fin de desarrollar una plataforma de desarrollo rica.

Hygeiorobot permite a los usuarios especificar toda la información necesaria para entregar medicinas o mensajes a alguna habitación específica o paciente. Los usuarios también pueden solicitar información acerca de los pacientes, como el teléfono o número de habitación de algún paciente. Además, permite poner tareas a la espera. Se desarrolló una mini gramática para cada estado que usa conocimiento sobre las enunciaciones que son más probables a usarse en cada estado para identificar y extraer piezas de información relevantes de las entradas del usuario. Conocer que palabras esperar en cada estado ayuda al reconocedor de voz. El manejador de diálogos se diseñó de forma que aceptara diálogos relativamente cortos, ya que la meta principal es asegurar que las entradas del usuario se interpreten correctamente. Sin embargo, existen subdiálogos de confirmación en puntos clave que permitan a los usuarios revisar la interpretación del robot de sus solicitudes, y repetirlos de ser necesario. Para el sistema de diálogos se utilizó el CSLU *toolkit*.

El trabajo realizado con robots guías descrito en esta sección, denota que por lo general el tipo de interfaz utilizada carece del uso de entradas mediante voz a excepción de los dos últimos trabajos reseñados. El SCDI está basado en la interacción mediante voz para hacer dicha interacción atractiva para los visitantes. En el caso de CERO y Hygeiorobot, están contemplados para interactuar con las mismas personas, en el primer caso con personal de una oficina y el último con el personal del hospital. En cambio el SCDI está considerado para interactuar con diferentes tipos de usuarios, que pueden incluir desde gente que casi no ha tenido contacto con computadoras hasta personas que la utilizan diariamente.

2.4.1 Percepción humana de los robots sociales

Una diferencia notable entre los robots convencionales y los socialmente interactivos, es que la forma en que cada humano percibe al robot establece las expectativas que guían su interacción con él. Esta percepción, especialmente relacionada con la inteligencia, autonomía y capacidades del robot, es influenciada por numerosos factores tanto intrínsecos como extrínsecos [Fong 2003a]. Claramente las ideas preconcebidas, el conocimiento y la previa exposición del humano al robot tienen una influencia muy fuerte. Además, los aspectos en cuanto a diseño del robot, como su personificación o diálogo, puede jugar un papel significativo.

A continuación se presenta brevemente algunos estudios que examinan como estos factores afectan la interacción humano-robot, particularmente la forma en que los humanos se relacionan y trabajan con robots sociales.

2.4.2 Actitudes hacia robots

En [Bumby 1999] se llevó a cabo un estudio para identificar como la gente, en específico niños, perciben a los robots y que tipo de comportamiento exhiben al interactuar con ellos. Encontraron que los niños tienden a concebir a los robots como formas geométricas con rasgos humanos. Además, los niños tienden a atribuirle a los robots autonomía, preferencias, emociones y género masculino, aún sin una guía externa.

En [Khan 1998] se describe una investigación sobre las actitudes de las personas hacia robots de servicio inteligentes. Para diseñar el cuestionario se utilizaron una revisión de robots en la literatura y películas, seguida de una entrevista. El estudio reveló que las actitudes de las personas están fuertemente influenciadas por la ciencia ficción. Dos descubrimientos importantes fueron: 1. Se prefiere a un robot con apariencia de máquina, personalidad seria y con forma redonda. 2. Se desea preferentemente una comunicación verbal utilizando una voz humana.

2.4.3 Efectos de la apariencia y el diálogo

Un problema con el diálogo es que puede desembocar en percepciones prejuiciosas. Por ejemplo, se pueden crear asociaciones de comportamiento estereotipado, lo cual puede conducir a los usuarios a atribuir cualidades al robot que son incorrectas. Los usuarios también pueden formar modelos incorrectos, o hacer suposiciones falsas acerca de cómo funciona el robot realmente. Esto puede guiar a consecuencias serias, la menor siendo error del usuario [Fong 2003b].

Kiesler y Goetz [Kiesler 2002] condujeron una serie de estudios para entender la influencia de la apariencia y diálogo de un robot. Una contribución importante de este trabajo, son las métricas para caracterizar los modelos mentales que usan las personas cuando interactúan con robots. Un descubrimiento significativo fue que ni las valoraciones ni las observaciones de comportamiento por si solas son lo suficientes para describir las respuestas humanas a los robots. Los autores concluyeron que el diálogo tiene mayor influencia al cambiar los modelos mentales y su desarrollo, que las diferencias en apariencia.

En [DiSalvo 2002] se investigó como las características y tamaño de los rostros de robots humanoides contribuyen a la percepción de humanidad. En este estudio, se analizaron 48 robots y se condujeron investigaciones para medir la percepción

de la gente. El análisis estadístico mostró que la presencia de ciertas características, las dimensiones de la cabeza, y el número de rasgos faciales tienen gran influencia en la percepción de humanidad. También se hace referencia a la teoría del *Uncanny Valley* o “valle extraño” (Figura 2.11), la cual establece que mientras más incrementa la apariencia humana en un robot, existe un punto en donde el robot no es 100% similar a los humanos, pero el balance entre la similitud a humanos y máquinas es incómodo.

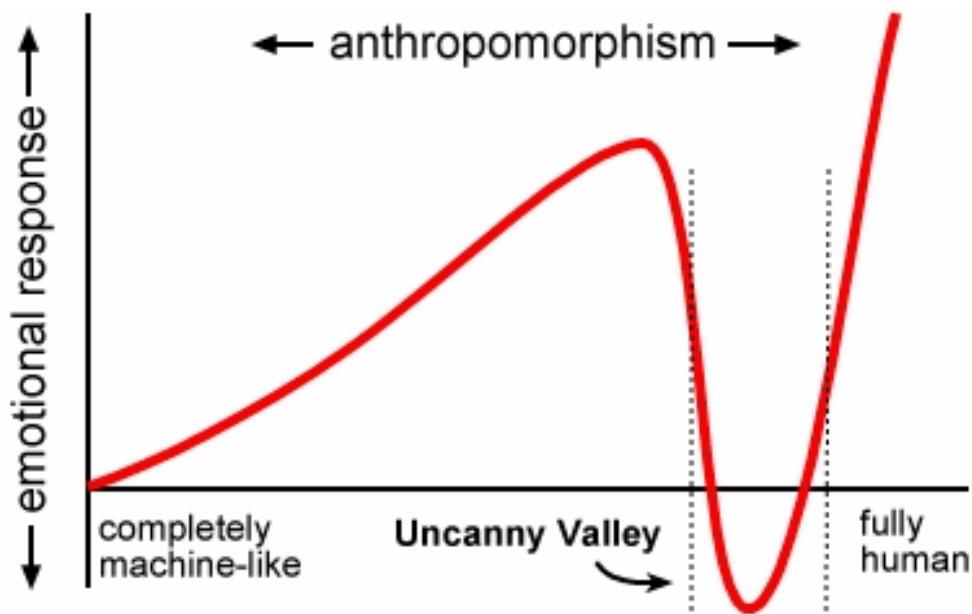


Figura 2.11 La teoría del Uncanny valley.

2.5 Sistemas para Despliegue de Información

Como ya se ha mencionado el SCDI presenta información a manera de imágenes y videos referentes a las distintas secciones del Campus. En el trabajo realizado por Alarcón [Alarcón 2004] se hace la analogía de que el Campus es un museo en sí y las distintas áreas son las salas de exhibición, por lo que el sistema va presentando la información respectiva a cada una de estas “salas”. Con el objetivo de exponer en que forma presentan información otros sistemas similares, se presenta a continuación una descripción del trabajo relacionado con sistemas de despliegue de información en ambientes de museo.

2.5.1 Museos virtuales

El museo Smithsonian [Smithsonian 2005] alberga en su página web, una sección para exhibiciones disponibles en la red. En esta sección se pueden visualizar imágenes o videos de obras expuestas en el museo con diferentes temáticas. Incluye una descripción de la obra tal como lo haría en la visita física al museo. El sitio permite ver un acercamiento de la imagen, así como ver una lista de los objetos disponibles. En la figura 2.12 se presenta una imagen del sitio.

El *Museum System* [Museum system 2005] es un sistema manufacturado por Gallery Systems, Inc. Está diseñado como una base de datos robusta para catalogar, almacenar y manejar colecciones de objetos, ya sean objetos de arte, presentaciones, imágenes digitales, audio y video. El *Art History Visual Resources Library* utiliza el *Museum System* para manejar su colección de presentaciones e imágenes digitales. La figura 2.13 presenta la interfaz del Museum system.

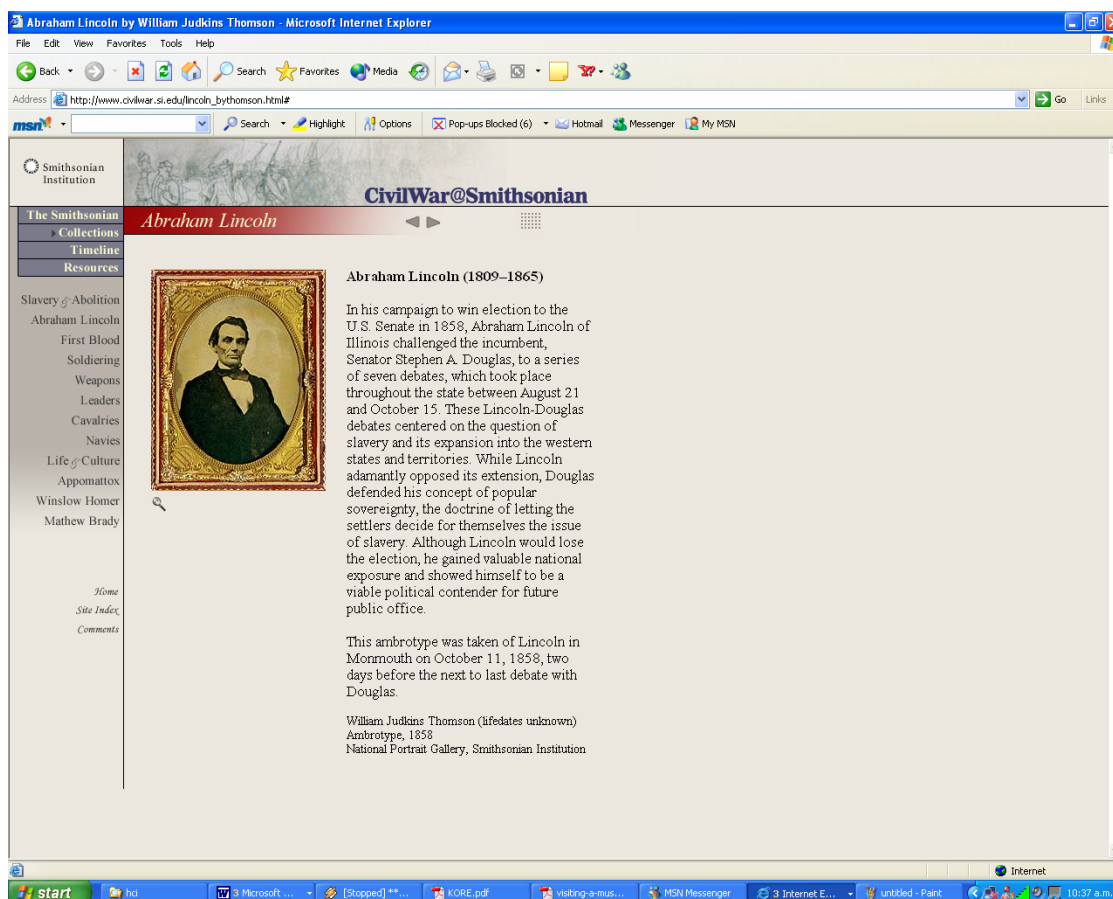


Figura 2.12 Exposición sobre la guerra civil norteamericana.



Figura 2.13 Museum system.

El *Virtual Museum Canada* [VMC 2005] permite conocer las distintas exhibiciones albergadas en diferentes museos de Canadá. Se pueden visualizar imágenes de las obras, junto con una descripción de la misma. También se encuentran videos y presentaciones *flash*. La figura 2.14 muestra algunas exhibiciones disponibles en el sitio.



Figura 2.14 Virtual museum Canada.

2.5.2 Museos interactivos y PDAs

KORE [Bombara 2001] es una arquitectura multiagente, parte de un proyecto de investigación para realizar un guía electrónico personalizado para un visitante de museo. La idea básica es proveer al visitante con la información referente a la obra de arte que esté observando. Dado que el usuario estaría equipado con algún dispositivo como PDA o teléfono celular, KORE es capaz de reconocer su posición y la obra que esté observando. La información se filtra y adapta mediante KORE en base al perfil de usuario del visitante. La figura 2.15 muestra el *KORE browser*.

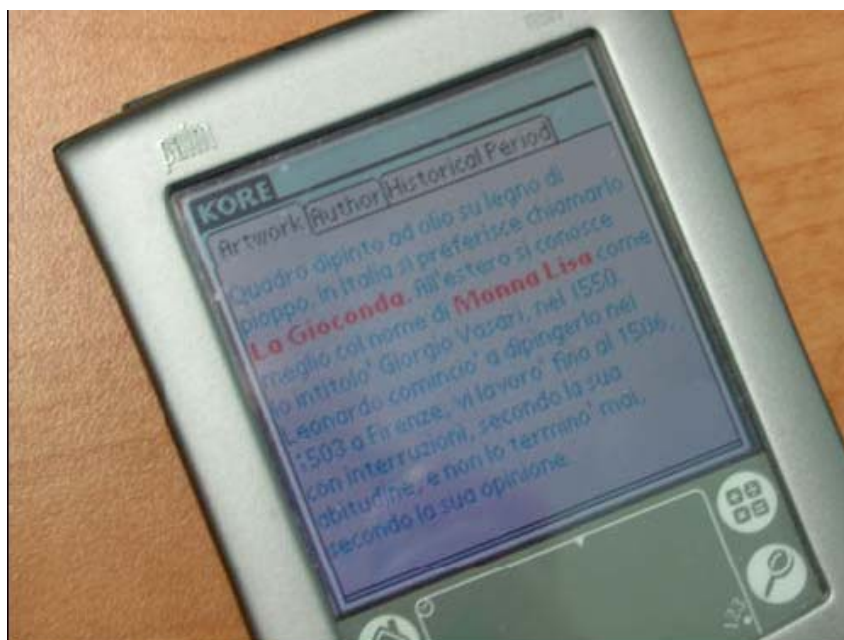


Figura 2.15 El KORE browser.

Para determinar el perfil de cada usuario, KORE maneja la siguiente información:

1. Preferencias personales, como lenguaje, edad, nivel de instrucción, tipo de arte preferido, etc.
2. Capacidades del dispositivo, como tamaño de pantalla, número de colores, etc.

Esta información la administra la aplicación cliente que se ejecuta en el dispositivo portable del usuario, o sea, el *KORE browser*. Las capacidades del dispositivo son detectadas por el *browser*, mientras que las preferencias se activan mediante el menú KORE. Los datos del perfil son almacenados en el dispositivo y usados para filtrar y adaptar la información. Los datos de preferencia son usados

por el *browser* para determinar a que categorías pertenece el usuario al compararlas con un conjunto de categorías predeterminadas.

Una aplicación desarrollada para el *Museo de Mármol* en Italia para PDAs [Ciavarella 2002] permite al usuario realizar tres tareas básicas:

- *Orientarse dentro del museo*, donde se provee un mapa del museo, un mapa de la sección donde se encuentre y por cada ambiente físico que componga una sección, un mapa con íconos indicando las obras de arte más relevantes y su ubicación. Al seleccionar dichos íconos, se despliega una imagen del elemento correspondiente e información básica junto con una descripción auditiva.
- *Control de la interfaz de usuario*, por ejemplo permitiendo ajustar el volumen del audio, detener y reproducir y navegar a través de los diferentes niveles de detalle de la descripción del museo.
- *Acceso a información del museo*, se provee a distintos niveles de abstracción como museo, sección, ambiente físico, obra de arte en particular.

La figura 2.16 muestra la interfaz de la aplicación descrita.



Figura 2.16 Aplicación para el museo de mármol.

La Interfaz Móvil para Visualización de Información en Ambientes Interactivos Mediante PDAs [Alarcón 2004] es una interfaz móvil para navegación en ambientes interactivos, para el ITESM Campus Cuernavaca. Esta aplicación permite visualizar un mapa interactivo del Campus mediante una PDA (figura 2.17) y de acuerdo al área seleccionada se presentan videos o imágenes relacionadas. Simula que el Campus es un museo de gran tamaño, donde los edificios y departamentos representan las salas de exhibición y objetos de exposición. Esta aplicación presenta información general, localización del visitante dentro del museo, información de las salas de exposición y objetos que en ellas se exhiben. Se permite proyectar estos elementos mediante proyectores digitales. La aplicación fue desarrollada bajo tecnología ASP utilizando IIS. La figura 2.18 muestra la pantalla donde se proyectan las imágenes y videos con esta aplicación.

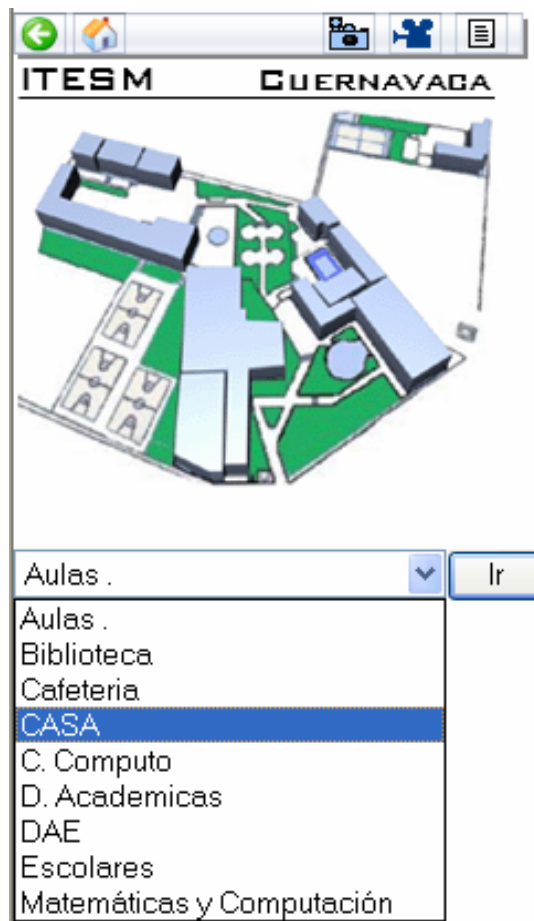


Figura 2.17 Pantalla del mapa general.



ADRIÁN A. ALARCÓN OSOLLO

MINIE - A MOBILE INTERFACE FOR NAVIGATION IN INTERACTIVE ENVIRONMENTS



Figura 2.18 Pantalla de MINIE desplegando una imagen de la biblioteca.

Los trabajos reseñados en esta sección describen la presentación de información en ambientes de museo. En la mayoría de los casos dicha información se despliega en forma de texto acompañando una imagen del objeto de exhibición. El SCDI además de imágenes presenta también videos, y la explicación de cada área se realiza en forma verbal, ya que a diferencia de los sistemas aquí descritos la interfaz no es gráfica si no mediante voz.

En el caso del sistema desarrollado por Ciavarella, se despliega una descripción auditiva junto con la información textual pero dicha descripción se realiza mediante una grabación, mientras que en el SCDI se hace uso de síntesis de voz. Por otro lado, se presentó la aplicación desarrollada por Alarcón que en sí sirvió como base para desarrollar SCDI. En esta aplicación el usuario se va guiando del mapa para poder acceder a la información, pero en el SCDI no se tiene dicho apoyo visual, así que el mapa se traduce en el menú de opciones de que SCDI va presentando. Cabe mencionar que algunos de los videos e imágenes utilizados en dicha aplicación se reutilizaron en SCDI.

2.6 Resumen del Capítulo

Este capítulo presentó una descripción del trabajo relacionado a sistemas que utilizan la interacción por voz. Como se pudo observar hay una vasta diversidad de aplicaciones que hacen uso de voz en distintos ámbitos, ya sea como medio principal de interacción o como medio auxiliar para otras interfaces. Dentro de las aplicaciones descritas también se hizo mención de las interfaces multimodales como medio para facilitar la interacción o para ofrecer versatilidad al sistema.

Sin embargo, el uso de voz como medio de interacción puede acarrear diversos problemas tanto en el reconocimiento como en la síntesis de voz, ya que suelen ser muy sensibles a ruidos externos del ambiente. Si los sistemas usan voz como único medio de entrada pueden representar problemas para el usuario al no contar con un medio visual que lo guíe a través de la interacción, si es que el sistema no brinda la suficiente retroalimentación; esto puede provocar una pobre recuperación de errores o disgusto por parte del usuario hacia la interfaz entre otras cosas. Una forma de agilizar la interacción en aplicaciones basadas en voz son los Procesos de Decisión de Markov, que se describen en el capítulo que sigue.

Capítulo III. Procesos de Decisión de Markov

En este capítulo se describe lo que es un Proceso de Decisión de Markov y los métodos más usuales para su resolución. También se presenta una reseña del trabajo realizado en la aplicación de MDPs y el manejo de diálogos.

3.1 Definición

Un Proceso de Decisión de Markov (MDP) es una formalización matemática de un problema en donde un agente debe decidir que acciones tomar para maximizar su recompensa esperada mientras interactúa con su ambiente. Se asume que no hay incertidumbre acerca del estado en que el agente se encuentra, es decir, el ambiente es completamente observable. Sin embargo, no se puede predecir el efecto de la elección de sus acciones [Sutton 1990]; [Koeing 1992]; [Puterman 1994]; [Littman 1995a].

Los MDPs son completamente observables, es decir, el agente conoce precisamente en que estado se encuentra. Sin embargo, si el agente no puede determinar en que estado se encuentra, se dice que su mundo es *parcialmente observable*. Para dichas situaciones se utiliza una generalización de los MDPs, denominado Proceso de Decisión de Markov Parcialmente Observable (POMDP). En un POMDP el agente no conoce precisamente el estado actual. Cuando actúa el estado de transición es igual que en un MDP. Sin embargo, después de actuar recibe una observación que proporciona información parcial acerca del estado actual [Littman 1995b]; [Littman 1995c]; [Meuleau 1999].

Un MDP se puede definir como una tupla (S, P, A, R) donde S es un conjunto finito de estados y A es un conjunto finito de acciones. Las acciones inducen transiciones entre estados con $P(s, a, s')$ denotando la probabilidad con la que se alcanza el estado s' cuando se ejecuta la acción a en el estado s . $R(s)$ es una función de recompensa que asocia a cada estado s su utilidad inmediata. Con un MDP el estado siguiente depende solo de la acción que fue tomada y del estado del cual proviene. Los estados o acciones anteriores no tienen ningún impacto. Esto se conoce como la propiedad Markoviana. La recompensa en cada estado depende de la acción ejecutada.

El objetivo de un agente es elegir una secuencia de acciones que maximicen su recompensa esperada. Una política mapea cada estado a una acción, la acción a tomar desde ese estado. Una política óptima proporciona la recompensa máxima esperada de cada estado [Jonas 2003].

3.1.1 Métodos de solución

Los problemas con un número finito de pasos se conocen como MDP de horizonte finito. Por otro lado los problemas en que puede haber un número infinito de pasos se conocen como MDP de horizonte infinito. Los métodos más usuales para la resolución de MDPs son: iteración de valor (Bellman 1957), iteración de política (Howards 1960) y programación lineal (Puterman 1994).

En el caso de horizonte infinito, se puede obtener la utilidad de los estados y la política óptima, mediante un método iterativo. El algoritmo de iteración de valor estima la utilidad de cada estado en cada iteración ($t+1$), basada en los valores de la iteración anterior (t). Cuando ($t \rightarrow \infty$), los valores de utilidad convergen a un valor estable [Sucar 2005].

Las probabilidades de transición están dadas por:

$$P_{ss'}^a = P \{s_{t+1} = s' \mid s_t = s, a_t = a\}$$

El valor de recompensa esperado es:

$$R_{ss'}^a = E \{r_{t+1} \mid s_t = s, a_t = a, s_{t+1} = s'\}$$

Donde:

s es el estado actual.

a la acción a ejecutar.

P es la probabilidad de pasar a un estado s' dado el estado presente s y la acción a .

R es el valor de la recompensa dado que se ejecuta la acción a en el estado s .

Se busca estimar las funciones de valor. Esto es, qué tan bueno es estar en un estado o realizar una acción. La noción de que tan bueno se define en términos de recompensas futuras o recompensas esperadas. Este material fue tomado ampliamente de [Morales 2004].

La política π es un mapeo de cada estado s y acción a a la probabilidad $\pi(s,a)$ de tomar la acción a estando en el estado s . El valor de un estado s bajo la política π , denotado como $V^\pi(s)$, es el refuerzo esperado estando en el estado s y siguiendo la política π .

Este valor esperado se puede expresar como:

$$V^\pi(s) = E_\pi \{R_t \mid s_t = s\} = E_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s \right\}$$

Y el valor esperado tomando una acción a en el estado s bajo la política π ($Q^\pi(s,a)$):

$$Q^\pi(s, a) = E_\pi \{R_t | s_t = s, a_t = a\} = E_\pi \left\{ \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a \right\}$$

Las funciones de valor óptimas se definen como:

$$V^*(s) = \max_\pi V^\pi(s) \text{ y } Q^*(s, a) = \max_\pi Q^\pi(s, a)$$

Las cuales se pueden expresar como las ecuaciones de optimalidad de Bellman:

$$V^*(s) = \max_a \sum_{s'} P^a_{ss'} [R^a_{ss'} + \gamma V^*(s')]]$$

y

$$Q^*(s, a) = \sum_{s'} P^a_{ss'} [R^a_{ss'} + \gamma V^*(s')]]$$

Si se conoce el modelo del ambiente, es decir, las transiciones de probabilidad y los valores de recompensas, las ecuaciones de optimalidad de Bellman representan un sistema de $|S|$ ecuaciones y $|S|$ incógnitas.

La función de valor V^π dada una política arbitraria π se calcula de la siguiente forma:

$$\begin{aligned} V^\pi(s) &= E_\pi \{R_t | s_t = s\} \\ &= E_\pi \{r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots | s_t = s\} \\ &= E_\pi \{r_{t+1} + \gamma V^\pi(s_{t+1}) | s_t = s\} \\ &= \sum_a \pi(s, a) \sum_{s'} P^a_{ss'} [R^a_{ss'} + \gamma V^\pi(s')]] \end{aligned}$$

donde $\pi(s, a)$ es la probabilidad de tomar la acción a en el estado s bajo la política π .

Podemos hacer aproximaciones sucesivas, evaluando $V_{k+1}(s)$ en términos de $V_k(s)$.

$$V_{k+1}(s) = \sum_a \pi(s, a) \sum_{s'} P^a_{ss'} [R^a_{ss'} + \gamma V_k(s')]]$$

Una de las razones para calcular la función de valor de una política es para tratar de encontrar mejores políticas. Dada la función de valor para una política dada, podemos probar una acción $a \neq \pi(s)$ y ver si su $V(s)$ es mejor o peor que el $V^\pi(s)$.

En lugar de hacer un cambio en un estado y ver el resultado, se pueden considerar cambios en todos los estados considerando todas las acciones de cada estado, seleccionando aquella que parezca mejor de acuerdo a una política *greedy*. Se puede calcular una nueva política $\pi'(s) = \operatorname{argmax}_a Q^\pi(s,a)$ y continuar hasta no mejorar. Esto significa partir de una política (π_0) y calcular la función de valor (V^{π_0}), con la cual encontrar una mejor política (π_1) y así sucesivamente hasta converger a π^* y V^* . Este procedimiento es denominado iteración de política (Howards 1960).

Uno de los problemas de iteración de política es que cada iteración involucra evaluación de política que requiere recorrer todos los estados varias veces. Sin embargo, el paso de evaluación de política se puede truncar sin perder la garantía de convergencia. Se puede detener después de recorrer una sola vez todos los estados. Esta forma se le llama iteración de valor. Se puede escribir combinando la mejora en la política y la evaluación de la política truncada:

$$V_{k+1}(s) = \max_a \sum_{s'} P^a_{ss'} [R^a_{ss'} + \gamma V_k(s')]$$

Se puede ver como expresar la ecuación de Bellman en una regla de actualización. Es muy parecido a la regla de evaluación de política, solo que se evalúa el máximo sobre todas las acciones.

Algoritmo de iteración de valor:

1. Inicializa: $V(s) = 0$ para toda $s \in S$
2. Repite
 - $\Delta \leftarrow 0$
 - Para cada $s \in S$
 - $v \leftarrow V(s)$
 - $V(s) \leftarrow \max_a \sum_{s'} P^a_{ss'} [R^a_{ss'} + \gamma V^*(s')]$
 - $\Delta \leftarrow \max(\Delta, |v - V(s)|)$
3. Hasta que $\Delta < \Theta$ (número positivo pequeño)
4. Regresa una política determinística tal que:

$$\pi(s) = \operatorname{argmax}_a \sum_{s'} P^a_{ss'} [R^a_{ss'} + \gamma V_k(s')]$$

Para espacios muy grandes, el ver todos los estados puede ser computacionalmente muy costoso, así que una opción es hacer estas actualizaciones al momento de estar explorando el espacio y por lo tanto determinando sobre que estados se hacen las actualizaciones.

3.2 Aplicación de MDPs en Diálogos

A continuación se describe el trabajo previo con respecto al uso de MDPs en diálogos, sobre todo su aplicación en robots de servicio como medio de interacción.

Roy, Pineau y Thrun [2000] proponen a Flo como asistente en un asilo de ancianos, cuya función es la de proporcionar la hora, el horario de medicamentos del paciente y la programación en las diferentes estaciones de televisión, así como responder a comandos de navegación. En la figura 3.1 se muestra una imagen de Flo.

Este proyecto aborda el problema de un manejador de diálogos utilizando tres enfoques: MDP, MDP aumentado y Proceso de Decisión de Markov Parcialmente Observable (POMDP). Realizan una comparación entre las tres técnicas y concluyen que POMDP tiene resultados más satisfactorios. Dado que el reconocedor de voz no siempre identifica de forma acertada las solicitudes del usuario, el POMDP permite realizar esta identificación la mayor parte de las veces en contraste con los otros dos enfoques. Flo utilizaba Sphinx II [Ravishankar 1996] como sistema de reconocimiento de voz y para la síntesis de voz el sistema Festival [Black 1999]. Las tareas determinadas para el robot consistían en que el usuario solicitara la hora, su horario de medicamentos, la programación en los distintos canales televisivos además de una lista de comandos de movimiento. Estas tareas se podían traducir dentro del problema de decisión en 13 estados, como se muestra en la figura 3.2. Se identifican 20 acciones de las cuales 10 corresponden a diferentes habilidades del robot, como ir a la cocina o dar la hora. Las 10 restantes son de aclaración o de confirmación, como reconfirmar el canal de TV deseado. La estructura de recompensas proporciona mayor recompensa por elegir acciones que satisfagan la petición del usuario. Estas acciones se redirigen de vuelta al estado inicial. La mayoría de las demás acciones son penalizadas con una cantidad equivalente negativa, mientras que las acciones de confirmación o aclaración se penalizan ligeramente con valores cercanos a 0, y los comandos de movimiento se penalizan fuertemente si se ejecutan en el estado incorrecto.



Figura 3.1 El robot Flo.

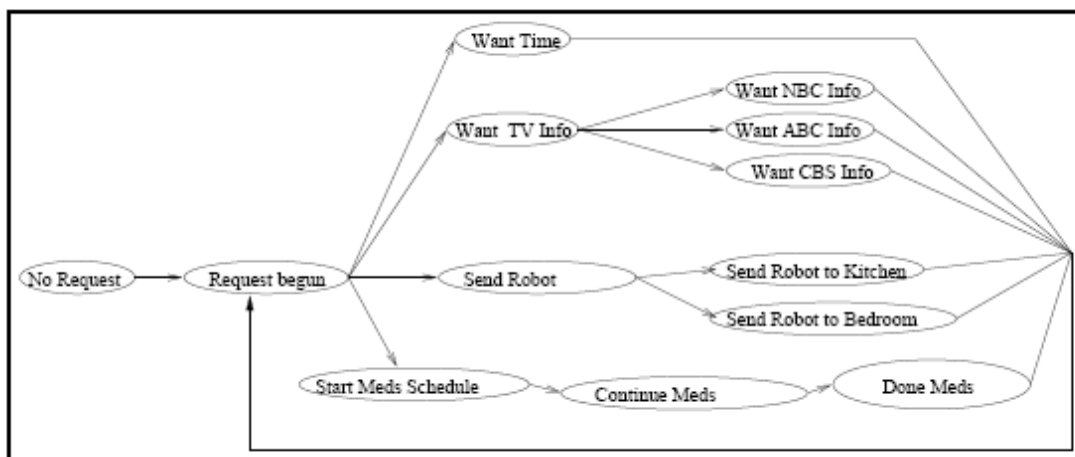


Figura 3.2 Grafo simplificado del MDP básico utilizado por el manejador de diálogos en Flo.

Pearl [Montemerlo 2002] es un robot para la asistencia de ancianos en un asilo y la versión mejorada de su antecesor Flo. Su función es la de proporcionar recordatorios a los usuarios, como sus citas médicas, y guiarlos a diferentes destinos dentro del propio asilo, por ejemplo a sus citas, terapias, eventos sociales o solo caminar para ejercitarse. Pearl desempeña estas funciones y se adapta a cada usuario en el sentido de que, lo cuestiona para saber si desea ser asistido y ajusta su velocidad a la del usuario monitoreando su progreso, pero también es tolerante a sus deficiencias verbales. Así también, proporciona información a solicitud del usuario como eventos comunitarios próximos, reportes del clima, horarios televisivos, etc. La figura 3.3 ilustra a Pearl interactuando con las personas en su ambiente.



Figura 3.3 Pearl interactuando con los ancianos en el asilo.

Pearl interactúa mediante voz y pantallas sensibles al tacto. Su módulo más importante con respecto a la interacción es un algoritmo probabilístico para control de alto nivel y manejo de diálogos. Esta arquitectura es una variante jerárquica de un POMDP. En el caso de Pearl, esta distribución incluye una multitud de estados probabilísticos multivaluados y variables de metas, como localización del robot, localización de la persona, status de la persona, meta de movimiento, meta recordatorio, meta iniciada por el usuario. En total existen 288 estados posibles. Las entradas del POMDP es una distribución de probabilidad factorizada sobre estos estados con incertidumbre que surge predominantemente de los módulos de localización y del sistema de reconocimiento de voz. Considerar la incertidumbre fue importante en este dominio puesto que los costos de errar en una respuesta eran altos. Los POMDPs del tamaño como el encontrado en este problema son de un orden de magnitud mayor al que pueden abordar los algoritmos actuales. Sin

embargo, el POMDP utilizado por Pearl está altamente estructurado en donde ciertas acciones se aplican solamente en ciertas situaciones. Para explotar esta estructura, se desarrolló una versión jerárquica de POMDPs, que descompone el problema de decisión en una colección de problemas menores y que pueden resolverse más eficientemente.

HOMER [Elinas 2004] es un robot móvil que comunica mensajes entre humanos en un ambiente de trabajo. HOMER consta de una pantalla con una cara animada, una cámara estéreo para poder navegar e identificar a los destinatarios de los mensajes, y un sistema de reconocimiento de voz para identificar los comandos. Este equipo propone un marco para la coordinación de tareas basado en MDPs múltiples que satisfagan diversos requerimientos. Éstos se denominan Procesos de Decisión de Markov Múltiplemente Seccionados (MS-MDPs). Utilizando una representación basada en MS-MDPs, se particiona la tarea en cierto número de subtareas, que se asignan cada una a un MDP para que se resuelva de manera independiente. Cada MDP controla un aspecto de la tarea global y se ejecutan concurrentemente, coordinados por variables de estado.

La tarea de entregar un mensaje se divide en tres subtareas, cada una controlada por su propio MDP. El navegador controla la navegación y localización del robot, el manejador de diálogos controla la interacción con la gente usando voz, y el generador de gestos controla la interacción con la gente usando gestos desempeñados por una cara animada. Las figuras 3.4 y 3.5 ilustran a HOMER interactuando con un usuario y el rostro de HOMER respectivamente.



Figura 3.4 HOMER.



Figura 3.5 Cabeza de HOMER.

Anteriormente se describió como el uso de MDPs en el reconocimiento de diálogos de Flo permite mejorar dicho reconocimiento, siendo el enfoque con POMDPs el que presentó resultados más satisfactorios. A pesar que el SCDI no hace uso de específicamente de POMDPs, este trabajo permitió tener una visión de las ventajas ofrecidas por MDPs y POMDPs en el manejo de diálogos.

El trabajo realizado con Pearl utiliza POMDPs no sólo para el manejo de diálogos, si no también para controlar otras funciones del robot, como la navegación y evitar obstáculos, por lo que abarca una problemática mucho más extensa que la abordada por los MDPs del SCDI, ya que en éste se contempló el manejo exclusivo de los diálogos. Además en Pearl se tiene una interacción multimodal al permitir entradas de voz y mediante *touchscreens*, a diferencia del SCDI donde sólo se tienen entradas por voz.

Con respecto al uso de MDPs en HOMER, se describe el uso de MS-MDPs en donde una tarea puede dividirse en subtareas y cada una ser resuelta por un MDP de manera simultánea. Este enfoque es distinto al utilizado por SCDI y mucho más ambicioso, pero también fue aplicado al manejo de diálogos entre otras funciones.

3.3 Resumen del Capítulo

En este capítulo se presentó la descripción del concepto de un Proceso de Decisión de Markov (MDP). Así también se enlistaron algunos de los trabajos realizados previamente en donde se ha hecho uso de MDPs dentro de la interacción mediante voz. En general los trabajos mencionados se refieren a robots de servicio cuya función es auxiliar a los humanos de alguna forma particular.

Con los trabajos expuestos aquí se puede observar que la aplicación de MDPs en los sistemas de diálogos puede proporcionar un mayor control y manipulación en el manejo de las entradas de los usuarios y por lo tanto se refleja en las acciones o respuestas que el sistema brinda. Por ende, la interacción se torna más ágil y certera en el cumplimiento de las solicitudes del usuario. Estos factores influyeron en la decisión de hacer uso de MDPs en el Sistema Conversacional para Despliegue de Información. En el capítulo siguiente se describirá el Sistema Conversacional para Despliegue de Información y en el capítulo 5 se presentará el uso de MDPs en el mismo sistema.

Capítulo IV. Descripción del Sistema Conversacional para Despliegue de Información

4.1 Sistema Conversacional para Despliegue de Información

SCDI es un *Sistema Conversacional para Despliegue de Información* diseñado para ser usado en un robot que servirá como guía en una locación determinada, inicialmente en nuestro Campus. El propósito del sistema es proporcionar información sobre las diferentes áreas del Campus a solicitud de los visitantes. El funcionamiento de SCDI inicia con un mensaje de bienvenida a la locación e invita al visitante a acompañarlo en un tour guiado del Campus. El sistema anima al visitante a solicitar información sobre una sección específica que desee conocer de una lista de áreas generales. Una vez que realizó su elección, el visitante recibe una descripción de las instalaciones del Campus y puede elegir ver videos o imágenes sobre ellos. El video o imagen que se haya elegido es desplegado en una pantalla por un cañón proyector.

SCDI cuenta con dos módulos principales: un módulo de interacción con el usuario y un módulo para el despliegue de la información. El módulo de interacción registra la solicitud del usuario y utiliza una serie de reglas para enviar la información apropiada al módulo de despliegue, quien responderá de vuelta al módulo anterior con un mensaje de confirmación para proseguir con la interacción. Esto se traduce en una nueva petición del usuario, donde se le puede preguntar si desea ver otro video o imagen de la misma área que fue seleccionada, visitar un área distinta o salir del tour.

El módulo a cargo de la interacción con el visitante se localiza a bordo del robot, y el que maneja el despliegue de imágenes y videos se localiza en un servidor externo. Ambos módulos se comunican utilizando *Active Server Pages* (ASPs) y el *Internet Information Service* (IIS) de *Microsoft* [IIS 2005], a través de una red local inalámbrica. La figura 4.1 muestra un diagrama del funcionamiento del SCDI.

4.1.1 Diseño del sistema

El primer paso fue diseñar los diálogos específicos para el robot guía, utilizando un diseño centrado en el usuario [Katz-Hass 2005]. El Diseño Centrado en el Usuario (UCD por sus siglas en inglés) es una filosofía y un proceso. Es una filosofía que ubica a la persona en el centro. Es un proceso que se enfoca en los factores cognitivos como la percepción, memoria, aprendizaje, solución de problemas, etc., al participar durante la interacción de las personas con las cosas.

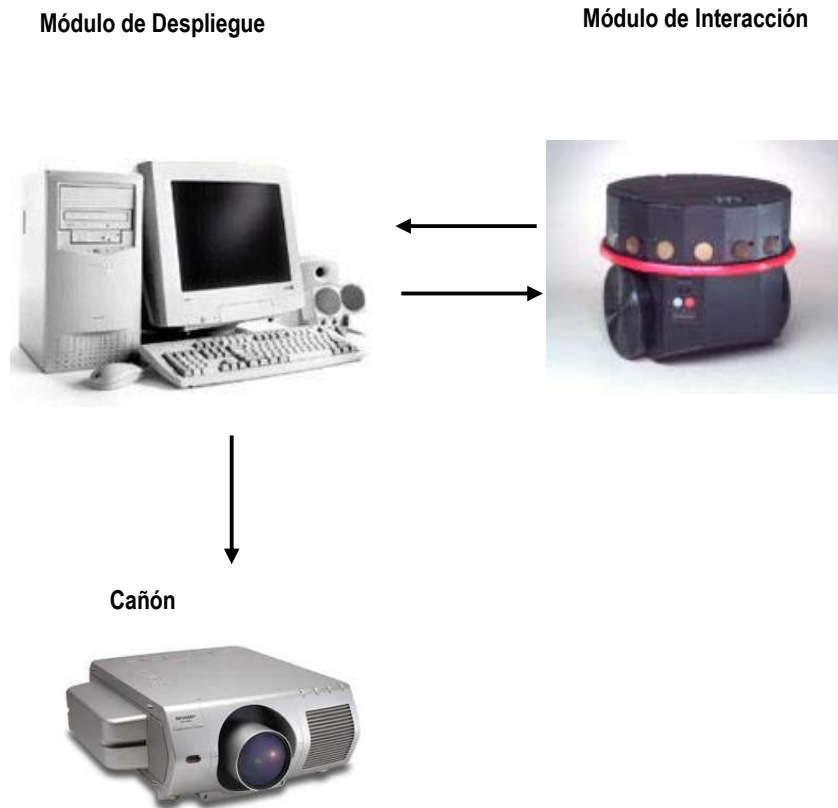


Figura 4.1. Funcionalidad de SCDI.

UCD busca responder preguntas acerca de los usuarios y de sus tareas y objetivos, para luego utilizar estos descubrimientos en la guía de desarrollo y diseño. El UCD buscar responder a preguntas tales como:

- ¿Quiénes son los usuarios de este sistema?
- ¿Cuáles son las tareas y objetivos de los usuarios?
- ¿Cuáles son los niveles de experiencia de los usuarios con este sistema y sistemas similares?
- ¿Qué funciones necesitan los usuarios que este sistema les provea?
- ¿Qué información podrían necesitar los usuarios y en qué forma la necesitan?
- ¿Cómo piensan los usuarios que este sistema deba funcionar?
- ¿Cómo puede el diseño de este sistema facilitar los procesos cognitivos de los usuarios?

El UCD puede mejorar la usabilidad y utilidad de cualquier cosa, desde objetos de uso diario hasta software, sistemas de información, procesos, etc., cualquier cosa con que la gente interactúe. Como tal el Diseño Centrado en el Usuario concierne tanto con la utilidad como la usabilidad.

Se consideró un enfoque de iniciativa mixta para la construcción de los diálogos, lo que permite que el usuario intervenga durante la interacción. La interacción por iniciativa mixta es un aspecto clave de una interacción humano-computadora efectiva. *Iniciativa mixta* se refiere a una estrategia de interacción flexible, donde cada participante, en este caso el sistema y el usuario, puedan contribuir a la tarea que se está realizando. En cualquier momento, uno de los participantes puede tener la iniciativa, controlando así la interacción, mientras que el otro trabaja para asistir al primero, contribuyendo a la interacción como se requiera. La clave de la interacción por iniciativa mixta, es que permite que los participantes trabajen de manera más eficiente como un equipo [Allen 1999].

Para estructurar la forma en que se presentaría la información al usuario, se elaboró un esquema de las áreas del Campus por ubicación física, es decir, los menús contenidos en el sistema se presentan por áreas generales que a su vez contienen subáreas y así sucesivamente. La estructura se realizó de esta manera, debido a que se basó en la distribución de las áreas del Campus presentada en los mapas de la aplicación descrita en [Alarcón 2004]. Por ejemplo, un área general es *edificio principal* que incluye la subárea de *servicios*, que a su vez se subdivide en los sitios específicos de biblioteca, cafetería y centro de cómputo (CEC). La figura 4.2 ilustra la estructura completa de los menús contenidos en el sistema.

Se consideró presentar el menor número de opciones posibles a la vez en el momento de las enunciaciones por parte del sistema, con la finalidad de no confundir al usuario y evitar que memorizara demasiada información, ya que podría conllevar a que el usuario olvidara las opciones disponibles. En caso de que el usuario no hubiera escuchado las opciones presentadas, éstas se le enuncian de nuevo después de un tiempo considerable. También se incluyó el despliegue de mensajes de confirmación para otorgar una retroalimentación al usuario.

4.1.2 Implementación

El sistema fue desarrollado bajo el marco del Voice eXtensible Markup Language (VXML) [VXML Manual 2004], un lenguaje de marcado para el desarrollo de aplicaciones de voz bajo ambientes *web*.

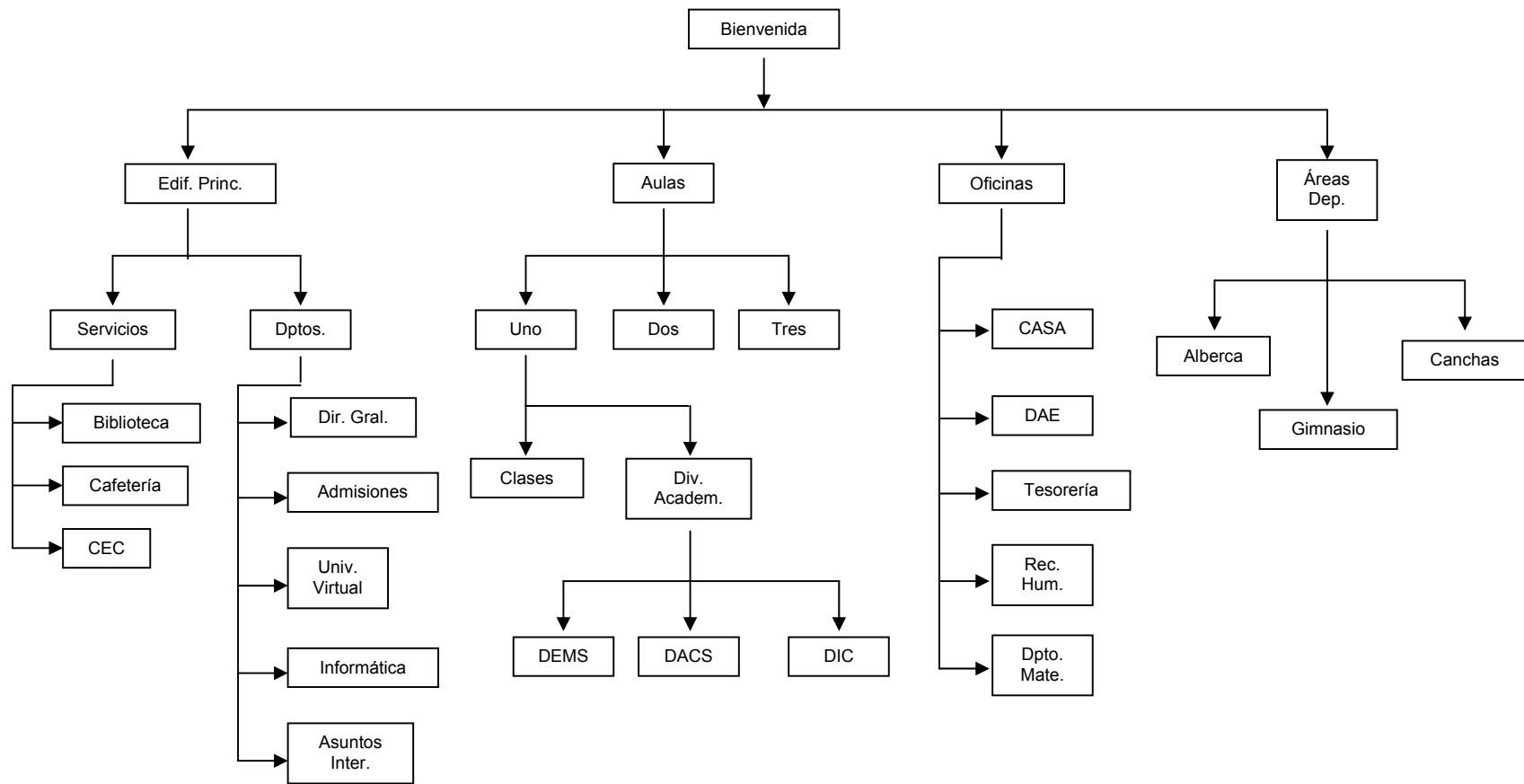


Figura 4.2 Estructura del Sistema Conversacional para Despliegue de Información

Dos técnicas que permitieron mayor flexibilidad en la aplicación y una interacción por iniciativa mixta, son: *barge-in* y enunciaciones alternativas [Larson 2003]. La primera permite al usuario interrumpir las enunciaciones del sistema, lo cual provoca una respuesta más pronta y mayor facilidad de acceso a la información. La segunda proporciona mayor libertad al usuario al decir palabras no necesariamente incluidas dentro de las opciones proporcionadas por el sistema. Esto significa que la gramática definida soporta una considerable cantidad de sinónimos de las palabras sugeridas por el sistema durante la interacción. Es preciso aclarar que el sistema no recibe entradas de frases o enunciados completos, ya que el entorno en donde se aplica es ruidoso, así que se consideró que se podría tener más control sobre las entradas del usuario si se usaban palabras aisladas o frases cortas para maximizar la capacidad de reconocimiento (consultar el Anexo B para ver los diálogos que soporta el sistema).

Para el despliegue de los videos e imágenes se optó por aprovechar el módulo desarrollado para la “Interfaz móvil para visualización de información en ambientes interactivos mediante PDAs” [Alarcón 2004], ya que cumple con el mismo fin de proyectar este tipo de información en páginas HTML. La comunicación con el módulo mencionado y el SCDI se hizo a través de ASP en VBscript. Para esto se envían los datos de que clase de información ha sido solicitada por el usuario, ya sea una imagen o un video, y se transfieren al módulo de despliegue, quien enviará un mensaje de confirmación al sistema conversacional cuando haya desplegado la información. Así SCDI puede continuar con la interacción normalmente. La figura 4.3 muestra el diagrama de bloques del SCDI.

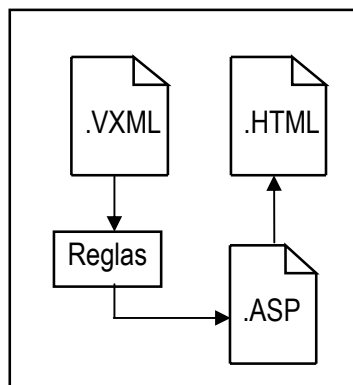


Figura 4.3 Diagrama de bloques de SCDI.

La estructura completa de los diálogos está elaborada por medio de menús jerárquicos, donde el usuario encuentra primeramente un menú de áreas generales, y cada opción puede dividirse en otros menús de subáreas. En este caso se cuenta con cuatro áreas generales, siendo éstas: *edificio principal*, *aulas*,

oficinas y áreas deportivas. Mientras el usuario interactúa con el sistema conversacional, basado en esta estructura jerárquica, puede acceder a áreas específicas de interés y visualizar su información detallada correspondiente.

Un ejemplo de los diálogos emitidos por el sistema se muestra a continuación:

Sistema: *Hola! Bienvenido al Campus. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?*

Visitante: *Edificio principal.*

Sistema: *El edificio principal se encuentra junto a la entrada. Alberga diferentes servicios y departamentos. ¿Qué deseas conocer?*

Visitante: *Departamentos.*

Sistema: *Los departamentos encontrados aquí son: dirección general, admisiones, universidad virtual, informática y asuntos internacionales. ¿Qué deseas conocer?*

Visitante: *Admisiones.*

Sistema: *En el centro de admisiones puedes conocer los programas que se ofrecen en el Campus y el proceso para ingresar a cualquiera de ellos. También ofrece información sobre becas y programas de financiamiento. Indica video o imagen para ver alguno de los videos o imágenes disponibles.*

Visitante: *Video.*

Sistema: *Tengo dos videos de admisiones. Indica uno o dos.*

Visitante: *Dos.*

(Se procesa la petición del usuario y se despliega la información solicitada).

Sistema: *¿Gustas ver otro video o imagen de admisiones?*

Visitante: *No.*

Como se mencionó previamente, el sistema se encuentra estructurado en menús, en donde el visitante va eligiendo de las diferentes opciones que se le proporcionan. Lo que significa que de un área general se va accediendo a subáreas.

Esto implica que el usuario deba pasar por diversas etapas para llegar a su objetivo final, es decir, que el sistema despliegue información sobre un área específica. Por lo tanto, es un aspecto importante que el desempeño de los diálogos sea óptimo, esto es, que el usuario no tenga que recorrer la estructura de menús completa para acceder al área que desea conocer. Para ello se consideró aplicar Procesos de Decisión de Markov (MDPs) que permiten tomar la secuencia mas adecuada de entre las opciones existentes, de acuerdo a la solicitud del usuario registrada.

4.2 Inclusión de un Avatar

4.2.1 Comunicación entre aplicaciones

El objetivo principal de integrar un rostro animado con el sistema conversacional, es el de alcanzar una interfaz más completa y amigable para que interactúe con el visitante. El hecho de tener un rostro para el robot permite que el visitante mejore su experiencia mientras interactúa con el robot. De acuerdo a [Green 2001] esto hace que el usuario esté más cómodo ya que provee un componente de interacción.

Se decidió que el rostro animado fuera parecido a un personaje de caricatura, ya que para realizar las enunciaciones se grabó una voz humana real, como se explicará posteriormente en esta sección. Boyce afirma en [Boyce 2000] que el tener un sistema que suene demasiado parecido a un humano puede crear que el usuario espere mucho más del sistema y provocar errores. El argumento principal utilizado por los investigadores de interacción humano-computadora en contra de los sistemas antropomórficos, es que pueden provocar que los usuarios esperen funciones no realistas del sistema; lo que significa que pueden esperar que el sistema entienda como un humano. Con base en esto se eligió utilizar un rostro animado caricaturesco.

Con respecto a la sincronización entre los diálogos y el rostro animado, se consideraron diferentes enfoques para el desarrollo de esta aplicación, incluyendo sockets [Stevens 1998] y ASPs [ASP 2005].

En este caso en particular, se decidió que el enfoque más apropiado sería el uso de ASPs, al igual que en la comunicación entre los módulos de SCDI. De esta manera se podría continuar con la misma línea de trabajo, ya que los ASPs desarrollados para SCDI proveen una plataforma firme en la cual basar el siguiente módulo de la aplicación. El primer paso fue el de sincronizar la animación de cada enunciado utilizado en SCDI con los archivos (WAV) de voz grabados. Este proceso requirió de la sincronización de cada palabra con su movimiento animado correspondiente. Para lograr esto se crearon diversos archivos de video en formato AVI que correspondieran a cada uno de los enunciados del sistema.

Debido a que estos archivos eran voluminosos y las necesidades del proyecto dictaban una aplicación que corriera en tiempo real, se encontraron dificultades con respecto a retrasos durante la ejecución de estos videos. La solución fue el convertir los videos en archivos ejecutables, los cuales se volvieron ligeros y por lo tanto más manejables.

Al incluir este avatar, llamado Jessica, la interacción del visitante se hace directamente con el rostro animado. Esto ubica al módulo de interacción de SCDI

en modo de escucha; una vez que el rostro animado enuncia su línea, SCDI procesa la respuesta del usuario tal como se describió anteriormente.

El módulo de interacción de SCDI y el módulo del rostro animado se encuentran a bordo del robot mientras que el módulo de despliegue se encuentra residiendo en un servidor externo conectado a un cañón proyector. La figura 4.4 muestra el funcionamiento de ambas aplicaciones en conjunto.

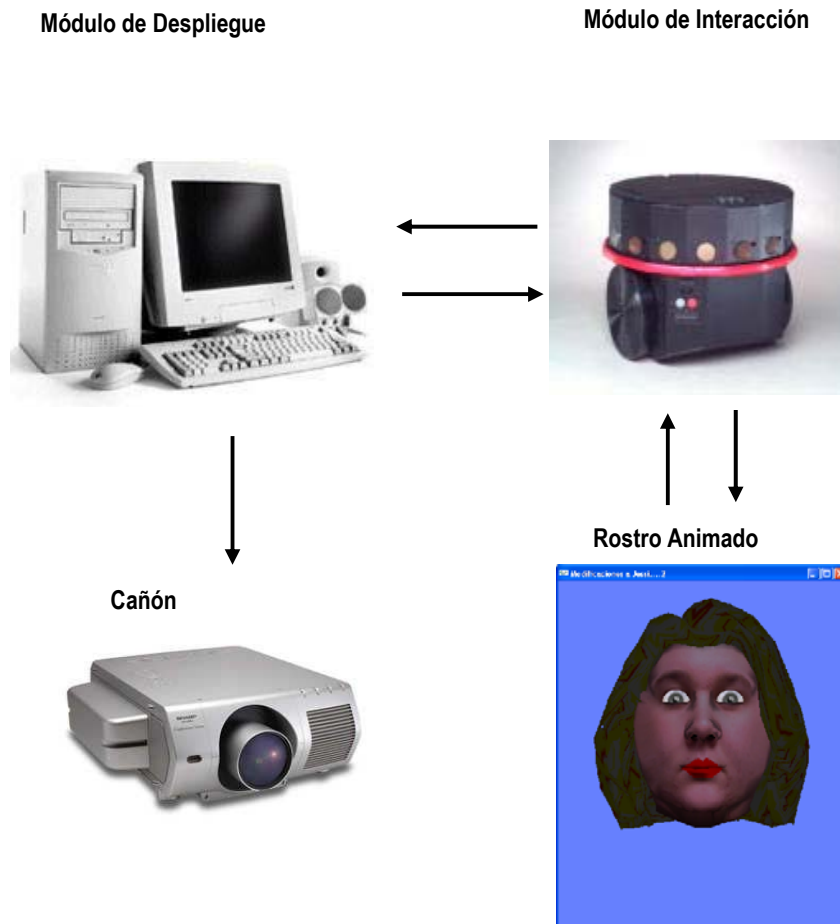


Figura 4.4. Funcionamiento de SCDI y Jessica.

4.2.2 Jessica: Un rostro animado

Jessica es un rostro animado [Ricaño 2005] utilizado como interfaz de usuario. El trabajo en dicha tesis consistió en un estudio de usabilidad para evaluar las preferencias de los usuarios con respecto a avatares como medio de interacción. El rostro está basado en un modelo pseudo-muscular convencional, adaptado

para uso directo con Puntos Faciales MPEG-4 (PFs). Este rostro se construye al incluir una estructura base sólida que no puede ser penetrada. Esta estructura es particularmente importante para la síntesis realista de la dinámica facial fuera de los PFs, y provee al modelo una apariencia de volumen que se encuentra ausente en otros enfoques de bajo costo. Jessica contiene una malla de 876 triángulos y 28 músculos que permiten expresiones faciales y movimiento.

Las características faciales de Jessica (ojos, boca y nariz) fueron hechas midiendo las distancias de los puntos basados en MPEG-4. Cada uno de los 28 músculos del rostro es asociado a los PFs correspondientes y a un área de influencia predeterminada. El área de influencia es una lista de aquellos vértices que son afectados por cada movimiento del músculo. La lista de vértices asociada a cada músculo se determinó una vez, pero puede ser modificada para crear nuevos modelos y movimientos del rostro. La figura 4.5 muestra el rostro animado.

4.2.3 Ajustes en el rostro animado para las enunciaciones del SCDI

En la tesis de Ricaño [Ricaño 2005] se trabajó con el SCDI descrito en este capítulo para probar a Jessica como medio de interacción. El mecanismo para integrar la interfaz por voz natural en el rostro animado se basa en la animación de las vocales, que constituyen los *visemas* más significativos en español. Así que los movimientos que simulan la modulación de las cinco vocales en español: *A, E, I, O, U*, fueron programadas para animación facial.

Después de que se programaron las animaciones de las vocales, se concatenaron secuencias de vocales para simular palabras. Se realizó el mismo proceso con las secuencias de palabras para simular la formulación de enunciados. Estos enunciados fueron formulados con base en los diálogos del sistema conversacional.

4.2.4 Resultados del estudio de usabilidad

Las pruebas realizadas con Jessica como interfaz gráfica del SCDI con un grupo de usuarios, consistieron en aplicar un cuestionario sobre las características deseables y/o indeseables para un rostro robótico animado. En seguida el usuario interactuó con el sistema sin Jessica, después interactuó con Jessica. Finalmente se aplicó un cuestionario sobre el desempeño del sistema con Jessica.

Los resultados obtenidos con el primer cuestionario, mostraron que las personas encuestadas prefieren ser guiados por un rostro robótico. Los encuestados respondieron que las características deseables del rostro son que tenga forma de caricatura, voz natural, agradable a la vista, amigable, atractivo y expresivo.



Figura 4.5 El rostro animado Jessica.

Los resultados del cuestionario número dos, después de la interacción con Jessica, revelaron que el rostro animado ayudó a entender mejor el sistema ya que daba la impresión de estar siendo atendidos por alguien y mejora la interacción con el sistema haciéndolo más amigable y realista. Más del 80 por ciento de los encuestados mencionaron que prefieren al SCDI con el rostro animado y que la interacción con él es mejor, pero en general la satisfacción de la interacción fue normal.

La opinión de los encuestados del rostro animado en general fue positiva. Mencionan que es amigable, que tiene buena coordinación el movimiento de la boca con la voz, que llama la atención y que es una buena herramienta de apoyo. También tuvo opiniones negativas al mencionar que el rostro es feo, que le faltan algunos movimientos y que distrae.

Las ventajas de utilizar a Jessica como interfaz gráfica conversacional es que atrae al usuario a interactuar con el sistema, llama la atención y facilita la comunicación. La desventaja es que distrae al usuario del objetivo principal, ya que se presta más atención al rostro que a la información que se brinda. Otra

desventaja es que la interacción con el sistema se hace más lenta ya que se necesita de muchos recursos computacionales para su ejecución.

4.3 Resumen del Capítulo

En este capítulo se describió el Sistema Conversacional para Despliegue de Información, su estructura, características y funciones principales. La estructura del SCDI mediante menús jerárquicos denota una cierta restricción en sus capacidades al limitar el acceso a la información, teniendo que recorrer primero todas las opciones contenidas en el menú para poder acceder al área deseada. De acuerdo a esta característica se consideró aplicar Procesos de Decisión de Markov (MDPs) que permitan agilizar la visualización de la información presentando más libertad para la interacción con el usuario, como se describirá en el capítulo siguiente.

Al incorporar el uso de Jessica como interfaz gráfica del sistema conversacional que simulara actuar como el rostro del robot anfitrión, proporcionó resultados importantes en las pruebas con usuarios, al revelar que un componente gráfico brinda una mejor interacción al sentirse orientados por un personaje.

Capítulo V. Aplicación de MDPs al Sistema Conversacional para Despliegue de Información

Como se describió en el capítulo anterior, el desempeño del sistema utilizando una estructura jerárquica de menús no era del todo satisfactorio y se deseaba optimizar el funcionamiento del sistema. Debido a esto se decidió aplicar Procesos de Decisión de Markov (MDPs) que permiten tomar la secuencia más adecuada de entre las opciones existentes, de acuerdo a la solicitud del usuario registrada.

Para esto, los diálogos son representados por estados, en donde dada una meta se ejecuta una serie de acciones que irán pasando por diversos estados. De un mismo estado puede haber transiciones a distintos estados dependiendo de la acción ejecutada. Estas acciones deben tener recompensas. Si la acción fue favorable, o sea, que permita acercarse más a la meta fijada, será recompensada con un valor grande. En cambio si la acción fue desfavorable, se le asignará una recompensa menor con el objetivo de que se sigan las rutas mas adecuadas de acuerdo a la meta establecida. Dicha meta esta dada por la solicitud que haya realizado el usuario.

5.1 Aplicando MDPs

Dado la extensión del sistema se optó por abordar el problema mediante MDPs jerárquicos. Lo que significa que se dividió el problema en cuatro MDPs que a su vez estarían controlados por un MDP principal. Esto es, el MDP principal tiene control del acceso a los cuatro MDPs restantes. Estos MDPs consisten de los cuatro submenús contenidos en la estructura original del sistema.

Esto se traduce en que el primer MDP es el submenú que contiene las opciones del *edificio principal*. El segundo MDP contiene las opciones del submenú *aulas*; el tercer MDP contiene las opciones del submenú de *oficinas*; y finalmente el cuarto MDP representa las opciones contenidas en el submenú de *áreas deportivas*.

Como se mencionó anteriormente, para poder acceder a cada uno de los diferentes MDPs se creó un MDP principal que lleva el control al llamado de cada uno de los MDPs que contienen las opciones del sistema, es decir, este MDP se encuentra en el nivel más alto de la jerarquía; tiene la tarea de presentar un mensaje de bienvenida al visitante y listar las cuatro áreas generales en las que se divide el Campus, por lo que cuenta con cinco estados: *inicio*, *edificio principal*, *aulas*, *oficinas* y *áreas deportivas*. El estado *inicio* representa el mensaje de

bienvenida y los demás estados representan los MDPs de las áreas restantes. En este MDP sólo se cuenta con dos acciones, siendo éstas:

1. Elija un área del Campus a conocer.
2. Reiniciar.

En la figura 5.1 se ilustra la estructura jerárquica de los MDPs que se mencionaron. A continuación se describen las consideraciones para modelar el primer MDP.

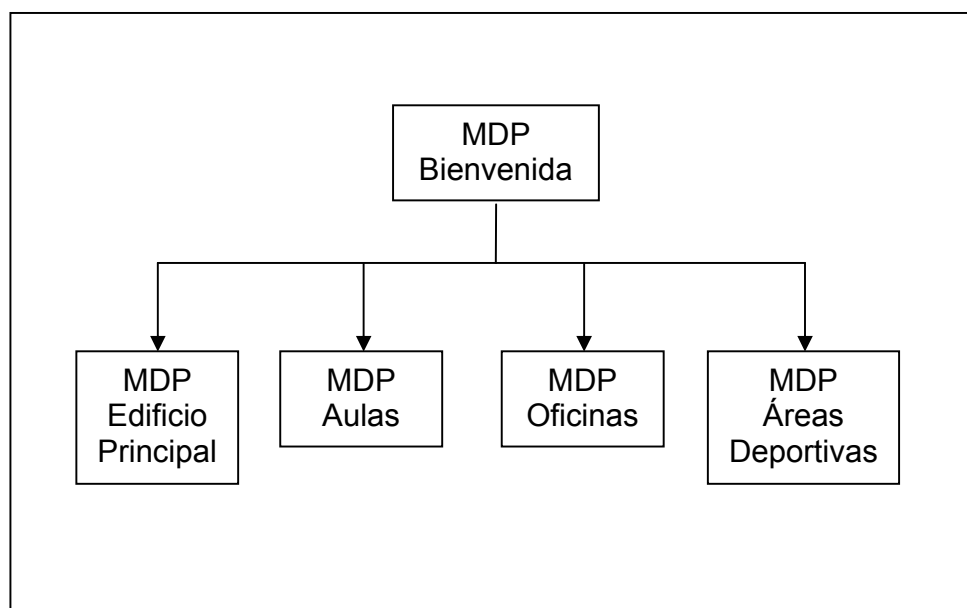


Figura 5.1 Estructura de los MDPs en SCDI.

5.1.1 Modelado del MDP de Edificio principal

Se determinó que los estados del sistema estarían dados de acuerdo a lo que el usuario puede enunciar. Como se mostró en el esquema de la estructura del sistema en el capítulo anterior (figura 4.2), el primer submenú que se encuentra es el del “*edificio principal*” el cual contiene las opciones de: *edificio principal, servicios, departamentos, biblioteca, centro de cómputo, cafetería, dirección general, admisiones, universidad virtual, informática y asuntos internacionales*; y opciones que se derivan de ellas. Estas opciones adicionales son: *video, imagen, uno, dos, tres, si y no*; se decidió incluir estas últimas opciones debido a que son respuestas que el usuario puede responder a las acciones del sistema en los casos en que se solicite visualizar información. Estas opciones fueron identificadas como los estados, lo cual da un total de 18 estados para este submenú en particular. La tabla 5.1 muestra los estados utilizados para este MDP.

Estados
1. Edificio principal
2. Servicios
3. Departamentos
4. Biblioteca
5. Centro de Cómputo
6. Cafetería
7. Dirección general
8. Admisiones
9. Universidad Virtual
10. Informática
11. Asuntos Internacionales
12. Video
13. Imagen
14. Uno
15. Dos
16. Tres
17. Si
18. No

Tabla 5.1 Estados para el submenú “edificio principal”.

En cuanto a las acciones, éstas se determinaron a partir de lo que el sistema decía. Se identificaron en total seis acciones, las cuales se enlistan en la tabla 5.2:

Acciones
1. Elija un área del Campus a conocer
2. Qué área desea conocer
3. Ver video o imagen
4. Video 1 ó 2
5. Imagen 1, 2 ó 3
6. Desea ver otro video o imagen

Tabla 5.2 Acciones para el submenú “edificio principal”.

Esto resultó en la creación de seis matrices de transición, una para cada acción determinada. Considerando el total de estados, cada matriz consta de 18X18 elementos. En la figura 5.2 se muestra la matriz de transición para la acción 1 (Ver anexo A para las matrices de transición de cada acción).

Estados	Edif. Princ	Serv	Dptos	Biblio	CEC	Cafe	Dir gral	Adm.	UV	Info.	Asuntos Int	Video	Img	1	2	3	Si	No
Edif. Princ	0.02	0.03	0.02	0.22	0.2	0.1	0.05	0.12	0.04	0.07	0.06	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Servicios	0.03	0.02	0.03	0.21	0.2	0.11	0.04	0.15	0.03	0.07	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Dptos	0.03	0.03	0.02	0.05	0.04	0.03	0.11	0.21	0.12	0.19	0.1	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Biblioteca	0.02	0.03	0.02	0.02	0.21	0.18	0.04	0.14	0.1	0.12	0.05	0.01	0.01	0.01	0.01	0.01	0.01	0.01
CEC	0.03	0.02	0.02	0.23	0.02	0.17	0.05	0.1	0.08	0.14	0.07	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Cafeteria	0.03	0.02	0.02	0.22	0.19	0.02	0.07	0.14	0.02	0.12	0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Dir gral	0.02	0.02	0.02	0.07	0.05	0.03	0.02	0.22	0.13	0.2	0.15	0.01	0.01	0.01	0.01	0.01		0.01
Adm.	0.03	0.02	0.02	0.08	0.07	0.03	0.12	0.02	0.15	0.21	0.18	0.01	0.01	0.01	0.01	0.01	0.01	0.01
UV	0.02	0.03	0.02	0.09	0.1	0.03	0.11	0.13	0.02	0.23	0.15	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Info	0.02	0.03	0.02	0.04	0.05	0.02	0.18	0.23	0.22	0.02	0.1	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Asuntos Int	0.02	0.03	0.02	0.05	0.04	0.02	0.2	0.23	0.2	0.1	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Video	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Imagen	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.01
1	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.01
2	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.01
3	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Si	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.01
No	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Figura 5.2 Matriz de transición para la acción 1 del menú edificio principal

La relación $A1(\text{estado_origen}, \text{estado_destino})$ de la matriz de transición de la acción 1 (elija un área del Campus) representa la probabilidad de transición del estado_origen al estado_destino, donde el primero se ubica en la fila y el segundo en la columna. Por ejemplo, en la relación $A1(2, 5)$ la probabilidad de que del estado *servicios* vaya al estado *cec*, es de 0.2, mayor a que si fuera al estado *edificio principal*, 0.03 (relación $A1(2, 1)$). La sumatoria de las probabilidades en cada fila da un total de 1. Estas probabilidades de transición se originaron planteando la respuesta más lógica del usuario en cada estado dada la acción a ejecutarse. Esta idea se aplicó en la creación de cada matriz de transición para cada una de las seis acciones identificadas. Cabe mencionar que la asignación de estos valores se realizó de manera subjetiva por parte del autor.

El criterio para la asignación de recompensas fue considerar la relación entre el estado actual y el estado más probable al que se pueda acceder. La recompensa asignada fue de 0.5, mientras que los estados restantes constan con una recompensa de 0.1. Tomando como ejemplo la acción 2 “que deseas ver”, se le designa mayor recompensa a los estados a donde se acceda a sitios específicos, es decir, los sitios finales en donde se puede visualizar información. Estos sitios son: biblioteca, cafetería, cec, dirección general, admisiones, universidad virtual, informática y asuntos internacionales. Una muestra es que cualquier relación $A2(\text{estado_origen}, \text{biblioteca})$ tendrá una recompensa de 0.5 (Ver anexo A para las matrices de recompensa de cada acción).

5.1.2 Modelado del MDP de Aulas

El segundo submenú que se encuentra es el de “aulas” el cual contiene las opciones de: *aulas*, *edificio uno*, *edificio dos*, *edificio tres*, *salones de clases*, *divisiones académicas*, *DEMS (División de Enseñanza Media Superior)*, *DACS (División de Administración y Ciencias Sociales)*, y *DIC (División de Ingeniería y Ciencias)*; y las opciones adicionales de: *video*, *imagen*, *uno*, *dos*, *tres*, *si* y *no*. Estas opciones fueron identificadas como los estados, lo cual da un total de 16 estados para este submenú.

Las acciones son las mismas que se identificaron anteriormente (tabla 5.2):

1. Elija un área del Campus a conocer.
2. Qué área desea conocer.
3. Ver video o imagen.
4. Video 1 o 2.
5. Imagen 1, 2 o 3.
6. Desea ver otro video o imagen.

Para determinar las matrices de transición para cada una de las seis acciones descritas, se siguió el mismo procedimiento indicado en la sección 5.1.1. En este caso las matrices constaron de 16X16 elementos cada una. Para ver todas las

matrices de transición consultar el Anexo A. De la misma manera las recompensas se asignaron considerando la relación entre el estado actual y el estado más probable al que se pueda acceder, con un valor de 0.5 y los estados restantes constan con una recompensa de 0.1. Las matrices de recompensa se pueden consultar en el Anexo A.

5.1.3 Modelado del MDP de Oficinas

El tercer submenú encontrado es el de “*oficinas*” el cual contiene las opciones de: *oficinas, CASA, DAE, tesorería, recursos humanos, y departamento de matemáticas*. Además de las opciones de: *video, imagen, uno, dos, tres, si y no*. Estas opciones fueron identificadas como los estados y dan un total de 13 estados para este submenú.

Las acciones son las mismas seis identificadas para los MDPs anteriores. Para determinar las matrices de transición para cada una de las seis acciones, se siguió el mismo procedimiento indicado en las secciones 5.1.1 y 5.1.2. En este caso las matrices constaron de 13X13 elementos cada una. Para ver todas la matrices de transición, consultar el Anexo A. Las recompensas se asignaron considerando la relación entre el estado actual y el estado más probable al que se pueda acceder, con un valor de 0.5 y los estados restantes constan con una recompensa de 0.1, al igual que en los casos de los MDPs previos (consultar Anexo A).

5.1.4 Modelado del MDP de Áreas deportivas

El cuarto submenú encontrado es el de “*áreas deportivas*” el cual contiene las opciones de: *áreas deportivas, alberca, gimnasio y canchas*. Además de las opciones de: *video, imagen, uno, dos, tres, si y no*. Estas opciones fueron identificadas como los estados y dan un total de 11 estados para este submenú.

Las acciones son las mismas seis identificadas para los MDPs anteriores. Para determinar las matrices de transición para cada una de las seis acciones, se siguió el mismo procedimiento indicado en las secciones 5.1.1, 5.1.2 y 5.1.3. En este caso las matrices constan de 11X11 elementos cada una. Para ver todas la matrices de transición, consultar el Anexo A. Las recompensas se asignaron considerando la relación entre el estado actual y el estado más probable al que se pueda acceder, con un valor de 0.5 y los estados restantes constan con una recompensa de 0.1, al igual que en los casos de los MDPs previos (ver Anexo A).

5.2 Implementación de MDPs

Una vez que se terminó el planteamiento de los MDPs, éstos se resolvieron por el método de iteración de valor, que es uno de los métodos más recurridos para la solución de estos problemas. El resultado, en cada uno, es una matriz indicando la política óptima para cada estado. Estos datos son accedidos por el sistema conversacional durante su ejecución. Por cada estado al que arribe, revisa que acción corresponde ejecutar de acuerdo a la solicitud que haya realizado el usuario. En la figura 5.3 se muestra la política óptima para el MDP *edificio principal*. Por ejemplo si del estado *CEC* se quiere llegar al estado *video*, la acción óptima a realizarse es la acción 3: *ver video o imagen*. Esto significa que la política óptima para acceder de un estado a otro se representa mediante la acción a ejecutar para poder llegar a dicho estado.

El modelado de los MDPs se realizó en Matlab haciendo uso del MDPToolkit del INRA [INRA 2004], el cual proporciona funciones ya implementadas de los métodos para la resolución de MDPs entre otras características. Se almacenó la matriz conteniendo la política óptima en un archivo de texto por cada MDP de manera independiente. Para acceder a cada archivo desde SCDI, se creó un ASP en VBscript que realizara esta labor e indicara las acciones a ejecutar según la política emitida por el programa en Matlab.

Este ASP recibe como parámetros desde SCDI el estado en el que se encuentra actualmente y el estado destino. Con estos datos se verifica la acción correspondiente de acuerdo a la política y la orden es procesada por SCDI. Este proceso se repite cada vez que el sistema accede a un nuevo estado.

5.3 Resumen del Capítulo

En este capítulo se describió el proceso de modelado e implementación de los MDPs propuestos para sustituir la estructura de menús jerárquicos con la que contaba el SCDI.

En el momento de finalizar la implementación se esperaba que el funcionamiento de esta nueva versión con MDPs fuera mucho mejor que la versión inicial de SCDI. Para comprobar esto, se decidió hacer un análisis comparativo de ambas versiones del sistema. Las pruebas con usuarios y sus resultados se describen en el capítulo siguiente.

Estados	Edif.. Princ.	Servicios	Deptos	Biblio	CEC	Café	Dir gral	Admisiones	UV	Informática	Asuntos Int	Video	Imagen	1	2	3	Si	No
Edif Princ	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1
Servicios	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1
Deptos	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1
Biblio	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
CEC	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
Café	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
Dir gral	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
Admisiones	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
UV	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
Informática	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
Asuntos Int	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
Video	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
Imagen	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
1	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
2	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
3	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
Si	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6
No	1	2	2	2	2	2	2	2	2	2	2	3	3	4	4	5	6	6

Figura 5.3 Política obtenida para el MDP de "edificio principal".

Capítulo VI. Experimentos y Resultados

Comúnmente las evaluaciones dentro de la interacción humano-computadora utilizan eficiencia, efectividad y satisfacción del usuario [HCI space 2005] como medidas al evaluar interfaces de usuario. La efectividad es una medida del porcentaje de una tarea que el usuario puede desempeñar a través de la interfaz. La eficiencia es una medida del tiempo que se lleva el usuario en completar una tarea. Las valoraciones de satisfacción se usan para evaluar cómo se siente el usuario al utilizar la interfaz [Scholtz, 2004]. Con base en lo anterior se consideró utilizar las mismas métricas para evaluar el sistema conversacional en sus dos versiones.

6.1. Proceso de Experimentación

Para comprobar que la versión del Sistema Conversacional para Despliegue de Información (SCDI) utilizando MDPs ofrece un mejor funcionamiento, se optó por realizar un estudio de usabilidad comparativo de ambas versiones del sistema, es decir, del SCDI en su primera versión y el SCDI con MDPs. Se propuso que la mitad de los usuarios utilizaran primero la versión inicial del SCDI y después la versión con MDPs, y la mitad restante de los usuarios interactuara con los sistemas en el orden inverso.

Las pruebas se realizaron con 20 usuarios, 8 mujeres y 12 hombres. 16 personas fueron estudiantes del programa de Maestría en Ciencias Computacionales del ITESM Campus Cuernavaca y las 4 restantes personas ajenas al Campus. El rango de edades de los usuarios fue de 22 a 29 años. Se les solicitó que exploraran libremente el sistema y después se les asignó la tarea específica de buscar información sobre la División de Ingeniería y Ciencias (DIC). Se les solicitó que realizaran esto para los dos sistemas. Cabe recordar que los menús presentados en el sistema están estructurados de acuerdo a la ubicación física de las áreas dentro del Campus.

El proceso llevado a cabo para la evaluación se describe a continuación:

1. Inicialmente se le explicó al usuario el funcionamiento del sistema, describiendo sus tareas principales y el entorno para el cual funcionaría.
2. Se aclararon las dudas que pudiera tener el usuario con respecto al funcionamiento del sistema.
3. Se le designa la tarea de explorar el Campus.
4. Se permitió al usuario interactuar con la primera versión de SCDI, realizando la tarea antes mencionada.

5. Se le designa la tarea de obtener información sobre la División de Ingeniería y Ciencias (DIC).
6. Se le explicó que ahora iba a interactuar con otra versión del sistema y las diferencias en cuanto a funcionalidad con respecto al sistema previo.
7. Se permitió al usuario interactuar con la segunda versión de SCDI.
8. Al finalizar la interacción con ambos sistemas se les pidió a los usuarios responder al siguiente cuestionario.

Efectividad

1. ¿Se cumplió el objetivo de explorar el Campus con el primer sistema?
 - Si
 - No
2. ¿Se cumplió el objetivo de explorar el Campus con el segundo sistema?
 - Si
 - No
3. ¿Se cumplió el objetivo de visualizar información sobre la división de ingeniería y ciencias (DIC) con el primer sistema?
 - Si
 - No
4. ¿Se cumplió el objetivo de visualizar información sobre la división de ingeniería y ciencias (DIC) con el segundo sistema?
 - Si
 - No
5. ¿Cuál de los dos sistemas permitió una interacción más libre?
 - El primero
 - El segundo

Eficiencia

6. ¿Cómo sientes que fue el tiempo de tu interacción con el primer sistema?
 - Veloz (poco tiempo)
 - Normal (tiempo adecuado)
 - Lento (mucho tiempo)
7. ¿Cómo sientes que fue el tiempo de tu interacción con el segundo sistema?
 - Veloz (poco tiempo)
 - Normal (tiempo adecuado)
 - Lento (mucho tiempo)

8. Para ti, ¿qué es mejor? ¿Por qué?

- Interactuar más rápido con el sistema, perdiendo algunos detalles.
- Interactuar más lento con el sistema sin perder detalles.

Satisfacción del usuario

9. ¿Estás satisfecho(a) con tu interacción con el primer sistema?

- Mucho
- Normal
- Poco

10. ¿Estás satisfecho(a) con tu interacción con el segundo sistema?

- Mucho
- Normal
- Poco

11. ¿Con qué sistema fue mejor la interacción? ¿Por qué?

- Primero
- Segundo

12. ¿Qué sugieres para mejorar el sistema?

6.2. Resultados

A continuación se resumen los resultados obtenidos.

Efectividad

1. ¿Se cumplió el objetivo de explorar el Campus con el SCDI?

18 personas respondieron “sí” y 2 respondieron “no”.

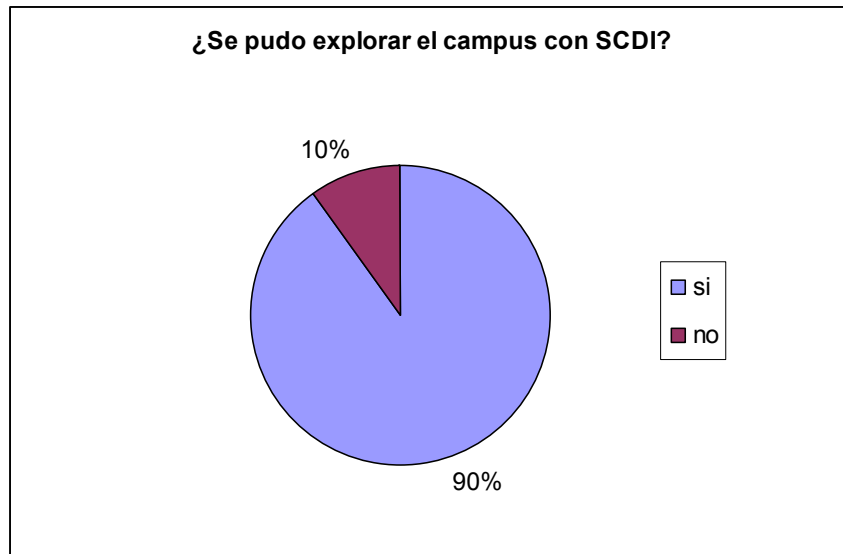


Figura 6.1 Respuesta a la pregunta uno sobre la efectividad del sistema.

2. ¿Se cumplió el objetivo de explorar el Campus con el SCDI con MDPs?

Todas las personas respondieron "si".

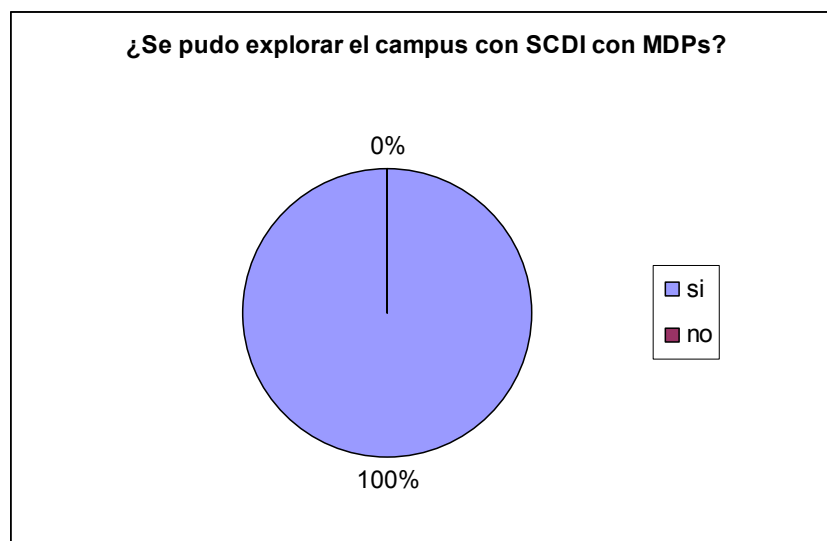


Figura 6.2 Respuesta a la pregunta dos sobre la efectividad del sistema.

3. ¿Se cumplió el objetivo de visualizar información sobre la división de ingeniería y ciencias (DIC) con el SCDI?

12 personas respondieron "si" y 8 respondieron "no".



Figura 6.3 Respuesta a la pregunta tres sobre la efectividad del sistema.

4. ¿Se cumplió el objetivo de visualizar información sobre la división de ingeniería y ciencias (DIC) con el SCDI con MDPs?

14 personas respondieron "sí" y 6 respondieron "no".

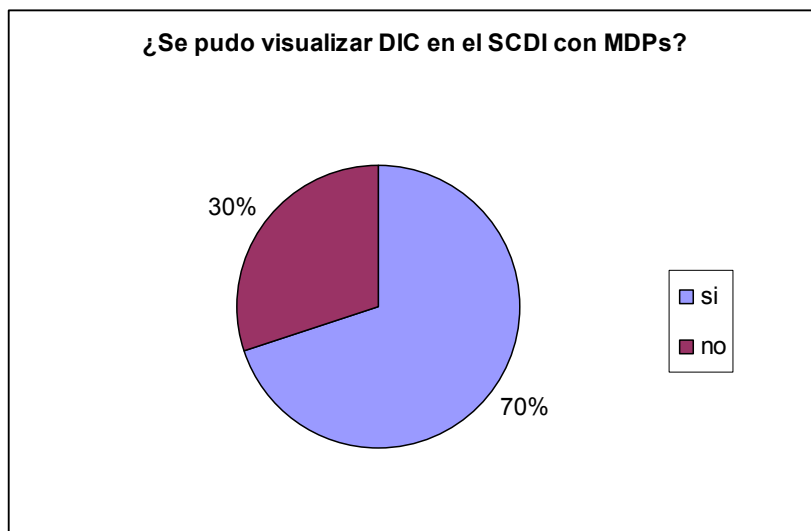


Figura 6.4 Respuesta a la pregunta cuatro sobre la efectividad del sistema.

5. ¿Cuál de los dos sistemas permitió una interacción más libre?

Todas las personas respondieron la versión del SCDI que usa MDPs.

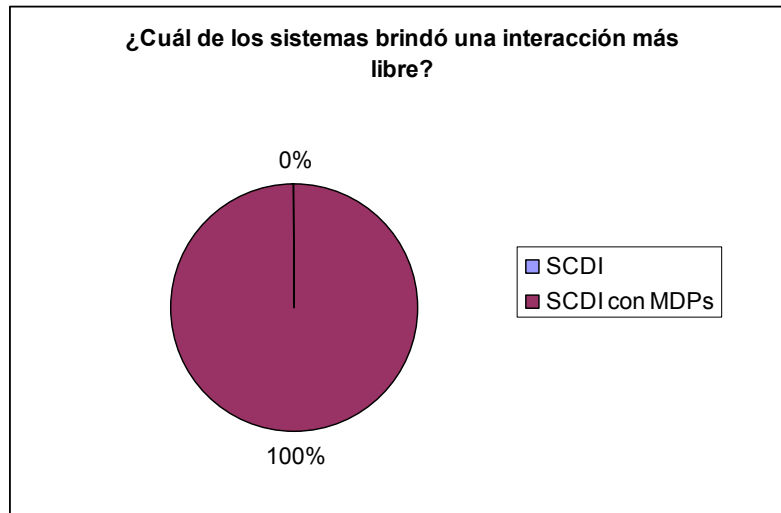


Figura 6.5 Respuesta a la pregunta cinco sobre la efectividad del sistema.

Eficiencia

6. ¿Cómo sientes que fue el tiempo de tu interacción con SCDI?

8 personas respondieron “normal” y 12 respondieron “lento”.

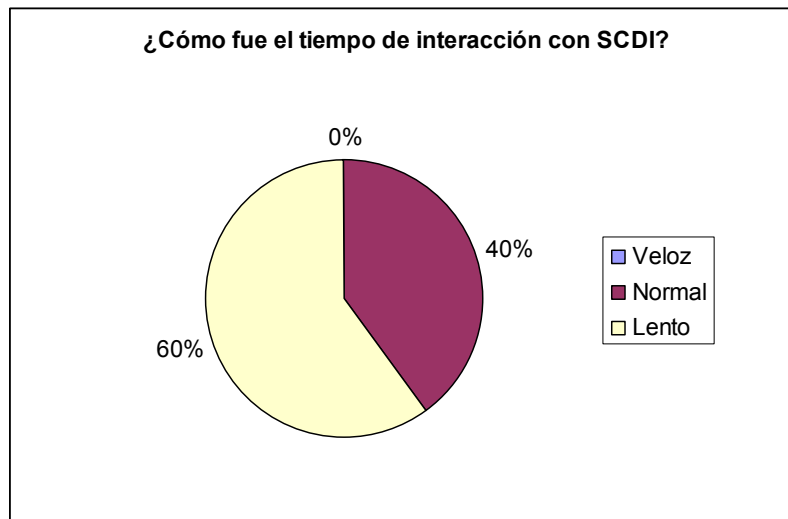


Figura 6.6 Respuesta a la pregunta seis sobre la eficiencia del sistema.

7. ¿Cómo sientes que fue el tiempo de tu interacción con SCDI con MDPs?

2 personas respondieron “veloz”, 16 respondieron “normal” y 2 respondieron “lento”.

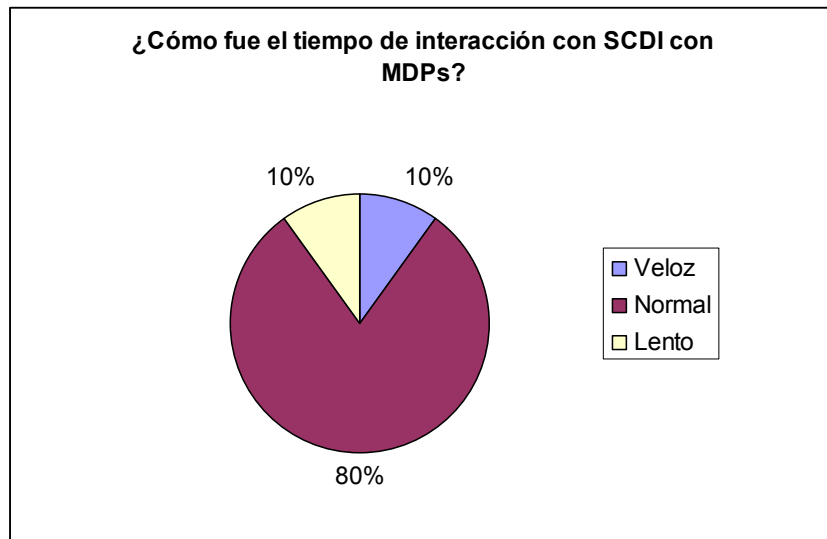


Figura 6.7 Respuesta a la pregunta siete sobre la eficiencia del sistema.

8. Para ti, ¿qué es mejor?

16 personas respondieron “interactuar más rápido con el sistema, perdiendo algunos detalles”, 2 personas “interactuar más lento con el sistema, sin perder detalles” y 2 “ambas”.

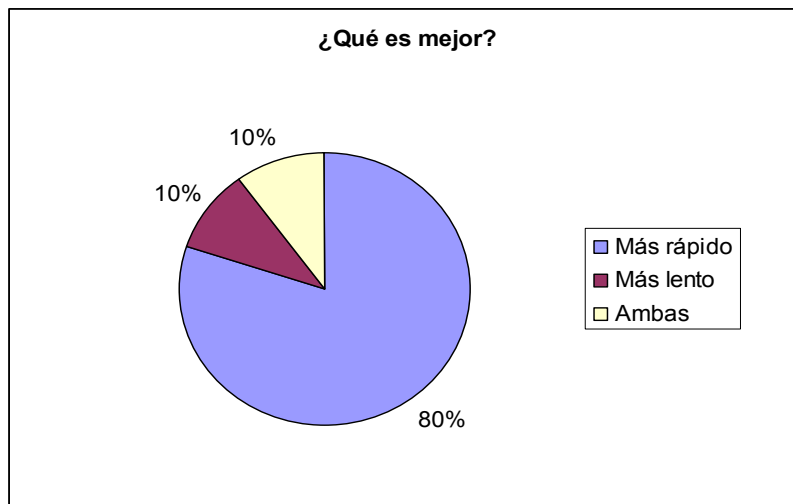


Figura 6.8 Respuesta a la pregunta ocho sobre la eficiencia del sistema.

¿Por qué?

Interactuar más rápido con el sistema, perdiendo algunos detalles.

- No se requiere información que no es de interés, sólo la que es acerca del área buscada (12 personas).
- Si se desea conocer más áreas resulta tedioso esperar mucho tiempo (4 personas).

Interactuar más lento con el sistema, sin perder detalles.

- La finalidad de la consulta es obtener toda la información necesaria (1 persona).
- Un sistema que proporciona información debe dar detalles (1 persona).

Ambas

- Las primeras veces que se interactúa con el sistema es importante conocer los detalles, pero una vez que se familiariza con él, los detalles pierden importancia (2 personas).

Satisfacción del usuario

9. ¿Estás satisfecho(a) con tu interacción con SCDI?

14 personas respondieron “normal” y 6 respondieron “poco”.

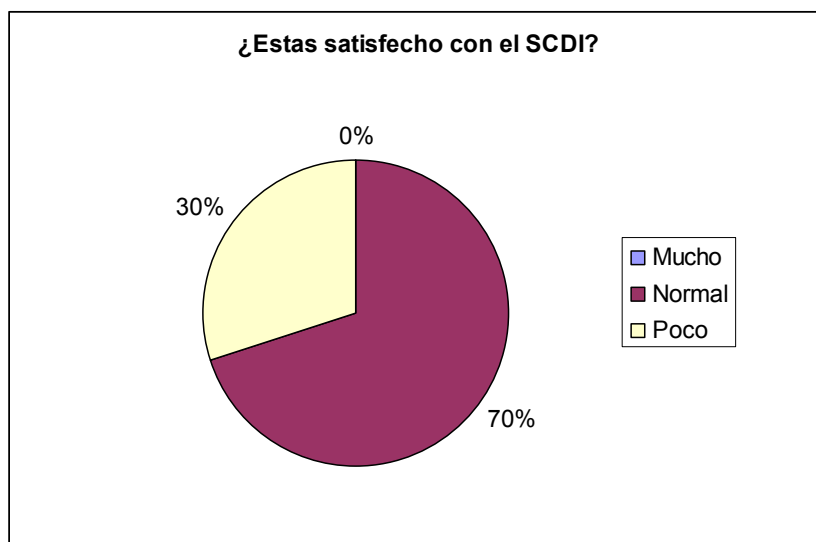


Figura 6.9 Respuesta a la pregunta nueve sobre la satisfacción del usuario.

10. ¿Estás satisfecho(a) con tu interacción con SCDI con MDPs?

6 personas respondieron “mucho” y 14 respondieron “normal”.

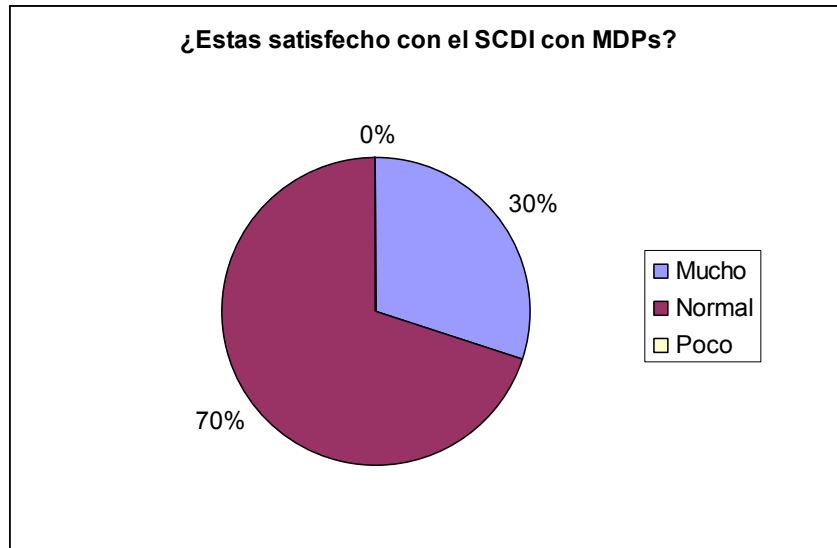


Figura 6.10 Respuesta a la pregunta diez sobre la satisfacción del usuario.

11. ¿Con qué sistema fue mejor la interacción?

Todas las personas respondieron “SCDI con MDPs”.

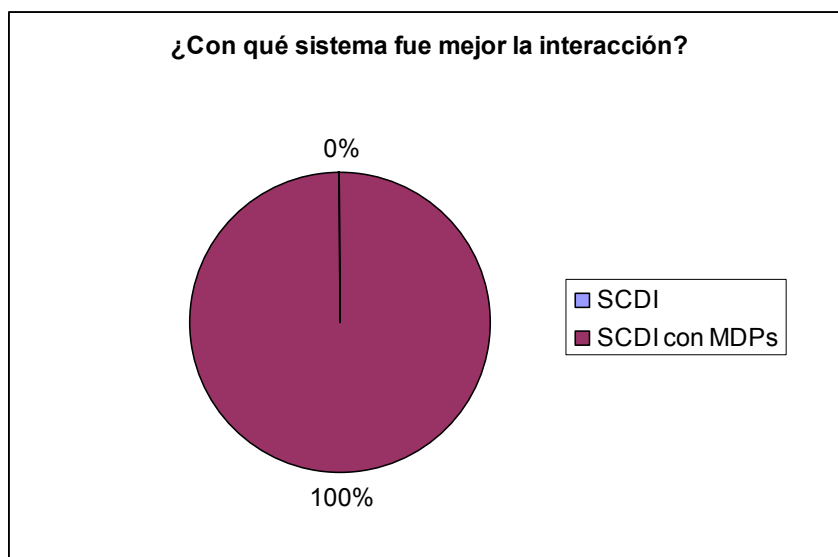


Figura 6.11 Respuesta a la pregunta once sobre la satisfacción del usuario

¿Por qué?

- Fue más flexible al permitir “saltos” dentro de los menús (14).
- Se podía llegar más rápido a la información solicitada (4).
- Permitted explorar mejor el Campus (2).

12. ¿Qué sugieres para mejorar el sistema?

- Visualizar en modo de texto o gráfico en que nivel del menú se encuentra el sistema.
- Tener una opción para repetir las opciones del menú.
- Mejorar la dicción del guía y su velocidad al hablar.
- Mejorar el reconocimiento de la voz del usuario.
- Reiniciar a una imagen neutral en la pantalla para cada cambio de nivel en el menú.
- Mayor rapidez en el despliegue de la información.
- Disminuir el volumen del audio de los videos u omitirlo, ya que no dejan escuchar con claridad al sistema.
- Los videos deberían visualizarse una sola vez.
- Los menús deberían ser menos profundos.
- Las explicaciones narradas deberían ser más a detalle.

Durante la interacción de los usuarios con las dos versiones del sistema conversacional, se monitorearon sus movimientos y se cronometró la actividad realizada por cada tarea. Los resultados en tiempo para el grupo de usuarios 1 (quienes interactuaron primero con SCDI y después con SCDI con MDPs) se muestran en las tablas 6.1 y 6.2, y los resultados para el grupo de usuarios 2 (quienes interactuaron primero con SCDI con MDPs y después con SCDI) se despliegan en las tablas 6.3 y 6.4, aquí cabe mencionar que el dato mostrado para el tiempo máximo de exploración del sistema fue de 22 minutos, sin embargo esto sucedió en el caso de un solo usuario.

Tarea	SCDI	SCDI con MDPs
1. Explorar el Campus		
Tiempo Max.	11 min	8 min
Tiempo Prom.	7 min	5 min
Tiempo Min.	6 min	4 min

Tabla 6.1 Tiempos de interacción para la tarea 1 del grupo de usuarios 1.

Tarea	SCDI		SCDI con MDPs	
	Número	Tiempo	Número	Tiempo
2. Buscar DIC				
Usuarios que completaron la tarea	4	5 min	6	2 min
Usuarios que abandonaron la tarea	6	5 min	4	4 min

Tabla 6.2 Tiempos de interacción para la tarea 2 del grupo de usuarios 1.

Tarea	SCDI con MDPs	SCDI
1. Explorar el Campus		
Tiempo Max.	22 min	7 min
Tiempo Prom.	7 min	4 min
Tiempo Min.	4 min	2 min

Tabla 6.3 Tiempos de interacción para la tarea 1 del grupo de usuarios 2.

Tarea	SCDI		SCDI con MDPs	
	Número	Tiempo	Número	Tiempo
2. Buscar DIC				
Usuarios que completaron la tarea	8	3 min	8	3 min
Usuarios que abandonaron la tarea	2	5 min	2	5 min

Tabla 6.4 Tiempos de interacción para la tarea 2 del grupo de usuarios 2.

Para ilustrar mejor la interacción de los usuarios con el sistema a continuación se presentan las transcripciones de las sesiones con dos usuarios, cada uno de distinto grupo sometido a las pruebas. Las transcripciones presentadas son sólo de la primera interacción con uno de los sistemas.

Usuario 1. Primera interacción con SCDI con MDPs.

Tarea 1: Explorar el Campus.

Sistema: *Hola! Bienvenido al Campus. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?*

Usuario: Edificio principal.

Sistema: El edificio principal...

(el sistema es interrumpido por el usuario)

Usuario: Cancelar.

(el sistema vuelve al punto anterior)

Sistema: Hola! Bienvenido al Campus. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?

Usuario: Aulas.

Sistema: En el Campus contamos con diferentes edificios de aulas. Puedes conocer el edificio uno, edificio dos o edificio tres. ¿Qué deseas conocer?

Usuario: Edificio uno.

Sistema: En el edificio uno podrás encontrar...

(el sistema es interrumpido por el usuario)

Usuario: Edificio dos.

Sistema: Este edificio se localiza enfrente de la alberca y consta de salones de clases en sus 3 pisos. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.

Usuario: Video

Sistema: Tengo dos videos del edificio dos de aulas. Indica uno o dos

Usuario: Dos

(se despliega el video correspondiente)

Sistema: ¿Gustas ver otro video o imagen del edificio dos?

Usuario: Si

Sistema: Indica Video o imagen para ver alguno de los videos o imágenes disponibles.

Usuario: Imagen.

Sistema: Tengo tres imágenes del edificio dos de aulas. Indica uno, dos o tres.

Usuario: Cancelar.

(el sistema vuelve al punto anterior)

Sistema: Indica video o imagen...

(el sistema es interrumpido por el usuario)

Usuario: Cancelar.

(el sistema vuelve al punto anterior)

Sistema: En el Campus contamos con...

(el sistema es interrumpido por el usuario)

Usuario: Salir.

(La tarea se realizó en un total de 6 minutos).

Tarea 2: Buscar DIC.

Sistema: Hola! Bienvenido al Campus. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?

(el sistema detecta ruido)

Sistema: No se entendió. Por favor indica que deseas ver.

(el sistema detecta ruido)

Sistema: Repite por favor.

Usuario: Cancelar.

(el sistema vuelve al punto anterior)

Sistema: Hola! Bienvenido al Campus. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?

Usuario: Oficinas

Sistema: En el Campus contamos con...

(el sistema es interrumpido por el usuario)

Usuario: Cancelar.

(el sistema vuelve al punto anterior)

Sistema: Hola! Bienvenido al Campus. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?

Usuario: Aulas

Sistema: En el Campus contamos con diferentes edificios de aulas. Puedes conocer el edificio uno, edificio dos o edificio tres. ¿Qué deseas conocer?

Usuario: Edificio dos

Sistema: Este edificio se localiza enfrente...

(el sistema es interrumpido por el usuario)

Usuario: Cancelar

(el sistema vuelve al punto anterior)

Sistema: En el Campus contamos con diferentes...

(el sistema es interrumpido por el usuario)

Usuario: Edificio tres

Sistema: Este edificio se encuentra a un lado de la alberca...

(el sistema es interrumpido por el usuario)

Usuario: Cancelar

(el sistema vuelve al punto anterior)

Sistema: En el Campus contamos con diferentes...

(el sistema es interrumpido por el usuario)

Usuario: Edificio uno

Sistema: En el edificio uno podrás encontrar salones de clases y las oficinas de las diferentes divisiones académicas. ¿Qué deseas conocer?

(el usuario dice no entender la dicción del locutor y decide cancelar para volver al estado anterior)

Sistema: En el Campus contamos con diferentes...

(el sistema es interrumpido por el usuario)

Usuario: Edificio uno

Sistema: En el edificio uno podrás encontrar salones...

(el sistema es interrumpido por el usuario)

Usuario: Divisiones académicas

Sistema: Las divisiones que se encuentran en este edificio son: dems, dacs y dic. ¿Qué deseas conocer?

Usuario: DIC

Sistema: La división de ingeniería y ciencias se localiza en el cuarto piso. Comprende los departamentos de ingeniería, ciencias básicas y posgrado en

informática. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.

Usuario: *Imagen.*

Sistema: *Tengo tres imágenes de dic. Indica uno, dos o tres.*

Usuario: *Uno.*

(se despliega la imagen correspondiente)

Sistema: *¿Gustas ver otro video o imagen del edificio dos?*

Usuario: *Salir.*

(El usuario completó la tarea en 5 minutos).

Usuario 2. *Primera interacción con SCDI.*

Tarea 1: *Explorar el Campus.*

Sistema: *Hola! Bienvenido al Campus. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?*

Usuario: *Edificio principal.*

Sistema: *El edificio principal se encuentra junto a la entrada. Alberga diferentes servicios y departamentos. ¿Qué deseas conocer?*

Usuario: *Servicios.*

Sistema: *Los servicios que se ofrecen son: biblioteca, cafetería y centro de cómputo. ¿Qué deseas conocer?*

Usuario: *Biblioteca.*

Sistema: *La biblioteca se localiza en la planta baja y cuenta con una amplia gama de libros, revistas y videos. Su horario es de ocho de la mañana a 11 de la noche. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*

Usuario: *Video.*

Sistema: *Tengo dos videos de la biblioteca. Indica uno o dos.*

Usuario: *Uno.*

(se despliega el video correspondiente)

Sistema: *¿Gustas ver otro video o imagen de la biblioteca?*

Usuario: *No.*

Sistema: *Ahora voy a dormir. Si me necesitas solo di, Robot. Si quieres apagarme di salir.*

Usuario: *Robot.*

Sistema: *Hola! Ya desperté. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?*

Usuario: *Áreas deportivas*

Sistema: *En las instalaciones del Campus contamos con diversas áreas deportivas. Puedes conocer la alberca, el gimnasio, o las diferentes canchas. ¿Qué deseas conocer?*

Usuario: *Gimnasio.*

Sistema: *El gimnasio se localiza del lado derecho a la entrada principal. Su horario es de 7 de la mañana a 10 de la noche. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*

Usuario: *Imagen.*

Sistema: *Tengo tres imágenes del gimnasio. Indica uno, dos o tres.*

Usuario: *Tres.*

(se despliega la imagen correspondiente)

Sistema: *¿Gustas ver otro video o imagen del gimnasio?*

Usuario: *Si.*

Sistema: *El gimnasio se localiza del lado derecho a la entrada principal. Su horario es de 7 de la mañana a 10 de la noche. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*

Usuario: *Imagen.*

Sistema: *Tengo tres imágenes del gimnasio. Indica uno, dos o tres.*

Usuario: *Dos.*

(se despliega la imagen correspondiente)

Sistema: *¿Gustas ver otro video o imagen del gimnasio?*

Usuario: *No.*

Sistema: *Ahora voy a dormir. Si me necesitas solo di, Robot. Si quieres apagarme di salir.*

Usuario: *Robot.*

Sistema: *Hola! Ya desperté. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?*

Usuario: *Edificio principal.*

Sistema: *El edificio principal se encuentra junto a la entrada. Alberga diferentes servicios y departamentos. ¿Qué deseas conocer?*

Usuario: *Departamentos.*

Sistema: *Los departamentos encontrados aquí son: dirección general, admisiones, universidad virtual, informática y asuntos internacionales. ¿Qué deseas conocer?*

Usuario: *Informática.*

Sistema: *El departamento de informática tiene como Objetivo General: Proveer de una adecuada infraestructura tecnológica y servicios de información a la comunidad del Campus, y atender las preocupaciones de los usuarios en un esquema de mejora continua. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*

Usuario: *Imagen.*

Sistema: *Tengo tres imágenes de informática. Indica uno, dos o tres.*

Usuario: *Dos.*

(se despliega la imagen correspondiente)

Sistema: *¿Gustas ver otro video o imagen de informática?*

Usuario: *No.*

Sistema: *Ahora voy a dormir. Si me necesitas solo di, Robot. Si quieres apagarme di salir.*

Usuario: *Robot.*

Sistema: *Hola! Ya desperté. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?*

Usuario: *Áreas deportivas..*

Sistema: *En las instalaciones del Campus contamos con diversas áreas deportivas. Puedes conocer la alberca, el gimnasio, o las diferentes canchas. ¿Qué deseas conocer?*

Usuario: *Alberca.*

Sistema: La alberca se encuentra entre el departamento de matemáticas y el auditorio. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.

Usuario: Imagen.

Sistema: Tengo tres imágenes de informática. Indica uno, dos o tres.

Usuario: Uno.

(se despliega la imagen correspondiente)

Sistema: ¿Gustas ver otro video o imagen de informática?

Usuario: Salir.

(La tarea se realizó en un total de 8 minutos).

Tarea 2: Buscar DIC.

Sistema: Hola! Bienvenido al Campus. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?

Usuario: Oficinas.

Sistema: En el Campus existen diferentes oficinas de servicios. Puedes conocer casa, dae, tesorería, recursos humanos o el departamento de matemáticas.

Usuario: Departamento de matemáticas.

Sistema: En el departamento de matemáticas se encuentran los cubículos de los catedráticos y podrás acudir ahí para asesorías. Se ubica a un costado de la alberca. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.

Usuario: Imagen.

Sistema: Tengo tres imágenes del departamento de matemáticas. Indica uno, dos o tres.

Usuario: Uno.

(se despliega la imagen correspondiente)

Sistema: ¿Gustas ver otro video o imagen del departamento de matemáticas?

Usuario: Salir.

Sistema: Ahora voy a dormir. Si me necesitas solo di, Robot. Si quieres apagarme di salir.

Usuario: Robot.

Sistema: Hola! Ya desperté. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?

Usuario: Oficinas.

Sistema: En el Campus existen diferentes oficinas de servicios. Puedes conocer casa, dae, tesorería, recursos humanos o el departamento de matemáticas.

Usuario: Dae.

Sistema: DAE es la dirección de asuntos estudiantiles. Aquí podrán informarte sobre las actividades culturales y deportivas que puedes realizar en el Campus. Las oficinas se encuentran detrás del estacionamiento. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.

Usuario: Imagen.

Sistema: Tengo tres imágenes de dae. Indica uno, dos o tres.

Usuario: Uno.

(se despliega la imagen correspondiente)

Sistema: ¿Gustas ver otro video o imagen de dae?

Usuario: No.

Sistema: Ahora voy a dormir. Si me necesitas solo di, Robot. Si quieres apagarme di salir.

Usuario: Robot.

Sistema: Hola! Ya desperté. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?

Usuario: Aulas.

Sistema: En el Campus contamos con diferentes edificios de aulas. Puedes conocer el edificio uno, edificio dos o edificio tres. ¿Qué deseas conocer?

Usuario: Edificio uno.

Sistema: En el edificio uno podrás encontrar salones de clases y las oficinas de las diferentes divisiones académicas. ¿Qué deseas conocer?

Usuario: Divisiones

Sistema: Las divisiones que se encuentran en este edificio son: dems, dacs y dic. ¿Qué deseas conocer?

Usuario: DIC

Sistema: La división de ingeniería y ciencias se localiza en el cuarto piso. Comprende los departamentos de ingeniería, ciencias básicas y posgrado en informática. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.

Usuario: Imagen.

Sistema: Tengo tres imágenes de dic. Indica uno, dos o tres.

Usuario: tres.

(se despliega la imagen correspondiente)

Sistema: ¿Gustas ver otro video o imagen del edificio dos?

Usuario: Salir.

(El usuario completó la tarea en 6 minutos).

6.3 Observaciones

A partir de los resultados presentados en la sección anterior se pueden indicar las siguientes observaciones.

La mayoría de los usuarios tendían a ubicar *DIC* en el menú de *oficinas*, otros buscaban en el menú de *edificio principal* y después en la opción de *departamentos*.

Otro factor que tuvo gran influencia en la interacción fue el ruido causado en su mayoría cuando se reproducían videos, en otras ocasiones el ruido fue externo, esto es, del ambiente como gente cercana hablando. En estos casos, los usuarios pedían asistencia al facilitador de la prueba. Incluso hubo usuarios que se distraían en algún punto de la interacción y no escuchaban las opciones del menú

aún cuando ya se habían repetido, en estos casos algunas personas solicitaron asistencia y otras reiniciaban la interacción.

Algunos usuarios hicieron comentarios adicionales, por ejemplo, que el sistema les agradó mucho puesto que era diferente al tipo de interfaz al que están acostumbrados. Otros usuarios dijeron que la interfaz era muy llamativa y que la experiencia de interacción había sido divertida.

Otros comentarios no fueron tan positivos, por ejemplo, el hecho de que el sistema no tuviera un buen reconocimiento de voz, y que la voz del guía en el sistema ocasionalmente iba muy rápido o su dicción era mala.

Un factor interesante durante el monitoreo de la interacción de usuarios, fue la diferencia en el género de la persona. La mayoría de los usuarios hombres tendían a interrumpir el sistema en diversas ocasiones con la intención de apresurar la interacción, mientras que las mujeres fueron más pacientes y esperaban a que el sistema terminara de “hablar”. Durante su primera interacción, las mujeres denotaban más curiosidad por explorar todas las opciones existentes, es decir, visitar la mayor cantidad de áreas posibles y procurar visualizar toda la información correspondiente. Aunque cabe mencionar que el usuario que interactuó durante el mayor período de tiempo registrado, fue un usuario de sexo masculino.

Un caso generalizado indistinto del género, fue que durante la segunda interacción los usuarios ya no se demoraban en explorar el sistema, puesto que ya tenían una idea general de sus funciones y capacidades, aún cuando se les explicó que los sistemas tenían características distintas.

6.4. Análisis

De acuerdo a los resultados arrojados por el cuestionario aplicado a la muestra de usuarios, es evidente que la interacción con el sistema mejora cuando se aplican MDPs, ya que permite que el uso del sistema sea más rápido y ágil. Esto se puede observar en las respuestas a las preguntas 5, 6 y 7, donde la mayoría de los usuarios percibieron tener una interacción más rápida y libre con el sistema que utiliza MDPs.

En lo referente a la pregunta 8, el 80% de los usuarios prefiere una interacción más rápida aunque implique el perder detalles de la información. Esta respuesta pudo ser afectada por la lentitud sobre todo en el despliegue de los videos.

En las respuestas que conciernen la satisfacción del usuario (preguntas 9, 10 y 11), éstas demuestran que la interacción fue abrumadoramente mejor en el sistema que utiliza MDPs, incluso el 30% de los usuarios consideraron que la interacción con este sistema fue muy buena.

Con respecto a las respuestas de la pregunta 12, es interesante observar que las sugerencias más recurrentes tienen que ver con el ruido causado por los videos, lo cual indica que sería mejor omitir el audio para la visualización de los mismos.

En cuanto a los tiempos realizados durante la primera interacción para ambos grupos de usuarios se puede observar que el tiempo promedio para la exploración del sistema fue de 7 minutos en los dos casos. Durante la segunda interacción este tiempo disminuyó para ambos grupos, en el grupo 1 la exploración fue de 4 minutos y en el grupo 2 de 5 minutos.

Para la tarea de encontrar información de DIC se observó que los usuarios del grupo 1 que completaron la tarea en su primer contacto con el sistema, lo hicieron en 5 minutos mientras que los usuarios del grupo 2 lo hicieron en 3 minutos. Esto denota que la interacción con el sistema que usa MDPs es mejor. Así también se puede observar que hubo mayor número de usuarios que completaron la tarea en el grupo 2. Las figuras 6.12 y 6.13 presentan imágenes de un usuario en el ambiente donde se ejecutaron las pruebas.



Figura 6.12 Usuario interactuando con SCDI.



Figura 6.13 Ambiente de realización de pruebas.

6.5. Resumen del Capítulo

En este capítulo se describió la aplicación de un estudio de usabilidad en donde se compara el desempeño de las dos versiones del sistema conversacional. Se presentó el procedimiento llevado a cabo para la prueba con usuarios. Los usuarios fueron divididos en dos grupos, uno que interactuara primero con el SCDI en su versión inicial y después con el SCDI con MDPs; y el segundo grupo que interactuara con los sistemas en el orden inverso. También se enlistaron las preguntas contenidas en el cuestionario aplicado a ambos grupos de usuarios y las respuestas obtenidas.

Los resultados obtenidos a partir del cuestionario demostraron que el sistema conversacional con uso de MDPs, tiene un mejor desempeño con respecto a la primera versión que utiliza menús jerárquicos.

Tal como se esperaba, la interacción fue más rápida con dicha versión y provocó mayor satisfacción para los usuarios. Así también, los resultados revelaron que la poca exposición de los usuarios a interfaces de voz, provoca algo de inseguridad, varios de ellos esperaban una respuesta visual del estado del sistema. Por otro lado la misma inexperiencia con este tipo de sistemas causó expectativa en los usuarios y muchos de ellos lo consideraron bastante atractivo.

Capítulo VII. Conclusiones y Trabajo Futuro

7.1. Conclusiones

Una interfaz es aquella parte de la tecnología con la cual las personas interactúan. La interacción entre una computadora y un usuario es de doble vía. A veces el usuario está dando información a la computadora, y a veces la computadora está dando información al usuario. Esta interacción puede incluir componentes de hardware como teclados o ratones, o componentes de software como una pantalla, ventana, menú, sonido o voz parlante. Una interfaz por voz describe una interfaz de software que emplea ya sea voz humana o la simulación de una voz humana. El uso de interfaces por voz está teniendo gran auge actualmente, ofrecen una alternativa y/o complemento a las interfaces gráficas comunes, además de proporcionar el medio de comunicación habitual entre las personas aplicado ahora a las interfaces computacionales.

En esta tesis se presentó el desarrollo de una aplicación comandada por voz para un robot anfitrión que proporcione información del Campus. El objetivo principal del proyecto fue desarrollar un sistema de diálogos óptimo mediante el uso de MDPs que permitiera interactuar con dicho robot de manera eficiente y con el propósito de presentar información multimedia. Aunque el utilizar aplicaciones basadas en el uso de voz implica problemas cuando se usan dentro de ambientes ruidosos, la manipulación del micrófono por parte del usuario probó tener buenos resultados a la hora de interactuar en estos medios. Esto es, que el usuario pueda prender y apagar el micrófono a voluntad cuando vaya a responder o comandar al sistema.

El diseño del sistema conversacional se basó en un diseño centrado en el usuario. Se intentó soportar dentro de la gramática del sistema una variedad de sinónimos a las opciones presentadas en los menús, con el objetivo de no “abusar” de la memoria a corto plazo del usuario y permitir una interacción más intuitiva. Se procuró presentar un número de opciones mínima en cada nivel del menú al mostrar entre tres y cuatro opciones, precisamente evitando que el usuario tuviera que recordar demasiada información y permitir una interacción más dócil.

Con el fin de imprimirle libertad y rapidez al sistema, se decidió utilizar MDPs que optimizaran su funcionamiento, por lo que la estructura original del sistema fue adaptada para el uso de MDPs jerárquicos, donde un MDP principal controla el acceso a cuatro diferentes MDPs que representan los menús conteniendo las distintas áreas del Campus que se pueden explorar. El uso de MDPs probó brindar una interacción más versátil al usuario, al permitirle regresar y adelantarse a las

opciones presentadas en los menús. Esto proporciona mayor control y velocidad al usuario durante la interacción.

Se realizó un estudio de usabilidad comparando la versión inicial del SCDI con el SCDI con MDPs; las métricas evaluadas fueron las de eficiencia, efectividad y satisfacción del usuario. Con base en los resultados obtenidos durante la aplicación de pruebas con diferentes usuarios, se puede concluir lo siguiente:

En el caso de la métrica de *efectividad* se pudo observar una mejor respuesta por parte de los usuarios hacia el sistema conversacional con uso de MDPs, ya que la mayoría de los encuestados dijeron haber culminado las tareas designadas con dicha versión en comparación con la versión del sistema conversacional original. De igual manera, el 100% de los usuarios afirmó tener una interacción más libre al usar el SCDI con MDPs.

Con respecto a la medida de *eficiencia*, se determinó que el SCDI con MDPs es más eficiente, ya que la mayoría de los usuarios afirmaron tener una interacción entre normal y veloz, mientras que sintieron que la otra versión proporcionaba una interacción entre lenta y normal. La totalidad de los usuarios afirmó tener una interacción más libre con el SCDI con MDPs.

En la medida de *satisfacción del usuario*, la mayoría de las personas sintieron que su interacción con el SCDI brindó una satisfacción entre lo normal y poco satisfactoria, mientras que su interacción con el SCDI con MDPs ofreció una satisfacción entre lo normal y muy satisfactoria.

Un factor importante en la percepción de los usuarios, fue la falta de un apoyo visual para determinar en que punto del menú se ubicaba el sistema. Además la opinión general fue que la voz del guía no tenía buena dicción y a veces hablaba muy rápido. Esto puede deberse en parte a que el sintetizador en español con el que cuenta el *Voice Server* es una versión de Español de España, no de México.

La presentación de la interfaz por voz para el sistema denotó ser novedosa para los usuarios, ya que están poco acostumbrados a trabajar con este tipo de sistemas, lo cual conlleva a inferir que sería muy llamativa también para el público general que visite el Campus.

Los recursos con los que contaba el equipo en donde se encontraba alojado el sistema conversacional estaban un tanto limitados y comprometieron en cierta medida el desempeño del sistema al retrasar su ejecución. Esto fue evidente sobre todo durante la proyección de los videos, ya que a veces existió un retraso al cargarse y desplegarse por primera vez, y durante su proyección tenían un leve efecto de congelamiento.

7.2. Trabajo Futuro

El trabajo que se puede desarrollar a partir del sistema conversacional para despliegue de información es extenso. Puesto que el uso de las interfaces por voz aún no prolifera en comparación con las interfaces gráficas, las mejoras que pueden aplicarse representan un campo muy vasto.

En primera instancia se puede replantear el diseño e implementación de los MDPs propuestos en este sistema. Como se describió a detalle en el capítulo cuatro, las recompensas fueron asignadas en cuanto a alcanzar la meta en la ruta más corta posible. Otro enfoque puede ser el considerar la meta en términos de satisfacción del usuario. Por ejemplo, esperar a recibir retroalimentación del usuario esperando su consentimiento con respecto al estado actual del sistema.

Un aspecto interesante sería modificar la estructura de los MDPs utilizados. Debido a que el diseño actual utiliza MDPs jerárquicos, permite que el usuario recorra un menú de manera fácil y rápida al brindar movimientos libres a través de él. Pero esta característica solo está disponible a lo largo de cada uno de los cuatro menús principales independientemente, por lo que no se permite “saltar” entre las opciones de todos los menús contenidos en el sistema. Por lo que podría proporcionarse más versatilidad y ventaja al sistema al habilitar esta característica.

Se puede enriquecer la gramática soportada por el sistema, ya que como se mencionó antes ayuda a que el usuario tenga una interacción más intuitiva al tener más opciones para acceder a la misma información mediante sinónimos. A pesar de que esto si está contemplado en la estructura gramatical del sistema, sería conveniente verificar que se cubra la mayor cantidad de opciones posibles.

La información multimedia es un tanto limitada en cantidad y en tipo. Para darle mayor atractivo al sistema, pueden incluirse animaciones multimedia con objetos *flash* y presentar información sobre el personal que labora en las distintas áreas del Campus; así como incorporar un mayor número de imágenes y videos.

Una característica que quedó pendiente integrar al sistema, es la inclusión de un mapa general del Campus como alternativa de consulta en la visualización de información. También se consideró presentar la opción de desplegar los mapas correspondientes a las subáreas en cada nivel del menú.

Un rubro que podría replantearse sería la forma de evaluar las métricas de efectividad, eficiencia y satisfacción del usuario en las pruebas de usabilidad, Podría aplicarse la variante de utilizar una escala de Likert (escala para medir actitudes) [Likert 2005] o la de diferencial semántico [Osgood 2005] (método de Osgood). En la técnica Likert se presenta un conjunto de declaraciones de actitudes, donde a los sujetos se les pide expresar si están de acuerdo o en

desacuerdo en una escala de cinco puntos. Sirve para medir actitudes, preferencias y reacciones subjetivas, así que ayudan a obtener las respuestas emocionales y preferenciales de las personas. El diferencial semántico mide las reacciones de la gente a palabras y conceptos de estímulo en términos de calificaciones en escalas bipolares definidas con adjetivos contrastantes en cada extremo, utiliza una escala de siete puntos.

Referencias

[Alarcón 2004] Alarcón A. "Interfaz móvil para visualización de información en ambientes interactivos mediante PDAs". Tesis de maestría. ITESM, Cuernavaca, México, 2004.

[Allen 1999] Allen J. "Mixed-Initiative Interaction". IEEE Intelligent Systems. Volume 15, Issue 5, pp. 14-16, septiembre 1999 .

[ASP 2005] Active Server Pages.
<http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnanchor/html/activeservpages.asp>. Consultado en septiembre de 2005.

[Bartneck 2002] Bartneck C. "eMuu: an emotional embodied character for the ambient intelligent home". Tesis doctoral, Technical University Eindhoven, The Netherlands, 2002.

[Billiard 1997] Billard A. y Dautenhahn K. "Grounding communication in situated, social robots". In Proceedings Towards Intelligent Mobile Robots Conference, Report No. UMCS-97-9-1, Department of Computer Science, Manchester University, 1997.

[Billiard 1998] Billard A. y Dautenhahn K. "Grounding communication in autonomous robots: An experimental study". Robotics and Autonomous Systems Number 24 (1-2) pp. 71-81, 1998.

[Billiard 1999] Billard A. y Dautenhahn K. "Experiments in learning by imitation: Grounding and use of communication in robotic agents", Adaptive Behavior Number 7 (3-4), 1999.

[Black 1999] Black A., Taylor P., Caley R. "The Festival Speech Synthesis System". 1.4 edition, 1999.

[Bombara 2001] Bombara M., Cali D., y Santero C. "KORE: a Multi-Agent System to Assist Museum Visitors". 2001.

[Boyce 2000] Boyce S. "Natural Spoken Dialogue Systems for Telephony Applications". Communications of the ACM vol. 43, no. 9, pp. 29-34, September 2000.

[Breazeal 2002] Breazeal C., "Designing Sociable Robots", MIT Press, Cambridge, MA, 2002.

[Breazeal 2003] Breazeal C., "Toward sociable robots", Robotics and Autonomous Systems Number 42 pp.167–175, 2003.

[Bumby 1999] Bumby K., Dautenhahn K., "Investigating Children's Attitudes Towards Robots: A Case Study", In Proceedings of the Cognitive Technology Conference, 1999.

[Burgard 1998] Burgard W., Cremers A.B., Fox D., Hähnel D., Lakemeyer G., Schulz D., Steiner W., and Thrun, S. "The interactive museum tour-guide robot", In Proceedings of the Fifteenth National Conference on Artificial Intelligence, Madison, Wi, 1998.

[Cahn 1999] Cahn J., "The generation of affect in synthesized speech", Journal of American Voice I/O Society Volume 8 pp. 1–19, 1999.

[Cassell 1999] Cassell J., "Nudge, nudge, wink, wink: Elements of face-to-face conversation for embodied conversational agents", In: J. Cassell, et al. (Eds.), Embodied Conversational Agents, MIT Press, Cambridge, MA, 1999.

[Ciavarella 2002] Ciavarella C., y Paternò F. "Design of a handheld interactive support for museum visitors". Proceedings of DSVIS-02, Rostock, Alemania. 2002.

[CMUSphinx 2005] CMUSphinx. "The Carnegie Mellon Sphinx Project". <http://cmusphinx.sourceforge.net/html/compare.php>. Consultado en septiembre de 2005.

[CU Communicator] CU Communicator. <http://cslr.colorado.edu/beginweb/cumove/cucommunicator.html>. Consultado en septiembre de 2005.

[Danis et al 1994] Danis C., Comerford L., Janke E., Davies K., DeVries J., y Bertran, A. "StoryWriter: A speech oriented editor". In Proceedings of CHI'94: Human Factors in Computing Systems: Conference Companion. ACM Press, New York, pp. 277-278., 1994.

[DiSalvo 2002] DiSalvo C., Gemperle F., Forlizzi J., Kiesler S. "All Robots are not Created Equal: The Design and Perception of Humanoid Robot Heads". In Proceedings of the Conference on Designing Interactive Systems, 2002.

[Edwards 1993] Edwards W. K., Mynatt E., and Rodriguez T., "The Mercator Project: A Nonvisual Interface to the X Window System". In The X Resource, Sebastopol, CA. Issue #7, 1993.

[Elinas 2004] Elinas P., Sucar E., Reyes A., y Hoey J. "A Decision Theoretic Approach for Task Coordination in Social Robots". IEEE RO-MAN, Japan. 2004.

[Fong 2001] Fong T., Thorpe C., Baur C., “Collaboration, dialogue, and human–robot interaction”, In Proceedings of the International Symposium on Robotics Research, 2001.

[Fong 2003a] Fong T. Nourbakhsh I. y Dautenhahn K. “A survey of socially interactive robots”, .Robotics and Autonomous Systems Number 42 pp. 143–166, 2003

[Fong 2003b] Fong T., Thorpe C., Baur C., “Robot, asker of questions”, Robotics and Autonomous Systems Volume 42 pp. 235–243, 2003.

[Graf 2000] Graf B., Schraf R., y Neugebauer J. “A Mobile Robot Platform for Assistance and Entertainment”. Proceedings of ISR-2000, Montreal, pp. 252-253.

[Green 2001] Green A., Severinson K. “Task-oriented Dialogue for CERO: a User-centered Approach”. In proceedings of Ro-Man'01 (10th IEEE International Workshop on Robot and Human Communication), Bordeaux-Paris, pp. 146-151, September 2001.

[Hall 1966] Hall E., “The hidden dimension: Man’s use of space in public and private”, The Bodley Head Ltd., 1966.

[HCI space 2005] HCI space: Usability metrics. <http://www.tau-web.de/hci/space/i8.html>. Consultado en septiembre de 2005.

[Heise 2005] Heise D. “The Semantic Differential and Attitude Research”. <http://www.indiana.edu/~socpsy/papers/AttMeasure/attitude.htm>. Consultado en septiembre de 2005.

[HTK 2005] HTK Speech Recognition Toolkit. <http://htk.eng.cam.ac.uk/>. Consultado en septiembre de 2005.

[INRA 2004] Institut National de la Recherche Agronomique: Markov Decision Process (MDP) Toolbox v2.0 for MATLAB. <http://www.inra.fr/bia/T/MDPtoolbox/>. Consultado en diciembre de 2004.

[IIS 2005] Internet Information Services <http://www.microsoft.com/WindowsServer2003/iis/default.msp>. Consultado en agosto de 2005.

[Johnston 2005] Johnston M., Bangalore S. “Multimodal Applications from Mobile to Kiosk”. In W3C Workshop on Multimodal Interaction 2004 <http://www.w3.org/2004/02/mmi-workshop/mmiwsshopjohnston.pdf>. Consultado en agosto de 2005.

[Jonas 2003] Jonas M., “Modeling Speech Using a Partially Observable Markov Decision Process”. Tesis, TUFTS University, Boston, USA, 2003.

[Katz-Hass 2005] Katz-Haas R., "Usability Techniques: User-Centered Design and Web Development", http://www.stcsig.org/usability/topics/articles/ucd%20_web_devel.html#what_is_UCD. Consultado en julio de 2005.

[Khan 1998] Khan Z., "Attitudes Towards Intelligent Service Robots", Technical Report No. TRITA-NA-P9821, NADA, KTH, Stockholm, Sweden, 1998.

[Kiesler 2002] Kiesler S, Goetz J., "Mental Models and Cooperation with Robotic Assistants". In Proceedings of CHI, 2002.

[Kiesler 2004] Kiesler S., Hinds P. "Human-Robot Interaction". Human-Computer Interaction Journal, Volume 19, Numbers 1 & 2, 2004.

[Koeing 1992] Koenig S., "Optimal probabilistic and decision-theoretic planning using Markovian decision theory". Technical Report UCB/CSD 92/685, Berkeley, May 1992.

[Lai 2000] Lai J. "Conversational Interfaces". Communications of the ACM vol. 43, no. 9, pp. 24-27, September 2000.

[Lamel 2002] Lamel L., Bennacef S., Gauvain J., Dartigues H., Temem J., "User evaluation of the MASK kiosk". Speech Communication No. 38 Issue 1, pp. 131-139, 2002.

[Landsdale, Ormerod 1994] Lansdale M. y Ormerod T., "Understanding Interfaces", Academic Press, New York, 1994.

[Larson 2003] Larson J. "VoiceXML: Introduction to Developing Speech Applications". Prentice Hall, New Jersey. Capítulo 10 pp. 213, 2003.

[Likert 2005] Likert Scale: measurement of attitudes. <http://www.cultsock.ndirect.co.uk/MUHome/cshtml/psy/likert.html>. Consultado en septiembre de 2005.

[Littman 1995a] Littman M., Dean T., Kaelbling L., "On the Complexity of Solving Markov Decision Problems," Proceeding of the Eleventh International Conference on Uncertainty in Artificial Intelligence, 1995.

[Littman 1995b] Littman M., Cassandra A., Kaelbling L., "Efficient dynamic programming updates in partially observable Markov decision processes," Brown University Technical Report CS-95-19, 1995.

[Littman 1995c] Littman M., Cassandra A., Kaelbling L., "Learning policies of partially observable environments: Scaling up," in Armand Prieditis and Stuart

Russell, eds., Proceedings of the Twelfth International Conference on Machine Learning, pp. 362-70, San Francisco, CA, 1995.

[Ly 1993] Ly E., "Chatter: A Conversational Telephone Agent" MIT Master's Thesis, 1993.

[Meuleau 1999] Meuleau N., Kim K., Kaelbling L., Cassandra A., "Solving POMDPs by searching the space of finite policies," In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, CA, pp. 417-426, 1999.

[Montemerlo 2002] Montemerlo M., Pineau J., Roy N., Thrun S., y Verma V. "Experiences with a Mobile Robotic Guide for the Elderly". American Association for Artificial Intelligence. 2002.

[Morales 2004] Morales E., Apuntes de la materia "Búsqueda, Optimización y Aprendizaje".
<http://dns1.mor.itesm.mx/~emorales/Cursos/Busqueda04/principal.html>. 2004

[Muller 1990] Muller M.J. y Daniel J.E. "Toward a definition of voice documents", Proceedings of the Conference on Office Information Systems. pp. 174-183, Cambridge, MA USA 1990.

[Murray 1993] Murray I. y Arnott J. "Towards the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion", Journal of Acoustic Society of America Volume 93 Number2, 1993.

[Museum system 2005] Museum system
<http://it.emory.edu/showdoc.cfm?docid=1541&fr=1062>. Consultado en agosto de 2005.

[Nakauchi 2000] Nakauchi Y. y Simmons R., "A social robot that stands in line", In Proceedings of the International Conference on Intelligent Robots and Systems, 2000.

[Nicolescu 2001] Nicolescu M. y Mataric M., "Learning and interacting in human-robot domains", IEEE Transactions on Systems, Man and Cybernetics Volume 31 Number 5, 2001.

[Nourbakhsh 1999] Nourbakhsh I., Bobenage J., Grange S., Lutz R., Meyer R., y Soto A. "An Affective Mobile Robot Educator with a Full-time Job". Artificial Intelligence Journal Vol. 114, pp. 95-124. 1999.

[Osgood 2005] Osgood: Semantic differential.
<http://www.cultsock.ndirect.co.uk/MUHome/cshtml/introductory/semdif.html>. Consultado en septiembre de 2005.

- [Puterman 1994] Puterman M. "Markov Decision Processes: Discrete Stochastic Dynamic Programming". John Wiley & Sons, Inc., New York, NY, 1994.
- [Ravishankar 1996] Ravishankar M. "Efficient Algorithms for Speech Recognition". Tesis de doctorado. Carnegie Mellon, 1996.
- [Ricaño 2005] Ricaño Y. "Un Rostro Animado para la Interacción Humano-Computadora". Tesis de maestría. ITESM, Cuernavaca, México, 2005.
- [Roy 2000] Roy N., Pineau J., y Thrun S. "Spoken Dialogue Management Using Probabilistic Reasoning". Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000), 2000.
- [Scholtz 2004] Scholtz J., Bahrami S. "Human-Robot Interaction: Development of an Evaluation Methodology for the Bystander Role of Interaction". National Institute of Standards and Technology, 2004.
- [Severinson 2003] Severinson-Eklund K., Green A., Hüttenrauch H., "Social and collaborative aspects of interaction with a service robot", Robotics and Autonomous Systems Volume 42 pp- 223–234, 2003.
- [Sidner 2003] Sidner C., Lee C., Lesh N. "The Role of Dialogue in Human Robot Interaction". In First International Workshop on Language Understanding and Agents for Real World Interaction, Hokkaido University, 2003.
- [SIGCHI 2005] ACM SIGCHI curricula for humana computer interaction. http://sigchi.org/cdg/cdg2.html#2_1. Consultado en agosto 2005.
- [Smithsonian 2005] Smithsonian National Portrait Gallery, Web-based exhibitions. <http://www.npg.si.edu/exhibit/web1.htm>. Consultado en agosto 2005.
- [Spiliotopoulos 2001] Spiliotopoulos D., Androutsopoulos I., y Spyropoulos C. "Human–robot interaction based on spoken natural language dialogue" Proceedings of the European Workshop on Service and Humanoid Robots, 2001.
- [Steels 2001a] Steels L., "Language games for autonomous robots". IEEE Intelligent Systems Volume 16 Number 5, 2001.
- [Steels 2001b] Steels L. y Kaplan F., "AIBO's first words. The social learning of language and meaning", In: H. Gouzoules (Ed.), Evolution of Communications, Volume 4, Number 1, Benjamin, New York, 2001.
- [Stevens 1998] Stevens R. "UNIX Network Programming. Networking APIs: Sockets and XTI". 2a ed. Prentice Hall, New Jersey. 1998.

[Stifelman 1993] Stifelman L., Arons B., Schmandt C. and Hulteen E., "VoiceNotes: A Speech Interface for a Hand-Held Voice Notetaker", Proceedings of INTERCHI, ACM, New York, pp. 179-186. 1993

[Strong 1993] Strong B., "Casper: Speech Interface for the Macintosh." Proceedings of EuroSpeech, Berlin, Alemania, 1993.

[Sucar 2005] Sucar E., Apuntes de la materia "Razonamiento con Incertidumbre". 2005.

[Sutton 1990] Sutton R., "Integrated Architectures for Learning, Planning, and Reacting Based on Approximating Dynamic Programming," in Proceedings of the Seventh International Conference on Machine Learning, pp. 216-24, 1990.

[Sutton 1998] Sutton S., Cole R., de Villiers J., Schalkwyk J., Vermeulen P., Macon M., Yan Y., Kaiser E., Rundle B., Shobaki K., Hosom P., Kain A., Wouters J., Massaro M. and Cohen M., "Universal Speech Tools: the CSLU Toolkit", in Proceedings of the International Conference on Spoken Language Processing (ICSLP), pp. 3221-3224, Proceedings of the International Conference on Spoken Language Processing (ICSLP), Sydney, Australia, Nov, 1998.

[Thrun 1999] Thrun S., Bennewitz M., Burgard W., Cremers A., Dellaert F., Fox D., Hähnel D., Rosenberg C., Roy N., Schulte J., y Schulz D. "MINERVA: A Second-Generation Museum Tour-Guide Robot" 1999

[Usability glossary 2005] Usability first: usability glossary. http://www.usabilityfirst.com/glossary/main.cgi?function=display_term&term_id=969. Consultado en septiembre de 2005.

[ViaVoice 2005] IBM ViaVoice. <http://www-306.ibm.com/software/voice/viavoice/>. Consultado en septiembre de 2005.

[VMC 2005] Virtual Museum Canada. <http://www.virtualmuseum.ca/>. Consultado en septiembre de 2005.

[VoiceXML Forum 2005] VoiceXML Forum. <http://www.voicexml.org/overview.html>. Consultado en octubre de 2005.

[VXML Manual 2004] Voice eXtensible Markup Language VoiceXML. Version: 1.00, <http://www.voicexml.org/specs/VoiceXML-100.pdf>. Consultado en noviembre de 2004.

[Yankelovich 1994] Yankelovich N., Levow G. and Marx M., "Designing SpeechActs: Issues in Speech Interfaces", Proceedings of CHI 95, Denver, CO, May, 1995.

[Zue 1994] Zue V., Seneff S., Polifroni J., Phillips M., Pao C., Goddeau D., Glass J. and Brill E., "Pegasus: A spoken Language Interface for On-Line Travel Planning." Presented at the ARPA Human Language Technology Workshop, Merrill Lynch Conference Center, Princeton NJ, March, 1994.

Anexo A. Modelo de los MDPs

En este apartado se presentan las matrices de transición y de recompensa para cada acción utilizadas para modelar cada uno de los MDPs empleados en el SCDI. Esto es, se presentarán dos matrices de transición y dos de recompensa para el MDP de bienvenida, además de incluir seis matrices de transición y seis de recompensa que corresponden a cada una de las acciones para los MDPs siguientes:

1. Edificio principal
2. Aulas
3. Oficinas
4. Áreas deportivas

Tabla A.1. Matriz de transición para la acción 1 del MDP Bienvenida

Acción1: Elija un área del campus a conocer

Estados	Inicio	Edif. Principal	Aulas	Oficinas	Areas Deportivas
Inicio	0.01	0.27	0.32	0.18	0.22
Edif. Princ.	0.1	0.01	0.35	0.24	0.3
Aulas	0.1	0.37	0.01	0.22	0.3
Oficinas	0.1	0.35	0.32	0.01	0.22
Areas Deportivas	0.1	0.32	0.3	0.27	0.01

Tabla A.2. Matriz de transición para la acción 2 del MDP Bienvenida

Acción2: Reiniciar

Estados	Inicio	Edif. Principal	Aulas	Oficinas	Areas Deportivas
Inicio	0.6	0.1	0.1	0.1	0.1
Edif. Principal	0.6	0.1	0.1	0.1	0.1
Aulas	0.6	0.1	0.1	0.1	0.1
Oficinas	0.6	0.1	0.1	0.1	0.1
Areas Deportivas	0.6	0.1	0.1	0.1	0.1

Tabla A.3. Matriz de recompensa para la acción 1 del MDP Bienvenida

Estados	Inicio	Edif. Principal	Aulas	Oficinas	Areas Deportivas
Inicio	0.1	0.5	0.5	0.5	0.5
Edif. Principal	0.1	0.5	0.5	0.5	0.5
Aulas	0.1	0.5	0.5	0.5	0.5
Oficinas	0.1	0.5	0.5	0.5	0.5
Areas Deportivas	0.1	0.5	0.5	0.5	0.5

Tabla A.4. Matriz de recompensa para la acción 2 del MDP Bienvenida

Estados	Inicio	Edif. Principal	Aulas	Oficinas	Areas Deportivas
Inicio	0.5	0.1	0.1	0.1	0.1
Edif. Principal	0.5	0.1	0.1	0.1	0.1
Aulas	0.5	0.1	0.1	0.1	0.1
Oficinas	0.5	0.1	0.1	0.1	0.1
Areas Deportivas	0.5	0.1	0.1	0.1	0.1

Anexo B. Diálogos del SCDI

Esta sección incluye la descripción de los diálogos incluidos en el SCDI y la gramática que soporta dicho sistema. En cualquier punto de la interacción están activos los comandos de inicio, cancelar y salir. *Inicio* lleva al sistema al diálogo inicial en donde se da la bienvenida al usuario. *Cancelar* lleva al sistema al punto anterior de la interacción. *Salir* pone el sistema a “dormir” en donde en realidad está en un estado de escucha esperando la palabra clave “robot” para reiniciar la interacción.

1. MDP de Bienvenida

Para el MDP de bienvenida se tiene una enunciación del sistema:

Hola! Bienvenido al Campus. Te puedo presentar cuatro áreas principales: el edificio principal, aulas, oficinas o áreas deportivas. ¿Qué deseas conocer?

La gramática que se soporta en esta etapa es: *Edificio, edificio principal, aulas, oficinas, áreas deportivas, deportes.*

2. MDP de edificio principal

Para el MDP del edificio principal se incluye la siguiente gramática: *edificio principal, servicios, departamentos, biblioteca, cafetería, cec, centro de computo, asuntos internacionales, admisiones, informática, universidad virtual, dirección general, dirección, imagen, video, uno, dos, tres, si, no.*

Las enunciaciones del sistema para el MDP de edificio principal:

- *El edificio principal se encuentra junto a la entrada. Alberga diferentes servicios y departamentos. ¿Qué deseas conocer?*
- *Los servicios que se ofrecen son: biblioteca, cafetería y centro de cómputo. ¿Qué deseas conocer?*
- *La biblioteca se localiza en la planta baja y cuenta con una amplia gama de libros, revistas y videos. Su horario es de ocho de la mañana a 11 de la noche. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de la biblioteca. Indica uno o dos.*
- *Tengo tres imágenes de la biblioteca. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de la biblioteca?*
- *La cafetería se encuentra en la entrada del edificio y su horario es de ocho de la mañana a 8 de la noche. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de la cafetería. Indica uno o dos.*
- *Tengo tres imágenes de la cafetería. Indica uno, dos o tres.*

- *¿Gustas ver otro video o imagen de la cafetería?*
- *El centro de cómputo cuenta con dos salas. Una en la planta baja y otra en el primer piso. Cuenta con servicio de impresión, escaneo, respaldo de información. Su horario es de ocho de la mañana a 11 de la noche. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos del centro de cómputo. Indica uno o dos.*
- *Tengo tres imágenes del centro de cómputo. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen del centro de cómputo?*
- *Los departamentos encontrados aquí son: dirección general, admisiones, universidad virtual, informática y asuntos internacionales. ¿Qué deseas conocer?*
- *La dirección general es donde se toman las decisiones más importantes para el Campus. Aquí puedes encontrar a nuestra rectora, la licenciada Alejandra Vilalta. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de la dirección. Indica uno o dos.*
- *Tengo tres imágenes de la dirección. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de la dirección?*
- *En el centro de admisiones puedes conocer los programas que se ofrecen en el Campus y el proceso para ingresar a cualquiera de ellos. También ofrece información sobre becas y programas de financiamiento. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de admisiones. Indica uno o dos.*
- *Tengo tres imágenes de admisiones. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de admisiones?*
- *La universidad virtual, ofrece una gran variedad de programas de maestría y de educación continua, así como de capacitación de profesores, tanto del Tecnológico de Monterrey como de los sistemas educativos de México. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de la universidad virtual. Indica uno o dos.*
- *Tengo tres imágenes de la universidad virtual. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de la universidad virtual?*
- *El departamento de informática tiene como Objetivo General: Proveer de una adecuada infraestructura tecnológica y servicios de información a la comunidad del Campus. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de informática. Indica uno o dos.*
- *Tengo tres imágenes de informática. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de informática?*
- *En el departamento de asuntos internacionales atendemos las necesidades de toda nuestra comunidad estudiantil que se encuentra de intercambio en nuestro Campus. También te informamos sobre*

oportunidades para que estudies en el extranjero. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.

- *Tengo dos videos de asuntos internacionales. Indica uno o dos.*
- *Tengo tres imágenes de asuntos internacionales. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de asuntos internacionales?*

3. MDP de aulas

Para el MDP de aulas se incluye la siguiente gramática: *aulas, edificio uno, edificio dos, edificio tres, salones, clases, divisiones, direcciones, divisiones académicas, direcciones académicas, dems, dacs, dic, imagen, video, uno, dos, tres, si, no.*

Las enunciaciones del sistema para el MDP de aulas:

- *En el Campus contamos con diferentes edificios de aulas. Puedes conocer el edificio uno, edificio dos o edificio tres. ¿Qué deseas conocer?*
- *En el edificio uno podrás encontrar salones de clases y las oficinas de las diferentes divisiones académicas. ¿Qué deseas conocer?*
- *Los salones de clases se encuentran en los 4 pisos del edificio y están destinados para el alumnado de preparatoria y profesional. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de los salones de clases. Indica uno o dos.*
- *Tengo tres imágenes de los salones de clases. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de los salones de clases?*
- *Las divisiones que se encuentran en este edificio son: dems, dacs y dic. ¿Qué deseas conocer?*
- *La división de educación media superior se localiza en el segundo piso y dirige el área relacionada con la preparatoria. Aquí encontraras la información sobre los planes de estudio ofrecidos en el Campus. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de dems. Indica uno o dos.*
- *Tengo tres imágenes de dems. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de dems?*
- *La división de administración y ciencias sociales se localiza en el tercer piso y se encarga de coordinar las carreras de administración de empresas, mercadotecnia, comercio internacional, contaduría y ciencias de la comunicación. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de dacs. Indica uno o dos.*
- *Tengo tres imágenes de dacs. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de dacs?*

- *La división de ingeniería y ciencias se localiza en el cuarto piso. Comprende los departamentos de ingeniería, ciencias básicas y posgrado en informática. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de dic. Indica uno o dos.*
- *Tengo tres imágenes de dic. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de dic?*
- *Este edificio se localiza enfrente de la alberca y consta de salones d clases en sus 3 pisos. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos del edificio dos. Indica uno o dos.*
- *Tengo tres imágenes del edificio dos. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen del edificio dos?*
- *Este edificio se encuentra a un lado de la alberca y junto al edificio dos de aulas. Además de contar con salones, en el se localiza el auditorio. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos del edificio tres. Indica uno o dos.*
- *Tengo tres imágenes del edificio tres. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen del edificio tres?*

4. MDP de oficinas

Para el MDP de oficinas se incluye la siguiente gramática: *oficinas, casa, dae| tesorería, recursos humanos, recursos, departamento de matemáticas, matemáticas, imagen, video, uno, dos, tres, si, no.*

Las enunciaciones del sistema para el MDP de oficinas son:

- *En el Campus contamos con diferentes oficinas de servicios. Puedes conocer casa, dae, tesorería, recursos humanos o el departamento de matemáticas. ¿Qué deseas conocer?*
- *CASA es el centro de atención y servicios a alumnos. Tiene como misión atender las preocupaciones y solicitudes de los alumnos y egresados. Se ubica a un costado de la alberca en la planta alta. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de casa. Indica uno o dos.*
- *Tengo tres imágenes de casa. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de casa?*
- *DAE es la dirección de asuntos estudiantiles. Aquí podrán informarte sobre las actividades culturales y deportivas que puedes realizar en el Campus. Las oficinas se encuentran detrás del estacionamiento. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de dae. Indica uno o dos.*
- *Tengo tres imágenes de dae. Indica uno, dos o tres.*

- *¿Gustas ver otro video o imagen de dae?*
- *En tesorería te podrán informar sobre el estatus de tus pagos para colegiaturas o tu saldo sobre adeudos. Se ubica detrás del gimnasio. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de tesorería. Indica uno o dos.*
- *Tengo tres imágenes de tesorería. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de tesorería?*
- *Recursos humanos es el departamento que lleva el control sobre el personal que labora en el Campus. Se ubica detrás del gimnasio. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de recursos humanos. Indica uno o dos.*
- *Tengo tres imágenes de recursos humanos. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de recursos humanos?*
- *En el departamento de matemáticas se encuentran los cubículos de los catedráticos y podrás acudir ahí para asesorías. Se ubica a un costado de la alberca. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos del departamento de matemáticas. Indica uno o dos.*
- *Tengo tres imágenes del departamento de matemáticas. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen del departamento de matemáticas?*

5. MDP de áreas deportivas

Para el MDP de áreas deportivas se incluye la siguiente gramática: *áreas, áreas deportivas, deportes, alberca, gimnasio, gym, canchas, imagen, video, uno, dos, tres, si, no.*

Las enunciaciones del sistema para el MDP de áreas deportivas son:

- *En las instalaciones del Campus contamos con diversas areas deportivas. Puedes conocer la alberca, el gimnasio, o las diferentes canchas. ¿Qué deseas conocer?*
- *La alberca se encuentra entre el departamento de matemáticas y el auditorio. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de la alberca. Indica uno o dos.*
- *Tengo tres imágenes de la alberca. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de la alberca?*
- *El gimnasio se localiza del lado derecho a la entrada principal. Su horario es de 7 de la mañana a 10 de la noche. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos del gimnasio. Indica uno o dos.*

- *Tengo tres imágenes del gimnasio. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen del gimnasio?*
- *En el Campus se cuenta con canchas para diferentes deportes. Como basket, tenis, volley y futbol. Indica Video o imagen para ver alguno de los videos o imágenes disponibles.*
- *Tengo dos videos de las canchas. Indica uno o dos.*
- *Tengo tres imágenes de las canchas. Indica uno, dos o tres.*
- *¿Gustas ver otro video o imagen de las canchas?*

Anexo C. Diseño e Implementación del SCDI

Este apartado consta de los detalles técnicos considerados para la implementación de SCDI en cuanto a programación y ubicación física de los módulos.

De acuerdo a [Alarcón 2004] el diseño de la aplicación “Interfaz móvil para visualización de información en ambientes interactivos mediante PDAs” requiere del uso del directorio *wwwroot* dentro de *Inetpub* incluido en IIS. En él se crean las carpetas de *admin*, *app*, *avis*, *db* y *images* para el uso de dicha aplicación. El módulo correspondiente a la proyección de la información se localiza en el directorio *db*. Para facilitar el uso del módulo mencionado, se optó por plantar el SCDI en ese mismo directorio.

Como se muestra en la figura C.1 dentro del directorio *db* se localizan los directorios *Bienvenida*, *MDP*, *MDP2*, *MDP3* y *MDP4*. El directorio de *Bienvenida* contiene el MDP de control de acceso a los MDPs restantes, éstos a su vez están localizados como sigue: en *MDP* el submenú de *edificio principal*, en *MDP2* el de *aulas*, en *MDP3* el de *oficinas* y en *MDP4* el de *áreas deportivas*.

Dentro de cada directorio de MDP se encuentra el archivo que presenta la política óptima para ese MDP correspondiente. Este es un archivo de texto denominado “política”. Para acceder a este archivo se utiliza el módulo “checa_política” que recibe como parámetros el estado origen y el estado destino para poder ubicar la acción correspondiente. Los estados origen y destino estarán dados de acuerdo a las enunciaciones del usuario. Para identificar los estados, se les determinó un número por orden de aparición en la estructura del menú, como se muestra en la tabla C.1, así si el sistema se encuentra en el estado “Servicios” y el usuario quiere acceder a la “Cafetería”, el estado origen corresponde al número 2 y el estado destino al número 6. Un ejemplo de esto se muestra en el fragmento de código del “edificio principal” de la figura C.2.

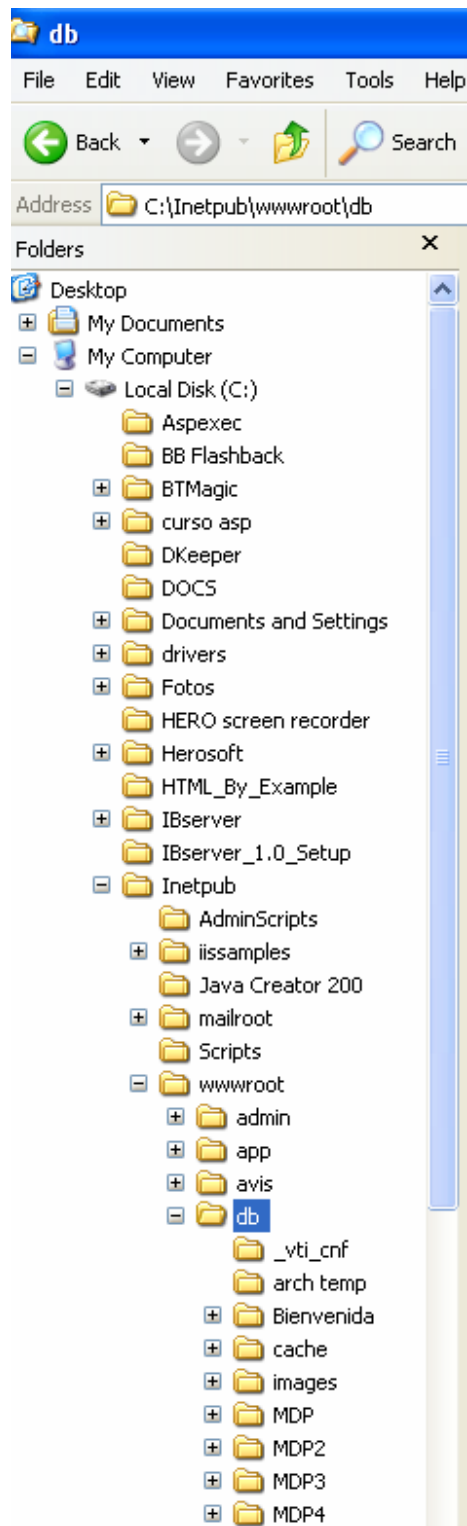


Figura C.1. Estructura de directorios para SCDI.

Estados
1. Edificio principal
2. Servicios
3. Departamentos
4. Biblioteca
5. Centro de Cómputo
6. Cafetería
7. Dirección general
8. Admisiones
9. Universidad Virtual
10. Informática
11. Asuntos Internacionales
12. Video
13. Imagen
14. Uno
15. Dos
16. Tres
17. Si
18. No

Tabla C.1. Asignación de números de identificación para los estados de “edificio principal”.

```

<if cond="edificio=='edificio principal'">
  <assign name="origen"          expr="1"/>
  <assign name="destino"        expr="1"/>
  <submit next= "http://dana/db/MDP/checa_politica.asp" method= "post" namelist= "origen
destino"/>
</if>
<if cond="edificio=='servicios'">
  <assign name="origen"          expr="1"/>
  <assign name="destino"        expr="2"/>
  <submit next= "http://dana/db/MDP/checa_politica.asp" method= "post" namelist= "origen
destino"/>
</if>
<if cond="edificio=='departamentos'">
  <assign name="origen"          expr="1"/>
  <assign name="destino"        expr="3"/>
  <submit next= "http://dana/db/MDP/checa_politica.asp" method= "post" namelist= "origen
destino"/>
</if>
<if cond="edificio=='biblioteca'">
  <assign name="origen"          expr="1"/>
  <assign name="destino"        expr="4"/>
  <submit next= "http://dana/db/MDP/checa_politica.asp" method= "post" namelist= "origen
destino"/>
</if>

```

Figura C.2. Fragmento de código de “edificio principal”.